

# Achieving the Potential of Health Care Performance Measures

## Timely Analysis of Immediate Health Policy Issues May 2013

Robert A. Berenson, Peter J. Pronovost, and Harlan M. Krumholz

*The United States is on the cusp of a new era, with greater demand for performance information, greater data availability, and a greater willingness to integrate performance information into public policy. This era has immense promise to deliver a learning health care system that encourages collaborative improvements in systems-based care, improves accountability, helps consumers make important choices, and improves quality at an acceptable cost. However, to curtail the possibility of unintended adverse consequences, it is important that we invest in developing sound measures, understand quality measures' strengths and limitations, study the science of quality measurement, and reduce inaccurate inferences about provider performance.*

### Introduction

There is a consensus that scientifically rigorous and valid measurement of performance can be instrumental in improving value in U.S. health care.<sup>1</sup> In particular clinical areas, such as cardiac and intensive care, measurement has been associated with important improvements in providers' use of evidence-based strategies and patients' health outcomes.<sup>2</sup> Perhaps most important, measures have altered the culture of health care delivery for the better, with a growing acceptance that clinical practice can be objectively assessed and improved. Nevertheless, despite notable successes and the recent cultural change, substantial shortcomings in the quality of U.S. health care persist.<sup>3</sup> Furthermore, the growth of performance measurement has been accompanied by increasing concerns about heterogeneity in the scientific rigor, transparency, and limitations of available measure sets, and how measures should be used to provide proper incentives to improve performance. The challenge ahead is to achieve the promise of measurement while avoiding the potential for unintended adverse consequences.

Many conceptual and operational measurement challenges have become apparent in recent years. The limited scope of available measures, defects in particular measures, and invalid inferences that have been made based on available measures have compromised the potential usefulness of some measurement efforts for consumers, health professionals, and payers. Many individuals and organizations have also expressed concerns about the application of measures in payment policies that do not

precisely discriminate differences in quality, leading to misclassification. Standards for measurement and their application for public policy are evolving, with controversies flaring over various technical issues. In an environment where both reputation and dollars depend on measured performance, it is often difficult to disentangle the legitimate concerns of those being measured from self-serving defenses of the status quo.

Despite these concerns, the promotion of public reporting and pay-for-performance is growing, even as a number of studies have shown that some of the most prominent applications of measures in the United States have not met their performance improvement objectives.<sup>4</sup> For example, the largest U.S. test of the combined use of public reporting and pay-for-performance, called the Medicare Premier Hospital Quality Incentive Demonstration, has had little or no impact on the value of care received for important clinical conditions; the demonstration neither reduced patient mortality nor cost growth.<sup>5</sup> Yet, based on face validity of the concept, expectations for success, and perhaps premature claims of cost savings,<sup>6</sup> Congress mandated a Hospital Value-Based Purchasing Program, under which hospital performance is rewarded or penalized with altered marginal payments.

In this paper, we first examine the measurement enterprise, including which organizations develop measures and how payers are using measures in their programs, with a special focus on Medicare, which some contend has been in the lead on using measurement. Next, we summarize the mechanics of

*“In an environment where both reputation and dollars depend on measured performance, it is often difficult to disentangle the legitimate concerns of those being measured from self-serving defenses of the status quo.”*

performance measurement by reviewing the characteristics of structure, process, and outcome measures, and the data required to calculate these measures. We also review the successes and failures of some current applications of performance measurement, with an emphasis on the lack of success of pay-for-performance approaches and the threat to intrinsic motivation that such an approach represents. Then, we assess the problems inherent in the United States’ current reliance on clinical process measures, and explore the substantial challenges of moving to outcome measures.

Based on these findings, we offer seven policy recommendations for achieving the potential of performance measurement. Specifically, we present the case that leaders in the public and private sectors need to:

1. Decisively move from measuring processes to outcomes;
2. Use quality measures strategically, adopting other quality improvement approaches where measures fall short;
3. Measure quality at the level of the organization, rather than the clinician;
4. Measure patient experience with care and patient-reported outcomes as ends in themselves;
5. Use measurement to promote the concept of the rapid-learning health care system;
6. Invest in the “basic science” of measurement development and applications, including an emphasis on anticipating and preventing unintended adverse consequences; and
7. Task a single entity with defining standards for measuring and reporting quality and cost data, similar to the role the Securities and Exchange Commission (SEC) serves for the reporting of corporate financial data, to improve the validity, comparability, and transparency of publicly-reported health care quality data.

## The Quality Measurement Enterprise

Measurement is vital to producing a health care system that achieves outstanding results. Without measurement and transparency, clinicians, institutions, patients, and society cannot readily evaluate the value being achieved in the health care system. A commonly quoted aphorism that encourages the measurement movement

states, “You can’t improve what you don’t measure.”<sup>i</sup> The United States is about 25 years into efforts to bring performance measurement into medicine.<sup>7</sup> A seminal event in this history was the decision by the Department of Health and Human Services’ Centers for Medicare & Medicaid Services (CMS) in 1992 to pivot from having experts review medical records to identify substandard practice in a small number of outlier health care organizations to shift to using standardized quality measurement aimed at understanding whether standard practice across the health care system could be improved.<sup>8</sup> What was novel about this shift was the focus on explicit, objective criteria rather than implicit, subjective expert opinions, and an intention to shift the curve of “mean” performance toward improvement, rather than just focusing attention on the poor performance “tail” of the quality bell curve.

After first briefly trying to rate hospitals based on outcomes, CMS launched an effort to characterize the overall performance of the nation’s hospitals, starting with acute myocardial infarction (heart attack) in the Cooperative Cardiovascular Project (CCP). The CCP, which started as a pilot project in four states in the early 1990s and then as a national project a few years later, was the first effort to measure performance uniformly across the country. It was a remarkably ambitious project, requiring the abstraction of more than 200,000 medical records drawn from all the hospitals caring for Medicare patients. The CCP produced vital information that served as the foundation for what became remarkable improvements in cardiovascular care ([see appendix for more on the CCP](#)).<sup>9</sup>

Following CMS’ Cooperative Cardiovascular Project, the Institute of Medicine released two seminal reports—*To Err is Human* and *Crossing the Quality Chasm*<sup>10</sup>—and researcher Elizabeth McGlynn and colleagues published an influential article documenting

---

<sup>i</sup> To which others respond, citing a quote incorrectly, but deliciously, attributed to Albert Einstein, “Not everything that can be counted counts, and not everything that counts can be counted.” In fact, the quote appears to be from William Bruce Cameron’s 1963 book, *Informal Sociology: A Casual Introduction to Sociological Thinking*.

deficiencies in U.S. quality of care when assessed against specific, evidence-based metrics.<sup>11</sup> With awareness of health care quality deficiencies rising, organizations began to focus more on quality: how to define it, how to measure it, how to collect data on it, and how to use those measures to improve it.

While the popularity of performance measurement in health care has grown, its ubiquity is creating challenges for the field. Non-profit and for-profit organizations actively develop and promote measures and measurement systems that vary widely in their rigor and transparency. Some measures' specifications are in the public domain while others' are considered proprietary, with a lack of transparency about how the measures and performance ratings are derived. Some measures are publicly reported, while others are only used internally. Some measures can be used free of charge, while other measure developers require institutions to pay for the right to promote their performance results, and do not have transparent evaluation or an independent endorsement of their methods for determining performance.

**Organizations that support measurement.** A number of organizations develop and evaluate quality measures, and an even larger number of organizations collect measures for the purpose of evaluating and reporting on the performance of providers. Public measure developers include CMS and the Agency for Healthcare Research and Quality (AHRQ), and non-profit private developers include the Joint Commission and the National Committee for Quality Assurance (NCQA); all use a transparent approach to give the public an opportunity to review and comment on their draft measures, refuse to use proprietary measures, and make transparent their measure scoring mechanisms ([see appendix for more on these organizations](#)). Many professional societies also develop measures, such as the American Heart Association, the American College of Cardiology, the Society for Thoracic Surgeons, and the American College of Surgeons, although methods may vary across the organizations. Once developed, quality measures may undergo evaluation by the National Quality Forum (NQF), a public/private, multi-stakeholder organization that endorses general standards for measurement and specific measures themselves after a rigorous and transparent validation process.

Numerous for-profit companies, including Healthgrades and *U.S. News and World Report*, have developed their own measures and use them to grade hospitals and other health care providers. However,

most such information brokers use measures not endorsed by NQF, and do not always explicitly disclose the methods by which they rank hospitals. A number of researchers have questioned the validity and reliability of such proprietary “report cards.”<sup>12</sup> Understandably, in the absence of transparent measurement standards, the correlation among these various report cards is low. For example, recently none of the 17 top hospitals listed in *U.S. News and World Report*'s “Best Hospitals Honor Roll” were identified as top hospitals by the Joint Commission in their 2010 list of hospitals that received at least a 95 percent composite score on a suite of key quality measures.<sup>13</sup> Proprietary ranking systems likely confuse more than clarify. Findings such as this suggest that the measurement of quality in health care by these private for-profit companies is not aligned with measures in the public domain; it is usually impossible to determine if they are accurate.<sup>14</sup>

In addition to measures developed primarily for public reporting purposes, many measures are also developed for use internally by a practice or facility for quality improvement purposes. Such measures can be constructed quickly by merely running a query in an electronic health record (EHR), or can be more formally specified using more rigorous methods. When used for internal quality improvement purposes and not publically reported on websites, measures need not be held to the same standards as those that are intended to be publicly reported. For example, these measures may have a lower specificity, meaning they result in more “false positive” indications of quality problems. When measures are only used internally to screen for quality issues, false positives are not a concern, since the next step is usually merely to investigate further; such investigation can determine whether, for example, a clinician's suboptimal performance is a reflection on her actions or factors outside of the clinician's control. Also, data for internal use should require less precise risk adjustment and allow for greater timeliness.

While such homegrown measures might be appropriate for internal use, many are being reported on hospital websites and in marketing materials and used to make inferences about the magnitude of quality improvements they may have achieved over a period of time—often without sufficient information to determine their methodology or accuracy. For example, one hospital advertised that it had no infections, without indicating which ones or for how long. Another reported that its quality improvement efforts had saved hundreds of lives, without discussing how the improvements or the saved lives were measured.<sup>15</sup> In short, the public may understandably be confused by

the array of measures that are now promoted in different places.

Despite the broad demand for performance measures and the recognized limitations of current measures, the United States lacks an organization charged with advancing the science of performance measurement, developing standards for performance measures, setting parameters for how accurate the measures must be before they are used in pay-for-performance or public reporting initiatives, and coordinating the development of the large number of measures required to inform patient choice and monitor performance—so that different entities don't develop duplicative yet different measures on the same topic. The closest thing we have to such an entity is NQF, which plays an important role by developing consensus standards for measures and validating measures submitted to it. However, given its mandate, NQF has a limited ability to support the development and pilot-testing of new measures itself or to attest to the accuracy of published measures that are not submitted to the NQF process.

## Measuring Structures, Processes, and Outcomes

Avedis Donabedian, an influential leader in the study of health care quality, developed a widely used, three-element model of quality measurement in 1966, which included measuring health care *structures* (the characteristics associated with a health care setting), *processes* (the activities done in a health care setting), and *outcomes* (the results achieved for a patient after a given set of interventions).<sup>16</sup>

*Structural measures* include requirements imposed by payers and regulators, such as specifications for the physical plant, management systems, board certification, and staffing ratios.

*Process measures* determine whether evidence-based care guidelines were followed, but do not indicate whether a patient's health actually improved. Process measures, in essence, are used on the assumption that better outcomes should result from evidence-based care processes. Examples of process measures include the rate at which patients experiencing a heart attack are administered aspirin and beta-blockers.

*Outcome measures* seek to determine whether the desired results are achieved. Examples of clinical outcome measures are whether a patient was readmitted to the hospital within 30 days of discharge and, for some conditions, whether the patient is alive at 30 days after admission.

So-called “intermediate” or “surrogate” outcome measures are those that, while not true outcomes, are assumed to be able to be used as proxies for patient outcomes. For example, hemoglobin A1C blood test results are used both in research and practice as an indicator of whether diabetes is under control, because the results of the test correlate with the likelihood of experiencing diabetes complications. Measuring hemoglobin A1C on a periodic basis is a *process measure*, whereas achieving desirable hemoglobin A1C blood levels is sometimes labeled an *intermediate outcome measure*.

Increasingly, quality experts also include various aspects of patients' experiences as important outcome measures. Examples of patient experience instruments include the Patient Reported Outcomes Measures Information System, which includes modules that address physical health, mental health, and social health; HealthActCHQ, which has developed pediatric quality of life questionnaires, among others; and the Consumer Assessment of Healthcare Providers and Systems (CAHPS) surveys developed under the auspices of AHRQ.

## Data Sources Used to Calculate Quality Measures

In general, the data needed for determining performance with established measures are obtained through three sources: administrative data, medical records, and patient surveys.

Administrative data are derived mostly from insurance claims and enrollment files. Such data are relatively easy and inexpensive to collect but lack the clinical detail needed to generate many desired measures. Reliance on administrative claims data therefore limits what and how accurately performance can be measured. Determining whether particular services were unnecessarily performed generally requires clinical detail to determine the appropriateness of the service in a particular patient's clinical circumstances—information that is not available from claims forms. For example, without knowing the patient's clinical history, current symptoms, and the results of images of her coronary arteries, it is impossible to determine whether a procedure involving inserting coronary artery stents into partially blocked arteries is appropriate. However, in some cases, the output from measures that use administrative claims have shown a high correlation with output from actual clinical data<sup>17</sup>—although administrative data can vary

substantially in accuracy compared with medical records, which are also far from perfect.

Medical record data provide substantially more detail about the care being provided, the patient's history, condition, and complications, but are substantially less standardized and in many cases less practical. They are also more expensive to use for quality measurement purposes, since they require expert staff to abstract and interpret them to determine if a particular care process was conducted or not. The quality of the data may also be variable, particularly across different sites, practices, or organizations, which has implications for profiling and benchmarking. The widespread adoption of EHRs should make medical record data collection substantially cheaper and easier in the future, although it will not fill all clinical data gaps and may not address problems with data quality. Moreover, it can be difficult to extract data from paper records or EHRs, as there are few common standards for documentation and many terms vary in their meaning. As examples, site-to-site variations in the use of terms (such as “shock”) or in the listing of contraindications to clinical strategies can lead to substantial bias in the assessment of performance.

Survey data are typically collected for the purpose of measuring patient experience with care. In the United States, the CAHPS survey is the most well-known of these surveys, and can be fielded among samples of patients by mail, phone, or email. The survey was developed by AHRQ, has been endorsed by NQF, is publicly reported by many health insurance plans, and is widely used by a range of organizations, including NCQA, which requires plans to field the survey to obtain certain types of plan certification, and also certifies survey vendors that organizations can hire to field the CAHPS survey for them. Unfortunately, survey data are expensive to obtain, and their interpretation as a quality measure can be compromised due to site variation in response rates, which can bias results. Initially developed to assess health plan enrollees' experience with their care, there are now CAHPS versions that focus on particular types of providers, including hospitals, doctor's offices, and dialysis facilities, and particular topics, such as the extent to which a provider is using health IT tools or delivering care in accordance with the patient-centered medical home model of care.<sup>18</sup> Efforts are underway to incorporate patient experience measures into public reporting of quality.<sup>19</sup> For example, data collected using the hospital version of the CAHPS survey are now publicly available for all U.S. hospitals on the CMS website.

To generate more robust quality measures, “hybrid” data collection is sometimes required, which refers to the combination of administrative data with information obtained from medical records or patient experience surveys. Such approaches can increase the number of data elements used to generate measure data, reduce the amount of data that must be extracted from medical records, or both.

## Primary Uses of Performance Measure Data

In the United States, performance measure data are predominantly used in public reporting and provider incentive programs as well as provider-led quality improvement efforts.

### *Public Reporting*

Measuring and reporting on the quality and cost of care serves several important functions, including: (1) enabling patients to make informed choices about their care and be more involved in medical decision-making; (2) allowing health care professionals to identify areas for improvement and providing them with the motivation to do so; and (3) providing consumers, purchasers, and taxpayers some level of accountability for their substantial expenditures on health care.<sup>20</sup>

While ample evidence exists to demonstrate how publicly reporting the performance of health care providers can spur quality improvements,<sup>21</sup> there is mixed evidence about how well public reporting informs consumer choice.<sup>22</sup> Public reports seem to have negligible impacts on the selection of providers by patients and families or their representatives, primarily because patients are often not aware that the quality information is available, the information provided in public reports is not what they need or value, the information is outdated, the information is not always available when they need it to make a decision, or the information is not presented in an easily understandable way.<sup>23</sup>

Commercial health plans often publicly report provider performance, and sometimes also combine quality measurement data with price and cost information to attempt to categorize providers, especially hospitals, into different value tiers, such that plan members face lower cost-sharing when selecting providers in favored tiers.<sup>24</sup> One of the most ambitious applications of performance measurement is in California, where the Integrated Healthcare Association collaborates with health plans and more than 200 medical groups and independent practice associations to maintain public

reporting and a pay-for-performance program using the Healthcare Effectiveness Data and Information Set (HEDIS), patient experience and satisfaction survey data, and data documenting the adoption and use of health information technology by practitioners.<sup>25</sup> In this context, performance measure data are provided to the public to help reassure them that quality is maintained even though these physicians—who are mainly paid capitated rates per patient by health maintenance organizations (HMOs)—have financial incentives that could result in stinting on care.

Medicare has also been a major producer and user of performance measure data, initially for the purpose of providing information to consumers to help them select providers and health plans. Medicare has used its own administrative datasets and has made extensive use of patient surveys on experience of care.<sup>ii</sup> The [Medicare.gov](http://www.Medicare.gov) website now provides comparative performance information for hospitals, nursing homes, home health agencies, dialysis facilities, Medicare Advantage health plans, and drug plans.<sup>26</sup>

### *Pay-for-performance*

Apart from promoting more informed consumer choice, CMS also uses performance measurement data in a number of its pay-for-performance initiatives, which provide direct financial rewards or penalties to health care providers based on their performance on quality measures. These initiatives include a suite of new “value-based purchasing” programs (Congress’s term for pay-for-performance) to reward providers who deliver better performance for beneficiaries at lower cost.<sup>27</sup> Some of these programs include the End-Stage Renal Disease (ESRD) Bundled-Payment and Quality Incentive Program, performance bonuses for Medicare Advantage (MA) plans based on star ratings, the Hospital Value-based Purchasing Program, and the Physician Value-based Payment Modifier.

One of the apparent success stories in the application of measures can be found in Medicare’s ESRD Quality Incentive Program; within two years of beginning this program, the majority of dialysis facilities showed significant improvement on the program’s three clinical

process measures.<sup>28</sup> Facilitating the success was the fact that the measures used to assess dialysis have been shown to be excellent intermediary outcome measures that reliably predict ESRD patient outcomes.<sup>29</sup>

Perhaps the most prominent use of pay-for-performance in Medicare results from the Affordable Care Act’s (ACA) new approach to paying quality bonuses to MA plans colloquially referred to as the “Medicare 5 star program.” For several years, CMS has posted quality ratings of MA plans online, using a 1 to 5 star scale, to provide beneficiaries additional information to inform their choice of plans. Under the ACA, Medicare now also pays plans differentially based on these star ratings and may limit enrollment in poorly-performing plans.

These quality scores are based on performance measures derived from CMS administrative data, HEDIS measure data provided by plans, and survey data collected directly from beneficiaries using AHRQ’s CAHPS survey and CMS’ Health of Seniors survey. A recent analysis found a positive association between beneficiary enrollment decisions and the star ratings, suggesting that the performance measures are an important factor in making health plan choices.<sup>30</sup>

Two important issues with the ratings relate to the limits of the measures and the regional variation associated with high-performing plans. First, while the star scale methodology culls from a reasonably broad set of measures, there are gaps in important areas of health plan performance, such as the health plan’s performance related to patients with acute, serious health care problems (which are obviously common in the Medicare population). For example, none of the measures relate to whether patients are informed about the advisability of referral outside of the MA plan’s provider network for patients with unique clinical circumstances, such as particular cancers best cared for in a specialized cancer center.

A further problem is the skewed geographic distribution of performance. More than half of enrollees in Massachusetts, Oregon, Washington, and Minnesota were in plans with four or five quality stars, whereas in 19 states fewer than two percent of enrollees were in this top tier,<sup>31</sup> implying that health plan quality performance mostly reflects the performance of the local providers who make up the health plan’s network. In short, while health plans generally have responded positively to improve their star ratings—for reputation and financial rewards—Medicare beneficiaries are likely getting only a partial picture of the value-added provided by any particular health plan.

---

<sup>ii</sup> Many of the measures that use administrative claims data have not been validated using measures based on medical record information, with the exception of mortality and readmission measures.

The CMS pay-for-performance approach that is now receiving physician attention is the new Value-Based Payment Modifier, established in the ACA. The assignment of physician value will be used to adjust Medicare payments to physicians based on measured performance on quality and cost starting in 2015. For the numerator of the value-based modifier calculation, CMS will use the measures in its Physician Quality Reporting System (PQRS),<sup>iii</sup> and is working on additional measures to assess their costs—which will make up the denominator in the value formula.

The challenges with producing a composite measure of physician value and implementing pay-for-performance for individual physicians are formidable. For example, family physicians, general practitioners, and internists treat nearly 400 different diagnostic categories, with about 70 categories making up 80 percent of their clinical episodes in a year.<sup>32</sup> Basing a payment modifier on performance on only a few PQRS quality measures will therefore not provide a meaningful assessment of the quality of a clinician's care. Further, the core of what some specialties do presents substantial measurement challenges—for example, we currently do not measure whether a physician made a correct diagnosis. The issues of assigning a cost measure to physicians are similarly difficult, in this case because of the problems of attributing costs generated by many clinicians and institutional providers to a single physician. Using such an approach to determine a physician's value, many physicians will likely be incorrectly assessed, with likely harmful effects on physicians' reputations and the measurement enterprise more broadly – and patients may be misled in choosing physicians.

---

<sup>iii</sup> Initiated as a voluntary program in 2007, PQRS provides incentive payments to eligible physicians and other practitioners who report quality data. CMS provided a 1 percent incentive payment in 2011, and will provide a 0.5 percent incentive payment in 2012 through 2014, for successfully reporting at least three measures that apply to the services furnished by that professional from a list of more than 200 measures that apply to all specialties. Penalties of up to 1 percent will begin in 2015 for those who do not satisfactorily submit quality data. Fewer than 30 percent of practices currently submit data under the PQRS program (See: Iglehart and Baron, 2012). For medical groups of more than 200 physicians, all 26 of the current NQF-endorsed quality measures for coronary artery disease, diabetes, heart failure, and preventive care services must be reported.

The physician value-based modifier is one of numerous pay-for-performance programs that Congress has mandated. One of the most prominent programs—Hospital Value-based Purchasing—is being launched despite the fact that the demonstration that informed the design of that program—the Premier Hospital Quality Incentive Demonstration Project—did not actually produce better results than comparison hospitals, which also demonstrated improved scores on what were mostly process measures. Indeed, two evaluations found little evidence that the demonstration's use of financial incentives to incentivize improved performance led to reduced mortality rates beyond those achieved with public reporting alone.<sup>33</sup> Various other studies of pay-for-performance for hospitals and physicians have produced mixed results, at best showing small, sometimes temporary, improvements in quality. ([See appendix for more details on the evidence base for pay-for-performance.](#)) Further, a few studies have questioned the common reliance on process measures to improve quality for hospital and physician care, although it seems likely that the details—such as the strength of the incentives, the number and selection of performance measures being used, the complexity of the care processes being improved, and restrictions on how bonuses can be used—may affect the success of pay-for-performance programs.<sup>34</sup> The message may be that we have not yet determined how such incentives can be most effectively applied, the extent to which they motivate hospital managers versus physicians, or even if they are sustainably effective in any form over the long run.

Although for some, pay-for-performance is a commonsense approach that would surely work to improve performance if the incentives are large enough,<sup>35</sup> in fact, there are both empirical and theoretical reasons why this approach might actually backfire. The approach has been used in other sectors of the economy without success,<sup>36</sup> perhaps the most visible being in education where the approach is being subjected to increasing criticism.<sup>37</sup>

Under principal-agent theory, the principal (in this case, the payer) offers the agent (a physician, hospital, or accountable care organization) incentives to make maximal effort to act in the principal's interests (i.e., to provide high quality to patients). But, according to one expert's interpretation of the theory, if an agent is expected to devote time and effort to some activity that cannot be measured, then incentive pay cannot be used effectively to encourage activities that *can* be measured.<sup>38</sup> Because most of what physicians do in caring for patients is not measured—and mostly cannot

feasibly be measured—rewarding a limited number of activities might lead to less effort—and reduced quality—in these other unmeasurable areas.<sup>39</sup> However, so far, the limited literature that finds a lack of positive impact on measured quality has not found stinting on other important areas of quality.

Behavioral economics offers some insights into why, despite intuitive appeal, pay-for-performance may have had a limited impact on improving quality of care. At root, economic incentives seek to change behavior through *extrinsic* motivation—yet most clinicians want the best outcomes for their patients based on an *intrinsic* motivation to act in their patient’s best interests. Some of the nation’s most effective quality improvement campaigns – such as those aimed at reducing central line infections and “door-to-balloon” times for heart attack patients requiring surgery to open up occluded arteries – were wholly based on intrinsic motivation combined with effective new strategies, without financial incentives.<sup>40</sup>

Further, there is evidence outside of health care that money may not be a solution—and in fact, it may backfire—particularly for cognitively challenging activities<sup>41</sup> performed by highly skilled persons needing to muster their skills to manage complexity and solve problems creatively.<sup>42</sup> While financial incentives are effective at changing behavior when the pathway from the incentive to the desired behavior is short and direct, the pathway from incentive to improving quality is long and indirect—and often times unknown. Experimental data demonstrate that financial incentives often crowd out intrinsic motivation.<sup>43</sup> In particular, tangible rewards, especially monetary ones, undermine motivation for tasks that are intrinsically interesting or rewarding and have their strongest negative impact when the external rewards are perceived as large, controlling, contingent on very specific task performance, or associated with surveillance, deadlines, or threats. In short, if intrinsic motivation is high and crowding out is strong, payment incentives may worsen performance.<sup>44</sup> It will be important to learn whether organizations like hospitals respond differently to payment incentives than professionals.

## Issues That Arise From Reliance on Structure and Process Measures

Structural measures, as described earlier, can include metrics such as the volume of a certain type of operation performed by a hospital. Such indicators can sometimes be a predictor of outcomes; for example, there is a literature that shows that for some procedures,

institutions that do more procedures achieve better health outcomes,<sup>45</sup> but the relationship between volume and outcomes is variable—by procedure and provider. Some quality ranking systems, like *U.S. News and World Report’s* rankings of hospitals, rely at least in part on structural criteria, such as nurse-to-patient bed ratios and availability of new technology. More commonly, structural criteria are included as survey questions to accredit or certify that a provider meets threshold standards to be included as a recipient of program funds. For example, “conditions of participation” establish the structural quality and safety standards that all U.S. hospitals must follow to participate in Medicare and Medicaid.

Meanwhile, process measures—which are the most common type of quality measures—calculate the rate at which a recommended clinical or care process is performed. By one estimate, of 78 HEDIS measures for 2010, all but five were clearly process measures, and none were true outcomes measures.<sup>46</sup> Process measures have several theoretical advantages over outcome measures.

First, calculating process measures is more straightforward because in some cases (for example, when evaluating physician prescribing) there may be less need for risk-adjustment to account for case mix differences that clearly affect outcomes. Yet, for process measures that evaluate patient adherence to treatment recommendations, for example, there may be a need for risk adjustment to account for relevant socioeconomic differences.<sup>47</sup>

Second, process measures typically reflect professional standards of care. As a result, they are most often subject to evidence-based, professional standard setting that is readily understood by clinicians. In contrast, the factors that often contribute to different outcomes across institutions include organizational culture, leadership, teamwork, technology, and other factors not part of professional standards of care that clinicians and other individuals can readily control.<sup>48</sup> Thus, process measures are “actionable”—that is, the measure itself prescribes the action that the clinician, institution, or health plan needs to take to improve performance.<sup>49</sup> Feedback to clinicians is more personally relevant and thus easier to act on.

Finally, practically, there is often a large research base that provides evidence on which processes reliably improve particular outcomes, although the studies do not always cover all populations of interest, especially the elderly.



Despite the theoretical advantages of process measures, reliance on them to assess quality presents several problems.

***There are major gaps in what process measures can measure.*** Currently, quality of care in the outpatient setting has become synonymous with preventive care and chronic disease management, with some growing interest in patient experience – virtually ignoring the very important quality issues of safety, effectiveness, coordination, and efficiency.<sup>50</sup> The result is that the available process measures may give a very misleading picture of quality for clinicians and organizations.<sup>51</sup> There are important clinical areas for which measures are lacking and are therefore, arguably, not being given the attention they deserve. For example, we have few measures to assess:

- diagnosis errors (which are alarmingly common and outnumber surgical errors as the leading cause of outpatient malpractice claims and settlements);<sup>52</sup>
- the appropriateness of diagnostic and therapeutic interventions;<sup>53</sup> and
- providers' ability to skillfully manage complex patients with varying combinations of multiple clinical and psychosocial problems.<sup>54</sup>

Furthermore, many of these gaps are not likely to be filled, given the limited types of data currently available from administrative claims and clinical records.

***Process measures do not always predict outcomes.*** Recent research suggests that even the longstanding and broadly accepted CMS process measures for heart failure, heart attack, and pneumonia did not predict overall short-term mortality in the Premier demonstration. Similarly, currently available information on CMS Hospital Compare website shows that the process measures used to assess surgical performance did not help patients identify hospitals with better outcomes for high-risk surgery.<sup>55</sup> This finding is consistent with other studies for non-surgical conditions.<sup>56</sup>

There are several possible reasons for the lack of relationship between process measures and short-term outcomes at the hospital level:

- Process measures tend to reflect quality for narrow actions for a small subset of patients with particular conditions. For example, most of the CMS process measures for heart attack apply to fewer than half of the patients admitted to the hospital with this

condition—and some apply to only 10 percent or fewer.<sup>57,iv</sup>

- Some process measures are only expected to provide a benefit over a long time horizon, so differences in early mortality would not be expected. For example, the use of beta-blockers after discharge for patients who survive a heart attack would not be expected to have a large effect in the subsequent 30 days, even though a benefit could become apparent for individual patients over the following year.
- Some conveniently available process measures, such as smoking cessation education, were never associated with reduced mortality, and so would not be expected to reduce mortality.
- Measurement error can weaken the association between a measured process and an outcome if the way a process is measured differs from how it was implemented in the original research.
- Process measures used for some conditions, such as treating a heart attack, typically do not capture overuse or inappropriate use of medications. An institution could appear to be performing highly on a measure even if they were indiscriminately administering a medication to patients for whom the drug is contraindicated.
- Process measures may not directly measure the effectiveness and appropriateness of actual care, even as it gives credit for performing a particular action. A measure might give credit for providing smoking cessation advice, no matter how perfunctory. A hospital might receive credit for administering a recommended medication, even if the wrong dose is administered, or used in a patient at risk for an adverse drug interaction.<sup>58,v</sup>

---

<sup>iv</sup> This means sample size could be hampering statistical analyses of the relationship between these process measures and outcomes. If compliance with the measure improves outcomes for a small portion of the patients, then the overall effect may be hard to detect.

<sup>v</sup> To illustrate, one commonly used performance measure is whether patients receive antibiotics immediately prior to surgery. In evaluating an initiative at Johns Hopkins Hospital that used this measure, evaluators found that 30 percent of patients did not receive the correct dose, mainly because overweight patients needed a higher dose, yet the hospital would have received credit on the performance measure regardless. In this case, the lack of association that was

**Teaching to the test and diverting resources.** A major concern with reliance on process measures rather than outcomes is that the hospital or medical practice being assessed could be diverting resources from other areas to ensure the requisite performance on the process measures, meanwhile ignoring problems in areas of care not being assessed, which also contribute importantly to hospitals' varying outcomes.<sup>59</sup> Borrowing from the critique of performance measures in education, this is commonly referred to as "teaching to the test."<sup>vi</sup>

**Practical Problems.** A practical limitation in using process measures relates to the high cost of data collection; that limitation produces a heavy reliance on laboratory tests and prescription drugs, which limits the care processes that can be measured. For example, the key work process improvements that reduce central line-associated blood stream infections (CLABSI) relate to a checklist of recommended activities, such as hand washing. However, experience has shown that self-reported compliance grossly overestimates performance. To obtain a valid measure would require having an anonymous observer actually watch central line placements. However, because these catheters are inserted at random times throughout the day, this type of data collection would be exceedingly expensive. In this case, fortuitously a valid outcome measure (the CLABSI rate) was ultimately developed to substitute for prior reliance on unreliable and costly collection of process measures. Whereas an individual hospital with high infection rates may want to collect process measures on compliance with the checklist periodically as part of its internal quality improvement effort, broadly collecting these process measures for public dissemination would be neither valid nor feasible.

Finally, updating process measures based on emerging evidence is often difficult and resource-intensive, resulting in the use of measures that may no longer meet recommended standards for process measures. In general, good measures should: have a strong evidence

---

ultimately found between performance on this process measure and the outcome of interest—reducing surgical site infections—may have resulted from the wrong thing being measured.

<sup>vi</sup> Most process measures also currently require manual data collection, which is typically performed by quality improvement staff, diverting their ability to participate in other efforts to improve care.

*“A major concern with reliance on process measures rather than outcomes is that the hospital or medical practice being assessed could be diverting resources from other areas to ensure the requisite performance on the process measures, meanwhile ignoring problems in areas of care not being assessed.”*

base showing that the measured care process leads to improved outcomes; capture whether the measured care process has, in fact, been provided with accuracy; address a process that has few intervening care activities that must also occur to achieve the desired outcome; and have little or no chance of inducing adverse consequences by their use.<sup>60</sup> Many process measures continue in wide use despite failing one or more of these criteria.

## The Challenges of Outcome Measures

Given these reasons to avoid an overreliance on structure and process measures, there is growing interest in measuring outcomes of care.<sup>61</sup> Patients are interested in surviving a serious illness and regaining optimal functioning, avoiding hospital admissions, having positive experiences, and minimizing symptoms—not the clinical processes providers use to achieve those desired outcomes. No set of process measures—even if they were accurate and important predictors of outcomes—can be comprehensive enough to serve as a substitute for actual outcomes.<sup>62</sup> When coupled with cost data, outcome measure data can also present patients with a useful measure of value as they choose providers.

Outcome measures are also attractive because there is growing recognition that hospitals can impact patient outcomes through factors beyond care processes—such as teamwork, leadership, and culture.<sup>63</sup> Moreover, the ever growing number of process measures—some of which are collected by manual medical record extraction—place an increasing administrative burden on providers, often for limited return in patient outcomes.

While many now call for migrating from measuring processes to outcomes, accomplishing the transition has proven devilishly difficult. Simply put, accurately measuring patient outcomes, while conceptually appealing, is very difficult to accomplish. Some of the key challenges associated with measuring outcomes are described below.

**Risk adjustment.** One reason measuring outcomes is challenging is because an individual patient's outcome is not simply the result of the effectiveness of medical care, but is also impacted by a patient's risk factors (i.e., how sick they are before receiving care, and how severe their current illness is) as well as chance events.<sup>64</sup> Social determinants of health may also be important, and it is unclear to what extent providers should be held accountable for outcome differences associated with such factors. To avoid penalizing hospitals and physicians who treat higher-risk patients, measuring outcomes requires using rigorous risk adjustment models to account for variation in patient characteristics and severity of illness that may importantly affect outcomes.<sup>65</sup> Unfortunately, while risk adjustment techniques have advanced in recent years, there is no standardized approach to adjusting outcomes for patient risk—different risk adjustment approaches make different operational decisions with different consequences for the measured performance on outcomes.<sup>66</sup> In an effort to improve risk adjustment approaches, the Council of the Presidents of Statistical Societies recently produced a consensus document recommending that CMS augment the patient-level attributes it uses to risk adjust data with the addition of race or other demographic variables.<sup>67</sup>

Risk adjustment models generally perform better when the patient population is narrow and well specified, such as patients having a specific type of surgery. Risk adjustment models for diverse patient populations, such as all hospitalized patients, perform less well and can often lead to inaccurate inferences.<sup>68</sup> As such, measures of overall hospital mortality generally should be avoided or used cautiously,<sup>69</sup> although a recently developed hospital-wide measure of all-condition readmission rates appears to perform well.<sup>70</sup>

**Data validity.** Other challenges associated with measuring outcomes include concerns about the validity of outcome measures—meaning, whether a measure correctly assesses the concept being measured. A measure can lack validity if it inappropriately excludes certain information, does not appropriately adjust or stratify the baseline risk of measured patients, uses multiple and inconsistent data sources or methods, uses incorrect data, or does not correctly capture the concept of quality that it is intended to measure.<sup>71</sup> Claims data—which are often used to calculate performance measures, due to their low cost—can introduce validity concerns, since they fail to identify preexisting conditions and complications that occur after hospital admission, making an accurate

assessment of baseline patient risk factors problematic.<sup>72</sup>

Public reporting can also introduce validity concerns about the accuracy of the data being used to calculate performance measures. When CMS stopped paying the costs of selected preventable adverse events under diagnosis related groups, there was a marked drop off in reporting of now unpaid complications from central line infections. Yet, a study based on clinical lab data finds no evidence that the nonpayment policy affected the true infection rate.<sup>73</sup> In general, measures of hospital infections and other complications calculated using administrative data correlate poorly with those calculated using medical record review<sup>74</sup> and other sources. Yet, medical records are far from a gold standard with respect to the patient's information. In short, there are considerable challenges in profiling institutions based on such source data.

**Surveillance bias.** Another factor in measuring certain outcomes is surveillance bias—the idea that more closely monitoring something can lead to higher rates of detecting something of interest—which can cause significant errors.<sup>75</sup> For example, one hospital found that their rate of deep venous thrombosis (DVT) increased ten-fold when doctors started looking harder for patients with this condition through greater use of routine ultrasounds.<sup>76</sup> As a result, the hospital went from having one of the lowest rate of DVTs to one of the highest, putting the hospital at financial and reputational risk and demoralizing the physicians who felt they were providing better care though the enhanced surveillance. The problem of surveillance bias has also been observed when attempting to measure rates of medical errors—where more conscientious programs to reduce errors lead to higher rates of detection and apparently worse performance.

**Sample sizes.** Another issue in measuring outcomes is that large samples are often needed to provide measures with acceptable random error. Many adverse outcomes are rare, and as such, measures of outcomes over a short period of time may have too few events to provide a stable measure. This challenge is especially acute in small hospitals that may have a low volume of specific procedures. One approach to addressing this problem is to consider cumulative performance over several years, rather than an annual measure, to increase the number of patients included in a measure's denominator. The downside, of course, is that accumulating data over years will compromise the objective of real-time appraisal of performance and make it more difficult to detect changes. Personnel may

have changed or new care processes adopted during the time period of the extended performance period, thereby compromising the accuracy of the measure. Another approach would be to aggregate small practices into larger groups, sometimes called “pods,” for statistical purposes. That approach increases statistical sensitivity but at the cost of specificity, as the group’s performance may not well reflect that of an individual practice.<sup>vii</sup>

\* \* \*

Increasingly, the problems with process measures are being acknowledged. CMS has indicated that it recognizes it needs to strengthen its portfolio of hospital measures, especially outcome measures, such as by emphasizing measures of 30-day mortality, hospital-acquired infections, cost, and patients’ experiences with care. And while there is growing interest in relying on outcome measures, since they better reflect what patients and providers are interested in, establishing valid outcome measures pose their own substantial challenges—including the need to risk-adjust results to account for patients’ baseline health status and risk factors, assure data validity, recognize surveillance bias, and use sufficiently large sample sizes to permit correct inferences about performance.

### Policy Recommendations

It should be clear by now that measuring the quality of health care, while worthwhile and potentially even transformative, is technically difficult and prone to error. Given this background and the important role that performance measurement can play in health care, we make several policy recommendations to advance the field.

1. Decisively move from measuring processes to outcomes;
2. Use quality measures strategically, adopting other quality improvement approaches where measures fall short;
3. Measure quality at the level of the organization, rather than the clinician;
4. Measure patient experience with care and patient-reported outcomes as ends in themselves;
5. Use measurement to promote the concept of the rapid-learning health care system;
6. Invest in the “basic science” of measurement development and applications, including an emphasis on anticipating and preventing unintended adverse consequences; and
7. Task a single entity with defining standards for measuring and reporting quality and cost data, similar to the role the SEC serves for the reporting of corporate financial data, to improve the validity and comparability of publicly-reported quality data.

#### 1. Decisively move from measuring processes to outcomes.

The operational challenges of moving to producing accurate and reliable outcome measures are daunting but worth the commitment. Patients, payers, policy-makers, and providers all care about the end results of care—not the technical approaches that providers may adopt to achieve desired outcomes, and may well vary across different organizations. Public reporting and rewards for outcomes rather than processes of care should cause provider organizations to engage in broader approaches to quality improvement activities, ideally relying on rapid-learning through root cause analysis and teamwork rather than taking on a few conveniently available process measures that are actionable but often explain little of the variation in outcomes that exemplifies U.S. health care.

However, given the inherent limitations of administrative data, which are used primarily for payment purposes, and even clinical information in EHRs, consideration should be given to developing a national, standardized system for outcome reporting.<sup>77</sup> A new outcome reporting system would not be simple or inexpensive, but current data systems may simply be insufficient to support accurate reporting of outcomes. An example is the National Health Care Safety Network system for reporting health care infections.<sup>78</sup>

Alternatively, EHR vendors could modify their products to allow them to be used to calculate validated quality measures. By standardizing which structured data elements they include in their products and the

---

<sup>vii</sup> An approach specific to patient safety outcomes is to aggregate multiple types of adverse events into a global measure of patient safety. Yet this approach, which relies on “triggers” (clues in the medical record that may indicate that an adverse event occurred) lacks sufficient rigor to measure rates of outcomes and to make inferences about quality and may lead to biased results. See: Mattsson TO, Knudsen JL, Lauritsen J, et al. “Assessment of the global trigger tool to measure, monitor and evaluate patient safety in cancer patients: reliability concerns are raised.” *BMJ Quality & Safety*, doi:10.1136/bmjqs-2012-001219, 2013.

metadata they use to describe these fields, vendors could allow for the calculation of validated quality measures, such as those collected by National Surgical Quality Improvement Program and the Society of Thoracic Surgeons. Once collected, clinical data would need to be evaluated for validity and quality. Prioritizing which measures require highly valid data and which do not may also help. It may be that for rare events, less accurate, although substantially less costly, administrative data would suffice, while for more common events and conditions, it would be more cost-effective to collect clinical data from clinical records. However, the quality of EHR data is also being questioned.<sup>79</sup>

An emphasis on measurement of outcomes, rather than care processes, need not ignore the contribution of specific processes that are associated with achieving better outcomes. In fact, achieving high reliability on process measures could be viewed as an internal tactic that providers might adopt as part of a comprehensive approach to achieve good outcomes, rather than as an end in itself.<sup>80</sup> Professional societies or governmental agencies could maintain a library of process measures that providers could select from to audit their own performance. But here the distinction between measures for quality improvement and for public reporting becomes important: publicly reported measures could emphasize the outcomes of interest, while measures used internally for quality improvement could emphasize the care processes that an organization is working on performing better.

A relatively small number of process measures, especially if linked with intermediate outcome measures, could serve as excellent measures for public reporting, mitigating the risks for surveillance bias, although the public would need to be educated about their clinical implications. Process measures (e.g., obtaining hemoglobin A1C levels in diabetics and properly taken blood pressure readings) could be linked to intermediate outcome measures (e.g., hemoglobin A1C level and blood pressure). The use of such measures in public reporting efforts could also educate patients and consumers about these important parameters of clinical care. However, caution should be used in using intermediary outcome measures, as demonstrated by the recent experience in which intensive treatment of patients to lower their hemoglobin A1C was recently shown not to be associated with the favorable outcomes expected. NCQA and others developed process measures favoring achievement of hemoglobin A1C levels below 7 percent. Yet, it was precisely this level that failed to

show improved outcomes in three recent randomized trials, ultimately leading to the abandonment of that process measure by NCQA.

In some clinical areas, process measures that assess the rate at which specific harmful medical errors occur also hold appeal. For harms that are almost entirely preventable—some of which are referred to as “never events”—risk adjustment and other statistical concerns should be unimportant.

A promising avenue for supporting a movement toward reliance on outcomes is greater use of patient-reported outcomes, which are derived using tools that measure what patients are able to do and how they feel through surveys. A wide variety of patient-level instruments to measure patient-reported outcomes related to physical, mental, and social well-being have been used in clinical research, such as within the National Institutes of Health’s Patient-Reported Outcomes Measurement Information System. Extending this research application for purposes of accountability and performance improvement will require additional work to address methodological and data challenges.<sup>81</sup>

## **2. Use quality measures strategically, adopting other quality improvement approaches where measures fall short.**

While working to develop a broad set of outcome measures that can be the basis for attaining the goals of public accountability and information for consumer choice, Medicare should ensure that the use of performance measures supports quality improvement efforts to address important deficiencies in how care is provided, not only to Medicare beneficiaries but to all Americans.

CMS’ current focus on reducing preventable rehospitalizations within 30 days of discharge represents a timely, strategic use of performance measurement to address an evident problem where there are demonstrated approaches to achieve successful improvement.<sup>82</sup> Physicians and hospital clinical staff, if not necessarily hospital financial officers, generally have responded quite positively to the challenge of reducing preventable readmissions. CMS has complemented the statutory mandate to provide financial incentives to hospitals to reduce readmission rates by developing new service codes in the Medicare physician fee schedule that provide payment to community physicians to support their enhanced role in assuring better patient transitions out of the hospital in order to reduce the likelihood of readmission.<sup>83</sup>

CMS recently announced that after hovering between 18.5 percent and 19.5 percent for the past five years, the 30-day all-cause readmission rate for Medicare beneficiaries dropped to 17.8 percent in the final quarter of 2012,<sup>84</sup> implying some early success with efforts to use performance measures as part of a broad quality improvement approach to improve a discrete and important quality and cost problem. However, this approach is not without controversy. Improvements have been modest, and some suggest that readmission rates are often outside the hospital's control, so CMS' new policy unfairly penalizes hospitals that treat patients who are the sickest.<sup>85</sup> And while readmission in surgical patients is largely related to preventable complications, readmissions in medical patients can be related to socioeconomic status. Also, some have even questioned the accuracy of CMS' seemingly straightforward readmission rate measure, finding that some hospitals reduce both admissions and readmissions—a desirable result—yet do not impact the readmission *rate* calculation.<sup>86</sup> And one of this paper's authors (R. Berenson) has suggested a very different payment model that would reward hospital improvement rather than absolute performance, thereby addressing the reality that hospitals' abilities to influence readmission rates do vary by factors outside of their control.<sup>87</sup>

We consider the current controversy around implementation of a readmissions penalty to be a healthy debate. Because the purpose for which the penalty was designed is so important, scrutiny and vigorous discussion can lead to improvements to CMS' payment policy and performance measures to address what remains an unacceptable failure in U.S. health care delivery. There clearly is a tension between getting the measures absolutely right and achieving a “good enough” status that can produce quality improvement. In the words of Jonathan Blum, deputy administrator and director for the Center of Medicare at CMS, “It's a very traumatic event to go back to the hospital. I'm personally comfortable with some imprecision to our measures.”<sup>88</sup>

With the growing evidence that Congress's value-based purchasing approach to measuring and rewarding hospitals only marginally improves patient outcomes, and possibly diverts attention from doing the hard work of making culture and work process improvements that actually would produce improved outcomes, Congress should refocus its directives to CMS to emphasize improving specific quality deficiencies—relying more on promoting collaborative quality improvement activities and new payment approaches that incorporate

performance measures than on public reporting and pay-for-performance *per se*. As an illustration, the nuclear industry has a robust approach to improving quality using peer-to-peer review, validated tools, and a focus on learning rather than judging.<sup>89</sup>

CMS on its own created the Partnership for Patients, a public/private partnership to improve the quality, safety, and affordability of health care for all Americans. The initiative promotes active collaboration by physicians, nurses, and other hospital personnel, as well as employers, patients and their advocates, and the federal and state governments to address tangible problems where approaches to quality improvement to improve outcomes exist but need broad-based adoption. Specifically, CMS is funding 26 hospital engagement networks to allow 3,700 hospitals to share best practices, and funding 82 sites to provide care transitions services to Medicare beneficiaries leaving the hospital through the agency's Community-Based Care Transitions Program; it is also encouraging patient engagement through both of these efforts.<sup>90</sup> The Partnership for Patients began in 2011, under the guidance of then acting CMS Administrator, Donald Berwick, and has targeted two basic areas for quality improvement with specific measureable outcome objectives:<sup>91</sup>

1. **Making Care Safer.** By the end of 2013, preventable hospital-acquired conditions would decrease by 40 percent compared to 2010.
2. **Improving Care Transitions.** By the end of 2013, preventable complications during transition from one care setting to another would be decreased so that all hospital readmissions would be reduced by 20 percent compared to 2010.

Unfortunately, this effort started without validated performance measures and currently lacks valid performance measures for most of the conditions. As a result, it will be exceedingly difficult to evaluate whether this program improved quality or safety for patients. Given the significant public investment in this program, rigorous evaluation should be a requirement.

A successful model of the strategic use of measures to accomplish substantial quality improvement can be found in the recent effort to reduce CLABSI ([see appendix for more information on CLABSI](#)). In this case, the primary motivation for physicians, nurses, and other hospital staff to participate in this activity was intrinsic—to reduce preventable mortality and morbidity caused by infections. One of the authors (P. Pronovost) who was instrumental in developing and leading the CLABSI-reduction programs believes that

public reporting of infection rates by states, *Consumer Reports*, the Commonwealth Fund, and, later, CMS had a generally positive effect on stimulating interest and action at senior levels of hospital management. Also contributing were the efforts of the Joint Commission with its national patient safety goals, and the Center for Disease Control and Prevention's (CDC) National Healthcare Safety Network and its work with state health departments to shine a spotlight on a problem that had a solution. The CDC recently reported that central-line bloodstream infections dropped by 41 percent between 2008 and 2011.<sup>92</sup>

Many opportunities for broad-based collaborations to improve hospital quality exist. CMS' current value-based purchasing efforts, requiring reporting on a raft of measures of varying usefulness and validity, should be replaced with the kind of strategic approach used in the national effort to reduce bloodstream infections.

Similarly, the current approach to improving the quality of care provided by physicians in Medicare needs to be reconsidered. Many physicians believe quality reporting on a few measures is being promoted as an end in itself, whether or not the particular measures chosen represent high priority for improvement, can accurately reflect the physician's actual quality of care, or are associated with meaningful patient outcomes. Drawing inferences about a physician's quality using a few measures peripheral to the physician's core professional activities may well be misleading and a diversion from the opportunity to engage physicians in substantive quality improvement activities.

Here, again, policymakers should be more strategic, focusing on clinical areas where measures are meaningful and valid, and where concerted multi-party collaboration could materially improve the health of the population. With this approach, it is likely that not all physicians in Medicare would be routinely measured; but much of what the public wants to know about physician competence and performance cannot be measured using the currently available measure sets. Strategies that work through peer assessment and fostering professionalism may also provide promising opportunities to improve quality and safety.

Observing the lack of "high leverage" processes of surgical care, particularly those specific to particular procedures, experts on surgical quality have suggested that surgeons be encouraged and supported to participate in surgical learning collaborative activities, with no reporting or rewards for individual performance.<sup>93</sup> Building on this suggestion, a more strategic approach would judge the effectiveness of

*"CMS' current value-based purchasing efforts, requiring reporting on a raft of measures of varying usefulness and validity, should be replaced with the kind of strategic approach used in the national effort to reduce bloodstream infections."*

care in terms of collective improvements in outcomes—on clinical quality, patient experience, and cost. Measurement would be integrated into quality improvement initiatives, such as those led by Regional Health Improvement Collaboratives,<sup>94</sup> national medical specialty societies,<sup>95</sup> national specialty boards,<sup>96</sup> and accountable care organizations (ACOs) as they come online. In short, Congress should allow CMS greater flexibility to provide physician incentives to actively participate in meaningful quality improvement collaboratives as an alternative or a complement to routine reporting and public reporting on a handful of quality measures.

### **3. Measure quality at the level of the organization, rather than the clinician.**

Historically, the physician has been viewed as the leader of medicine, with responsibility for the care and outcomes of patients; in iconic photographs and paintings, the physician is seen as a lone, heroic figure. Such a view has led to natural interest in the measurement of individual physicians' performance. It is therefore not surprising that some of the information brokers, including the *U.S. News and World Report* and many city magazines like the *Washingtonian* provide ratings of "top doctors," often based mostly on reputation, warranted or not.

However, this focus on the individual is flawed for most measures of quality and presents substantial technical challenges. Systems-based care is emerging as a key value within health care and a vital component of high quality care, while the notion that an individual health professional can be held accountable for the outcomes of patients in isolation from other health professionals and their work environment is becoming an outdated perspective. For example, better intensive care unit staffing sometimes mitigates the evidence that surgeons who perform more procedures achieve better outcomes.<sup>97</sup> The communication and coordination of services across providers is required to ensure that patients, many of whom have multiple conditions, are assisted through various health care settings.<sup>98</sup> For some aspects of care, such as diagnosis errors and patient experience, measuring at the individual

physician level might be considered. Nevertheless, focusing measurement on an individual runs counter to our goals in promoting teamwork and “systemness” as core health care delivery attributes.

For some professionals whose individual performance does matter, such as a surgeon in the operating room, there are rarely meaningful and valid process measures that reflect their individual performance anyway. In contrast, surgical outcomes depend crucially on the performance of the entire surgical team and the facility in which the procedure takes place. Also, consistent with the discussion of intrinsic motivation earlier, it is plausible that individuals respond differently to payment incentives than do organizations; assessment and pay-for-performance at the organizational instead of the individual level should be less likely to crowd out health professionals’ intrinsic motivations to provide high-quality care.

In addition to the conceptual issues with measuring an individual clinician’s performance, technical and statistical issues are also prominent. The attribution of a particular care process or outcome to a particular clinician is often difficult, if not impossible, to make. For example, several specialists, hospitalists, nurses, technicians, and others will typically care for a patient with a heart attack. Good estimates of performance require that the individual or group being evaluated have a sufficient number of observations to make inferences about their performance that are precise enough to be meaningful. Yet, many physicians and other health care professionals often lack sufficient volumes of certain types of patients to permit valid inferences about their performance. By focusing assessment on the organization, hospital unit, or clinic, rather than the individual clinician, measures can assess and promote team-based care while addressing many of the technical issues that can undermine the value of measurements. For virtually every performance measure evaluated (e.g., safety culture, patient experience, hand hygiene, infection rates, process measures) there is usually substantially greater variation among units within a hospital than among hospitals. The unit or clinic is therefore often the most effective focus for improvement.

While measuring at this level is conceptually right and technically easier than measuring a single individual’s performance, it nevertheless presents challenges. For example, it makes strategic sense to measure the quality of ACOs, especially to guard against the possibility that ACOs would stint on care as they receive increasing incentives to limit spending. Yet, recently, 31 Pioneer

ACOs participating in a major CMS demonstration sent CMS a letter criticizing both the agency’s use of measures that “are not yet mature” and the way in which CMS determined the thresholds for acceptable performance.<sup>99</sup> We expect they will work through the differences and arrive at a reasonable result.

Finally, measuring at the level of the organization does not imply that substandard individual performance should be tolerated. CMS and its contractors should aggressively use performance measures to identify such unacceptable performance and sanction or otherwise limit the ability of these practitioners to serve Medicare beneficiaries. But the role of measurement for “policing” the performance of individuals is different from public reporting to inform patient choice or to provide financial incentives to improve performance.

#### **4. Measure patient experience with care and patient-reported outcomes as ends in themselves.**

Performance measurement has too often been plagued by inordinate focus on technical aspects of clinical care—ordering a particular test or prescribing from a class of medication—such that the patient’s perspective of the care received may be totally ignored. Moreover, many patients, even with successful treatment, too often feel disrespected. Patients care not only about the outcomes of care but also and their personal experience with care. There is marked heterogeneity in the patient experience, and the quality of attention to patients’ needs and values can influence their course, whether or not short-term clinical outcomes are affected. Some patients have rapid recovery of function and strength, and minimal or no symptoms. Other patients may be markedly impaired, living with decreased function, substantial pain, and other symptoms, and with markedly diminished quality of life. It would be remiss to assume that these two groups of patients have similar outcomes just because they have avoided adverse clinical outcomes such as death or readmission.

In recommending a focus on measuring outcomes rather than care processes, we consider surveys or other approaches to obtaining the perspectives of patients on the care they receive to be an essential component of such outcomes. When designed and administered appropriately, patient experience surveys provide robust measures of quality, and can capture patient evaluation of care-focused communication with nurses and physicians.<sup>100</sup> And while patient-reported measures appear to be correlated with better outcomes, we believe they are worth collecting and working to improve in their own right, whether or not better



experiences are associated with improved clinical outcomes.<sup>101</sup>

We believe that measuring patient experience is not only important because it can facilitate care that improves clinical outcomes, but also because it represents an important outcome in its own right. If our health system is truly to commit itself to the goal of delivering patient-centered care, it requires assessment of patients' experiences with the care they receive and self-reported health status and functioning—whether or not they are associated with commonly-measured outcomes such as mortality, complications, errors, and avoidable readmissions. With the growing array of scientifically rigorous surveys of patient experiences with care,<sup>102</sup> we now have the capacity to incorporate standardized assessments of that experience into the measurement enterprise, increasing our sensitivity to the detection of differences in the results that are being achieved by provider organizations, assuming that we can adequately take into account baseline differences in patient characteristics. Given the inevitable gaps in both process and outcome measures for specific areas of clinical care, it is important to realize that patient experience is ubiquitous and can be drawn upon to measure a broad range of performance.

### **5. Use measurement to promote the concept of the rapid-learning health care system.**

Initiatives to promote performance measurement need to be accompanied by support to improve care. Quality measure data should not only be technically correct, but should be organized such that their dissemination is a resource to aid in quality improvement activities. As such, quality measurement should be viewed as just one component of a learning health care system that also includes advancing the science of quality improvement, building providers' capacity to improve care, transparently reporting performance, and creating formal accountability systems.

There are several strategies to make quality measure data more actionable for quality improvement purposes. For example, for publicly reported outcome measures, CMS provides hospitals with lists of the patients who are included in the calculation. Since the outcomes may occur outside the hospital for mortality and for readmissions that are at other hospitals, this information is often beyond what the hospitals already have available to them. These data give providers the ability to investigate care provided to individual patients, which in turn can support a variety of quality improvement efforts.

In addition, collaborative activities among institutions can produce insights that may elude them individually. Measures can help identify top performers, and detailed analysis can identify what distinguishes those who excel. As an example, the marked improvement nationwide in the “door-to-balloon” time it takes patients experiencing symptoms of a heart attack to receive a treatment to open up occluded coronary arteries was largely a result of relevant and valid measurement of provider-specific timeliness, followed by intense investigation of the features of top performance, and only then a national campaign to transform practice using the best practices uncovered by the top performers – all facilitated by the intrinsic motivation of health professionals on the front lines to improve patient outcomes.

To facilitate a learning health care system, investments are also needed to advance quality improvement sciences and to build capacity among provider organizations to practice these sciences. For example, although root causes analysis is a promising tool, its full potential has not been realized in health care; a likely explanation, at least in part, is that health care is one of the only risky industries in which lawyers and practitioners, rather than safety experts with formal training, investigate adverse events. Promising efforts to improve quality and safety are based on adherence to professional norms and include peer-to-peer review, a technique borrowed from the nuclear industry.<sup>103</sup> In addition, EHR vendors and other medical device manufacturers will need to agree to share their data and open it for analysis.

### **6. Invest in the basic science of measurement development and applications, including an emphasis on anticipating and preventing unintended adverse consequences.**

In describing the problems with process measures and the challenges with outcome measures above, the unfortunate reality is that there is no body of expertise with responsibility for addressing the science of performance measurement. NQF comes closest, and while it addresses some scientific issues when deciding whether to endorse a proposed measure, NQF is not mandated to explore broader issues to advance the science of measure development, nor does it have the financial support or structure to do so. An infrastructure is needed to gain national consensus on: what to measure, how to define the measures, how to collect the data and survey for events, what is the accuracy of EHRs as a source of performance, the cost-effectiveness of various measures, how to reduce the

costs of data collection, how to define thresholds for measures regarding their accuracy, and how to prioritize the measures collected (informed by the relative value of the information collected and the costs of data collection).

Despite this broad research agenda, there is little research funding to advance the basic science of performance measurement. Given the anticipated broad use of measures throughout the health system, funding can be a public/private partnership modeled after the Patient-Centered Outcomes Research Institute or a federally-funded initiative, perhaps centered at AHRQ. Given budgetary constraints, finding the funding to support the science of measurement will be a challenge. Yet, the costs of misapplication of measures and incorrect judgments about performance are substantial.

Moreover, the science of performance measurement and improvement needs an academic home. While many medical and health policy societies and associations have sections on quality or quality measurement, no professional society primarily focuses on the science of quality measurement and improvement. Such an entity could set standards for and advance the science of quality measurement, thereby moving the policy discussion from whether measures are good enough to use despite their flaws to a more fundamental discussion of how to achieve good measures, how to assess whether current measures measure up, and whether the costs of attaining good measures are worth the benefits. Professional societies, such as the American Heart Association, have an important role in speaking authoritatively about the science of clinical issues; performance measurement lacks a similar authoritative voice.

Such an endeavor needs to explicitly consider the unintended, yet harmful, consequences of misapplication of performance measures, whether resulting from the measures themselves, in how they are reported and assessed, or in the costs of collecting invalid performance data. There is substantial literature detailing such untoward consequences,<sup>104</sup> some from measures experts who promote the use of performance measurement. For example, some have expressed concern that unless carefully designed, public reporting and pay-for-performance programs will increase racial and ethnic disparities.<sup>105</sup>

**7. Task a single entity with defining standards for measuring and reporting quality and cost data, similar to the role the SEC serves for the reporting of corporate financial data, to improve the validity,**

**comparability, and transparency of publicly-reported health care quality data.**

There is a plethora of health care quality data being pushed out to the public, yet no rules to assure the accuracy of what is being presented publicly. The health care industry lacks standards for how valid a quality measure should be before it is used in public reporting or pay-for-performance initiatives, although some standards have been proposed. The NQF does a good job of reviewing and approving proposed measures presented to it, but lacks the authority to establish definitive quantitative standards that would apply broadly to purveyors of performance measures. Yet, as discussed earlier, many information brokers publically report provider performance without transparency and without meeting basic validity standards. Indeed, even CMS, which helps support NQF financially, has adopted measures for the PQRS that have not undergone NQF review and approval. Congress now is considering “SGR repeal,” or sustainable growth rate legislation, that would have CMS work directly with specialty societies to develop measures and measurement standards, presumably without requiring NQF review and approval.<sup>106</sup>

Without industry standards, payers, policy makers, and providers often become embroiled in a tug-of-war; with payers and policy-makers asserting that existing measures are good enough, and providers arguing they are not. Most often, neither side has data on how good the contested measures actually are. Most importantly, the public lacks valid information about quality, especially outcomes, and costs.

Indeed, most quality measurement efforts struggle to find measures that are scientifically sound yet feasible to implement with the limited resources available. Unfortunately, too often feasibility trumps sound science. In the absence of valid measures, bias in estimating the quality of care provided will likely increase in proportion to the risks and rewards associated with performance. The result is that the focus of health care organizations may change from improving care to “looking good” to attract business. Further, conscientious efforts to reduce measurement burden have significantly compromised the validity of many quality measures, making some nearly meaningless, or even misleading. Unfortunately, measurement bias often remains invisible because of limited reporting of data collection methods that produce the published results. In short, the measurement of quality in health care is neither standardized nor consistently accurate and reliable.

---

In short, while the number of performance measures is growing, the health care field lacks an entity to create the rules for reporting quality and cost data; as a result, the great variation in performance measure specifications is slowing efforts to advance quality—at times creating conflict over opposing findings.

The field of quality measurement could advance significantly if providers and policy-makers agreed on validity thresholds and transparently reported the validity of their quality measure data. Before the SEC was created in the aftermath of the Wall Street Crash of 1929, when looking at companies' financial data, the information provided by one business could not be compared to another; there were no standard rules for reporting performance. Congress established the SEC as an independent, nonpartisan government entity to, among other things, help ensure standards in the disclosure of financial information, make financial performance transparent, audit businesses, ensure compliance with rules, and apply penalties for transgressions.

Policymakers will need to consider whether such an entity should be housed at AHRQ; should be a public-private partnership, such as NQF; or should be a separate, new government entity. Such a commission could promote standardization, transparency, and auditing of the reporting of quality and cost measures. Consistent with First Amendment guarantees of free speech, we would not provide such an entity regulatory authority to require adherence to standards. Rather, we would anticipate that organizations would voluntarily

seek to comply with the applicable standards for reporting performance measures. Under the model, this entity would set the rules for the development of measures and the transparent reporting of performance on these measures, analyze progress (with input from clinicians, patients, employers, and insurers), and audit publicly-reported quality measure data. Private sector information brokers could then conduct secondary analyses of the reports, much like happens in the financial industry through companies like Bloomberg. This SEC-like model would thus ensure that all publicly-reported quality measure data are generated from a common basis in fact and allow apples-to-apples comparisons across provider organizations.

## Conclusion

The interest in promoting a health care system that rewards performance needs to be balanced with the practical challenges faced when measuring performance. Improvement requires substantial investments in the underlying science of measurement, greater care in communicating measurement results, greater attention to the role of measures in quality improvement efforts, and using performance data in more strategic ways. The adoption of flawed measurement approaches that do not accurately discriminate between providers can undermine professional and public support for provider accountability, reward indiscriminately, and divert attention from more appropriate and productive quality improvement efforts.

---

*The views expressed are those of the authors and should not be attributed to the Robert Wood Johnson Foundation or the Urban Institute, its trustees, or its funders.*

## **About the Authors and Acknowledgments**

Robert A. Berenson, MD is an institute fellow at the Urban Institute. Peter J. Pronovost, MD, PhD is the director of the Armstrong Institute for Patient Safety and Quality at Johns Hopkins, as well as Johns Hopkins Medicine's senior vice president for patient safety and quality. Harlan M. Krumholz, MD, is the director of the Yale-New Haven Hospital Center for Outcomes Research and Evaluation, director of the Robert Wood Johnson Foundation Clinical Scholars program at Yale University, and the Harold H. Hines, Jr. professor of cardiology, investigative medicine, and public health. The authors thank Lawrence Casalino, MD, PhD, chief of the Division of Outcomes and Effectiveness Research and an associate professor at Weill Cornell Medical College, and Andrea Ducas, MPH and Anne Weiss, MPP of the Robert Wood Johnson Foundation for their helpful comments on this paper. This research was funded by the Robert Wood Johnson Foundation.

## **About the Urban Institute**

The Urban Institute is a nonprofit, nonpartisan policy research and educational organization that examines the social, economic, and governance problems facing the nation.

## **About the Robert Wood Johnson Foundation**

The Robert Wood Johnson Foundation focuses on the pressing health and health care issues facing our country. As the nation's largest philanthropy devoted exclusively to health and health care, the Foundation works with a diverse group of organizations and individuals to identify solutions and achieve comprehensive, measurable, and timely change. For 40 years the Foundation has brought experience, commitment, and a rigorous, balanced approach to the problems that affect the health and health care of those it serves. When it comes to helping Americans lead healthier lives and get the care they need, the Foundation expects to make a difference in your lifetime. For more information, visit [www.rwjf.org](http://www.rwjf.org). Follow the Foundation on Twitter [www.rwjf.org/twitter](https://twitter.com/rwjf) or Facebook [www.rwjf.org/facebook](https://www.facebook.com/rwjf).

## The Role of Performance Measurement in Improving Cardiovascular Care

Twenty years ago, many patients with heart disease were not being treated in accordance with available evidence-based best practices. For example, among patients discharged from the hospital after an acute myocardial infarction, only about half were treated with beta-blocker drugs and only about two-thirds with aspirin.<sup>1,2</sup> Many other evidence-based treatments were similarly underused,<sup>3</sup> and treatment of patients presenting to the hospital with an acute myocardial infarction was often delayed.<sup>4</sup> There was also troubling regional variation, with some areas of the country performing markedly worse than the national average on the measures being used, which was already low.

The past two decades have seen a remarkable transformation in cardiovascular care. In the past decade alone, hospitalizations for acute myocardial infarction have dropped by more than 25 percent and hospitalizations for heart failure have fallen by more than 30 percent.<sup>5,6</sup> Mortality after hospitalizations for acute myocardial infarction has also decreased by more than 20 percent.<sup>7</sup> These improvements have occurred in an era without the introduction of new blockbuster drugs, but with a strong emphasis on performance measurement and quality improvement.

The key change began with the decision by CMS to support the explicit measurement of care provided to patients with an acute myocardial infarction. First with the Cooperative Cardiovascular Project pilot, launched in the early 1990s, and then with the national Cooperative Cardiovascular Project, which followed a few years later, the agency exposed gaps in the quality of care and supported efforts to improve. This performance measurement provided objective information about the quality of care being delivered.<sup>8</sup>

Of note, this broad-based change in practice occurred without financial incentives. Instead, the motivation derived from intrinsic motivation related to professionalism (clinicians' desires to provide the best care they could and to safeguard their reputations). Supportive organizations, including the American College of Cardiology, the American Heart Association, Medicare's Quality Improvement Organizations, consortia of hospitals, and others merely encouraged health care professionals to embrace the responsibility to improve care.

A prime example of measurement stimulating improvement through these programs is the experience with delays in treatment, which is measured as "door-to-balloon" time—the period from when the patient arrives at the hospital with symptoms of an acute myocardial infarction to the time that blood flow in a blocked artery is restored with an emergency percutaneous coronary intervention. The longer the delay, the more damage is done and the more likely the patient is to die. Measurement of door-to-balloon time, later required by CMS, revealed that less than a third of patients were being treated within the guideline-recommended time of 90 minutes.<sup>9</sup> National measurement, through an industry-sponsored registry, enabled the identification of exemplary hospitals that were treating patients faster than the vast majority. With funding from the National Institutes of Health, research then identified the strategies employed by the top performers.<sup>10</sup> A national campaign to disseminate those strategies ensued, resulting today in more than 90 percent of patients being treated within 90 minutes.<sup>11</sup>

<sup>1</sup> Krumholz HM, Radford MJ, Wang Y, et al. "National Use and Effectiveness of Beta-blockers for the Treatment of Elderly Patients After Acute Myocardial Infarction: National Cooperative Cardiovascular Project." *JAMA*, 280(7): 623-629, 1998.

<sup>2</sup> Krumholz HM, Radford MJ, Ellerbeck EF, et al. "Aspirin for Secondary Prevention After Acute Myocardial Infarction in the Elderly: Prescribed Use and Outcomes." *Annals of Internal Medicine*, 124(3): 292-298, 1996.

<sup>3</sup> Ellerbeck EF, Jencks SF, Radford MJ, et al. "Quality of Care for Medicare Patients With Acute Myocardial Infarction: A Four-State Pilot Study From the Cooperative Cardiovascular Project." *JAMA*, 273(19): 1509-1514, 1995.

<sup>4</sup> McNamara RL, Herrin J, Bradley EH, et al. "Hospital Improvement in Time to Reperfusion in Patients with Acute Myocardial Infarction, 1999 to 2002." *Journal of American College of Cardiology*, 47(1): 45-51, 2006.

<sup>5</sup> Chen J, Normand SL, Wang Y, et al. "National and Regional Trends in Heart Failure Hospitalization and Mortality Rates for Medicare Beneficiaries, 1998-2008." *JAMA*, 306(15): 1669-1678, 2011.

<sup>6</sup> Wang OJ, Wang Y, Chen J, et al. "Recent Trends in Hospitalization for Acute Myocardial Infarction." *American Journal of Cardiology*, 109(11): 1589-1593, 2012.

<sup>7</sup> Krumholz HM, Wang Y, Chen J, et al. "Reduction in Acute Myocardial Infarction Mortality in the United States: Risk-standardized Mortality Rates from 1995-2006." *JAMA*,

302(7):767-773, 2009.

<sup>8</sup> Ellerbeck et al., 1995; Marciniak TA, Ellerbeck EF, Radford MJ, et al. "Improving the Quality of Care for Medicare Patients with Acute Myocardial Infarction: Results from the Cooperative Cardiovascular Project." *Journal of the American Medical Association*, 279(17):1351-1357, 1998.

<sup>9</sup> McNamara et al., 2006.

<sup>10</sup> Bradley EH, Herrin J, Wang Y, et al. "Strategies for Reducing the Door-to-Balloon Time in Acute Myocardial Infarction." *New England Journal of Medicine*, 355: 2308-2320, 2006.

<sup>11</sup> Krumholz HM, Bradley EH, Nallamothu BK, et al. "A Campaign to Improve the Timeliness of Primary Percutaneous Coronary Intervention: Door-to-Balloon: An Alliance for Quality." *Journal of the American College of Cardiology: Cardiovascular Interventions*, 1(1): 97-104, 2008; Krumholz HM, Herrin J, Miller LE, et al. "Improvements in Door-to-Balloon Time in the United States, 2005 to 2010." *Circulation*, 124(9): 1038-1045, 2011.

[Go back to the paper](#)

## Key Players in the Quality Measurement Enterprise

Key entities involved in quality measurement include the National Committee for Quality Assurance (NCQA), the National Quality Forum (NQF), The Joint Commission, and the Agency for Healthcare Research and Quality (AHRQ).

**NCQA** is a private, nonprofit institution that has been reviewing and accrediting health insurance plans since 1991. More recently, NCQA developed accreditation and certification programs for a range of health care entities, including groups of provider organizations that want to become accountable care organizations and practices that want to be patient-centered medical homes. In 1992, NCQA took responsibility for maintaining a set of newly-developed quality measures called HEDIS, which had been developed by a group of employers and HMOs the year before. In 1995, NCQA used these measures to release the first-ever report card on health plan performance. Today, HEDIS measures are used by a range of organizations to measure performance at both the plan and provider level, and are largely focused on outpatient care. The full HEDIS set includes 80 quality measures divided into five domains of care and is updated every year. NCQA follows a standardized process for developing its measures, which includes multiple stages of internal and external review by a range of advisory groups. NCQA uses three overarching criteria to determine the desirability of adding a new measure: relevance, scientific soundness, and feasibility. Operationally, numerous other criteria help define these major criteria.<sup>1</sup> NCQA is governed by an independent 15-member board of directors, and receives support through grants and corporate sponsorships<sup>2</sup> and through revenues from certification fees it charges plans and providers.

**NQF** is a private, nonprofit membership-based organization that builds consensus around quality improvement priorities and goals, evaluates and endorses quality standards and measures submitted to it by a variety of types of organizations, and conducts education and outreach activities around quality improvement and performance measurement. NQF's membership includes consumer organizations, public and private purchasers, physicians, nurses, hospitals, accrediting and certifying bodies, supporting industries, and healthcare research and quality improvement organizations. NQF's primary role in the quality landscape is evaluating measures that other organizations develop; Many HEDIS measures, for example, are endorsed by NQF. To date, the organization has endorsed nearly 700 measures, all of which are publicly accessible in their [database](#). NQF evaluates all submitted standards according to four major criteria: importance, scientific acceptability, feasibility, and usability (although if the standard does not meet the first two criteria, it is not considered against the other criteria). Despite the fact that NQF assesses measures against these criteria, it does not establish specific standards that payers and information brokers must adhere to when publicly reporting measures or applying payment incentives to providers in pay-for-performance programs using NQF-endorsed measures. NQF does not endorse proprietary measures, for which the specifications or performance are not in the public domain. NQF is governed by a 33-member board, and receives funding from both public and private sources, including grants from foundations, corporations, and contracts from the federal government, particularly the Department of Health and Human Services' (HHS's) Centers for Medicare and Medicaid Services (CMS).

**The Joint Commission**, formerly the Joint Commission on Accreditation of Healthcare Organizations (JCAHO), is an independent, not-for-profit organization that accredits more than 20,000 health care organizations and programs in the United States. This volume stems in part from the fact that states and CMS require hospitals and other health

care organizations to be accredited by the Joint Commission in order to be eligible to participate in Medicaid and Medicare. The Joint Commission provides accreditation and certification services for hospitals, home health care organizations, nursing homes, behavioral health care organizations, ambulatory care providers, and independent or freestanding clinical laboratories. It develops performance standards that address elements of operation, such as patient care, medication safety, infection control, and consumer rights. In 1997, the Joint Commission introduced the ORYX initiative, which includes outcomes and other performance measurement data into the accreditation process. In 2010, it categorized its performance measures into accountability and non-accountability measures, placing more emphasis on an organization's performance on accountability measures, which focus on research, proximity, accuracy, and adverse events. The organization is governed by a 32-member Board of Commissioners that includes physicians, administrators, nurses, employers, a labor representative, quality experts, a consumer advocate and educators. It receives support through accreditation fees, as well as corporate sponsorships.<sup>3</sup>

AHRQ is a federal agency within the Department of Health and Human Services. AHRQ's mission is to improve the quality, safety, efficiency, and effectiveness of health care nationwide.<sup>4</sup> AHRQ's Consumer Assessment of Healthcare Providers and Systems (CAHPS) program is a multi-year initiative to support and promote the assessment of consumers' experiences with health care. Through the CAHPS program, AHRQ has developed standardized patient experience surveys that are widely used by health plans, doctor's offices, and dialysis facilities,<sup>5</sup> and maintains a benchmarking database containing the results of various organizations' administrations of this survey.<sup>6</sup> The various versions of the CAHPS surveys ask patients and their caregivers to report on and evaluate their experiences with health care. These surveys focus on elements of care that consumers deem most important, as well as aspects of quality that consumers are best qualified to assess, such as the communication skills of providers and ease of access to health care services.<sup>7</sup> AHRQ also maintains a clearinghouse of a variety of types of quality measure specifications and quality improvement resources.<sup>8</sup>

<sup>1</sup> *Desirable Attributes of HEDIS*. Washington, DC: National Committee for Quality Assurance, [www.ncqa.org/tabid/415/Default.aspx](http://www.ncqa.org/tabid/415/Default.aspx) (accessed April 2013).

<sup>2</sup> *Current Sponsors*. Washington, DC: National Committee for Quality Assurance, [www.ncqa.org/Sponsorship/CurrentSponsors.aspx](http://www.ncqa.org/Sponsorship/CurrentSponsors.aspx) (accessed April 2013).

<sup>3</sup> *Facts About The Joint Commission*. Oakbrook Terrace, IL: The Joint Commission, [www.jointcommission.org/about\\_us/fact\\_sheets.aspx](http://www.jointcommission.org/about_us/fact_sheets.aspx) (accessed April 2013).

<sup>4</sup> *AHRQ at a Glance*. Rockville, MD: Agency for Healthcare Research and Quality, [www.ahrq.gov/about/mission/glance/index.html](http://www.ahrq.gov/about/mission/glance/index.html) (accessed April 2013).

<sup>5</sup> *CAHPS Surveys and Guidance*. Rockville, MD: Agency for Healthcare Research and Quality, [www.cahps.ahrq.gov/surveysguidance.htm](http://www.cahps.ahrq.gov/surveysguidance.htm) (accessed April 2013).

<sup>6</sup> *CAHPS*. Rockville, MD: Agency for Healthcare Research and Quality, <http://cahps.ahrq.gov> (accessed April 2013).

<sup>7</sup> *About CAHPS*. Rockville, MD: Agency for Healthcare Research and Quality, [www.cahps.ahrq.gov/about.htm](http://www.cahps.ahrq.gov/about.htm) (accessed April 2013).

<sup>8</sup> *Quality Measure Tools and Resources*. Rockville, MD: Agency for Healthcare Research and Quality, [www.ahrq.gov/professionals/quality-patient-safety/quality-resources/index.html](http://www.ahrq.gov/professionals/quality-patient-safety/quality-resources/index.html) (accessed April 2013).

[Go back to the paper](#)

## Does Pay-for-Performance Work?

As required by statute, pay-for-performance programs are being launched by the Centers for Medicare & Medicaid Services (CMS) even though quality of care on the CMS core measure set is already improving substantially without additional financial incentives.<sup>1</sup> Since October 2012, under the Hospital Value-based Purchasing Program, 1 percent of hospital payments are adjusted based on performance on specific process, patient experience, and outcome measures—rising to 2 percent in October 2017. Extra payments are provided to hospitals for both achievement and improvement in performance. To promote hospital activity to perform even better, some have called for a much greater percentage at risk based on performance to increase the financial stakes.<sup>2</sup>

Yet, the demonstration that informed the design of the hospital VBP program—the Premier Hospital Quality Incentive Demonstration Project (HQID)—did not actually produce better results than other hospitals, which have also demonstrated improved scores on what were mostly process measures. Indeed, two evaluations found little evidence that the demonstration's use of financial incentives to incentivize improved performance led to reduced

mortality rates beyond those achieved with public reporting alone.<sup>3,4</sup> Another hospital pay-for-performance program implemented in Medicaid in Massachusetts, with much larger financial incentives than in the Premier demonstration, also showed that pay-for-performance had no effect on health outcomes.<sup>5</sup>

In contrast, a pilot in the northwest region of England, built on the Premier demo approach, found that mortality for conditions in the pay-for-performance program—pneumonia, heart failure, and heart attack—decreased, although statistically significant only for pneumonia. There was a small increase in mortality for the larger number of conditions not being rewarded, although the increase did not achieve statistical significance.<sup>6</sup> Of note, participating hospitals adopted a range of quality improvement strategies in response to the performance incentives, to attempt to accomplish systemic change. Also, this incentive program offered larger bonuses and a greater likelihood of achieving bonuses than the U.S. HQID prototype—leading some to speculate that stronger incentives using more measures might achieve a better result from pay-for-performance.<sup>7</sup>

Meanwhile, the findings on pay-for-performance for *physicians* are mixed. In 2004, the United Kingdom introduced a major pay-for-performance approach—the Quality Outcomes Framework—with 136 measures for general practitioners. Payments were generous, adding up to 25 percent more to general practitioners' (GPs') income; more than 99 percent of eligible physicians participated.<sup>8</sup> Analysis showed that the approach did accelerate improvement on measured performance for asthma and diabetes, but not coronary heart disease in the short term; in addition, once targets were reached, improvement in the quality slowed, while the quality of care for two conditions not linked to incentives actually declined, as did scores on measures assessing continuity of care.<sup>9</sup> Further analyses were mixed. One showed improvement in process performance among GPs led to outcome improvements for diabetes, coronary heart disease, stroke, epilepsy, and hypertension, whereas another found that reported improvements in blood pressure control did not reduce stroke, heart attack, or all-cause mortality as would be expected.

In the United States, a major pay-for-performance effort has been carried out by the Integrated Healthcare Association (IHA), an organization with broad representation by health plans, medical groups, purchasers, and consumers. In contrast to the UK approach, IHA has been providing small bonuses for almost a decade to medical groups based on performance on individual measures in the areas of clinical quality, patient experience, and health information technology use. Studies<sup>10,11</sup> have also shown mixed results, with one concluding that medical groups responses to the pay-for-performance incentives “did not translate into the breakthrough improvement in quality desired by plans and purchasers.”<sup>12</sup>

Overall, studies do not provide much support for reliance on process measures to improve quality for hospital and physician care, although it seems likely that the details—such as the strength of the incentives, the number and selection of performance measures being used, and restrictions on how bonuses can be used—may affect the success of pay-for-performance programs.<sup>13</sup> The message may be that we have not yet determined how such incentives can be most effectively applied, the program theory for how they work, the extent to which they motivate hospital managers versus physicians, or even if they are sustainably effective in any form over the long run.

---

<sup>1</sup> Van Lare JM and Conway PH. “Value-Based Purchasing—National Programs to Move from Volume to Value.” *New England Journal of Medicine*, 367: 292-295, 2012.

<sup>2</sup> Werner RM and Adams Dudley R. “Medicare’s New Hospital Value-Based Purchasing Program Is Likely To Have Only A Small Impact On Hospital Payments.” *Health Affairs*, 31(9): 1932-1940, 2012.

<sup>3</sup> Ryan AM. “Effects of the Premier Hospital Quality Incentive Demonstration on Medicare Patient Mortality and Cost.” *Health Services Research*, 44(3): 821-842, 2009.

<sup>4</sup> Jha AK, Joynt KE, Orav EJ, et al. “The Long-Term Effect of Premier Pay for Performance on Patient Outcomes.” *New England Journal of Medicine*, 366(17): 1606-1615, 2012.

<sup>5</sup> Ryan AM and Bluestein J. “The Effect of the MassHealth Hospital Pay-for-Performance Program on Quality.” *Health Services Research*, 46(3): 712-728, 2011.

<sup>6</sup> Sutton M, Nikolova S, Boaden R, et al. “Reduced Mortality with Hospital Pay for Performance in England.” *New England Journal of Medicine*, 367(19): 1821-1828, 2012.

<sup>7</sup> Epstein AM. “Will Pay for Performance Improve the Quality of Care? The Answer Is in the Details.” *New England Journal of Medicine*, 367: 1852-1853, 2012.

<sup>8</sup> Campbell SM, Reeves D, Kontopantelis E, et al. “Effects of Pay for Performance on the Quality of Primary Care in England.” *New England Journal of Medicine*, 361(14): 368-378, 2009.



<sup>9</sup> *ibid.*

<sup>10</sup> Rosenthal M. *Major Developments in Results-Based Financing (RBF) in OECD Countries: Country Summaries and Mapping of RBF Programs*. Oakland, CA: Integrated Healthcare Association. (Forthcoming 2013).

<sup>11</sup> Mullen KJ, Frank RG and Rosenthal MB. “Can You Get What You Pay For? Pay for Performance and the Quality of Healthcare Providers.” *RAND Journal of Economics*, 41(1): 64-91, 2010.

<sup>12</sup> Damberg CL, Raube K, Teleki SS, et al. “Taking Stock of Pay for Performance: A Candid Assessment from the Front Lines.” *Health Affairs* 28(2): 517-525, 2009.

<sup>13</sup> Epstein, 2012.

[Go back to the paper](#)

## Central Line-Associated Blood Stream Infections

Central line-associated bloodstream infections (CLABSI) killed nearly 31,000 inpatients in the United States in 2002.<sup>1</sup> In response to growing awareness of this problem, health providers, hospitals, and payers have mounted various activities which together have produced major reductions in mortality rates among intensive care unit (ICU) patients, although not among other inpatients.<sup>2</sup> The major success can be attributed to collaborations among ICU clinicians to adopt evidence-based practices known to prevent such infections. A pilot project in one ICU at Johns Hopkins<sup>3</sup> was expanded to the statewide Keystone collaborative in Michigan and reduced CLABSIs by 66 percent in 103 ICUs.<sup>4,5</sup> Hospital mortality in Michigan decreased significantly once the collaborative was implemented,<sup>6</sup> with an estimated cost savings of \$1.1 million per year.<sup>7</sup> Recent estimates by the Centers for Disease Control and Prevention attribute a 58 percent reduction in ICU-related CLABSIs between 2001 and 2009 to large scale programs, such as the Keystone project, and the spread of the culture and CLABSI interventions to every state. Over 1,100 hospitals participated in this unique AHRQ-funded collaborative effort among Johns Hopkins physicians, the Michigan Hospital Association, the American Hospital Association, and many state affiliates and individual hospitals. Participating hospitals reduced CLABSI rates by 40 percent, achieving a mean infection rate of 1.1 per 1000 catheter days, a rate previously believed to be unattainable.

<sup>1</sup> Klevens RM, Edwards JR, Richards CL Jr, et al. “Estimating Health Care Associated Infections and Deaths in U.S. Hospitals, 2002.” *Public Health Reports*, 122(2):160-166, 2007.

<sup>2</sup> Centers for Disease Control and Prevention. “Vital Signs: Central Line Bloodstream Infections—United States, 2008 and 2009.” *Morbidity and Mortality Weekly Report*, 60:243-248, 2011.

<sup>3</sup> Brenholz SM, Pronovost PJ, Lipsett PA, et al. “Eliminating Catheter-Related Bloodstream Infections in the Intensive Care Unit.” *Critical Care Medicine*, 32(10):2014-2020, 2004.

<sup>4</sup> Pronovost PJ, Needham DM, Berenholz, SM, et al. “An Intervention to Decrease Catheter Bloodstream Infections in the ICU.” *New England Journal of Medicine*, 355(26):2725-2732, 2006.

<sup>5</sup> Pronovost PJ, Goeschel CA, Colantuoni E, et al., “Sustaining Reductions in Catheter Related Bloodstream Infections in Michigan Intensive Care Units: Observational Study.” *British Medical Journal*, 340:c309, 2010.

<sup>6</sup> Lipitz Snyderman A, Steinwachs D, Needham DM, et al. “Impact of a Statewide Intensive Care Unit Quality Improvement Initiative on Hospital Mortality and Length of Stay: Retrospective Comparative Analysis.” *British Medical Journal*, 342:d219, 2011.

<sup>7</sup> Waters HR, Korn R Jr, Colantuoni E, et al., “The Business Case for Quality Economic Analysis of the Michigan Keystone Patient Safety Program in ICUs.” *American Journal of Medical Quality*, 26(5):333-339, 2011.

[Go back to the paper](#)

## Notes

<sup>1</sup> Brook RH, McGlynn EA and Cleary PD. “Quality of Health Care. Part 2: Measuring Quality of Care.” *New England Journal of Medicine*, 335(13): 966-970, 1996; Porter ME. “What Is Value in Health Care?” *New England Journal of Medicine*, 363(26): 2477-83, 2010.

<sup>2</sup> Marciniak TA, Ellerbeck EF, Radford MJ, et al. “Improving the Quality of Care for Medicare Patients with Acute Myocardial Infarction: Results from the Cooperative Cardiovascular Project.” *Journal of the American Medical Association*, 279(17):1351-7, 1998; Krumholz HM, Wang Y, Chen J, et al. “Reduction in Acute Myocardial Infarction Mortality in the United States: Risk-Standardized Mortality Rates From 1995-2006.” *Journal of the American Medical Association*, 302(7): 767-73, 2009; Chassin MR, Loeb JM, Schmaltz SP, et al. “Accountability Measures—Using Measurement to Promote Quality Improvement.” *New England Journal of Medicine*, 363: 683-688, 2010.

<sup>3</sup> Pronovost PJ and Lilford R. “A Road Map For Improving The Performance Of Performance Measures.” *Health Affairs (Millwood)*, 30(4): 569-73, 2011.

<sup>4</sup> Ryan AM, Nallamouth BK and Dimik JB. “Medicare’s Public Reporting Initiative On Hospital Quality Had Modest Or No Impact On Mortality From Three Key Conditions,” *Health Affairs*, 31(5):585-592, 2012.

<sup>5</sup> Ryan AM. “Effects of the Premier Hospital Quality Incentive Demonstration on Medicare Patient Mortality and Cost.” *Health Services Research*, 44(3): 821-842, 2009; Jha AK, Orav JE, Zhonghe L, et al. “The Inverse Relationship Between Mortality Rates And Performance In The Hospital Quality Alliance Measures.” *Health Affairs*, 26(4): 1104-1110, 2007.

<sup>6</sup> Remus D. “Exploring the Nexus of Quality and Cost: Methodology and Preliminary Findings.” Premier, Inc., 2006, [www.premierinc.com/p4p/press/quality-cost-methods-paper3.pdf](http://www.premierinc.com/p4p/press/quality-cost-methods-paper3.pdf) (accessed July 2007).

<sup>7</sup> Relman AS. “Assessment and Accountability: The Third Revolution in Medical Care.” *New England Journal of Medicine*, 319(18): 1220-2, 1988.

<sup>8</sup> Jencks SF and Wilensky GR. “The Health Care Quality Improvement Initiative. A New Approach to Quality Assurance in Medicare.” *Journal of the American Medical Association*, 268(7): 900-903, 1992.

<sup>9</sup> Ellerbeck EF, Jencks SF, Radford MJ, et al. “Quality of Care for Medicare Patients with Acute Myocardial Infarction. A Four-state Pilot Study from the Cooperative Cardiovascular Project.” *Journal of the American Medical Association*, 273(19): 1509-1514, 1995.

<sup>10</sup> Kohn LT, Corrigan JM and Donaldson MS (eds). *To Err is Human: Building a Safer Health System*. Washington, DC: National Academy Press, 2000; Richardson, WC, Berwick, DM, Bristow, LR, et al. (eds). *Crossing the Quality Chasm: A New Health System for the 21<sup>st</sup> Century*. Washington: National Academy Press, 2001.

<sup>11</sup> McGlynn EA, Asch SM, Adams J, et al. “The Quality of Health Care Delivered to Adults in the United States.” *New England Journal of Medicine*, 348: 2635-2645, 2003.

<sup>12</sup> Mulvey GK, Wang Y, Lin Z, et al. “Mortality and Readmission for Patients with Heart Failure Among U.S. News & World Report's Top Heart Hospitals.” *Circulation, Cardiovascular Quality and Outcomes*, 2(6): 558-565, 2009; Krumholz HM, Rathore SS, Chen J, et al. “Evaluation of a Consumer-Oriented Internet Health Care Report Card: The Risk of Quality Ratings Based on Mortality Data.” *Journal of the American Medical Association*, 287(10): 1277-1287, 2002; Chen J, Radford MJ, Wang Y, et al. “Performance of the ‘100 Top Hospitals’: What Does the Report Card Report?” *Health Affairs*, 18(4): 53-68, 1999.

<sup>13</sup> Sack K, “Report Finds Improved Performance by Hospitals,” *New York Times*, September 14, 2011.

<sup>14</sup> Pronovost, PJ, Miller, M and Wachter RM. “The GAAP in Quality Measurement and Reporting.” *Journal of the American Medical Association*, 298(15):1800-1802, 2007.

<sup>15</sup> *ibid.*

<sup>16</sup> Donabedian A. “Evaluating the Quality of Medical Care.” *The Milbank Memorial Fund Quarterly*, 44(3): 166-206, 1996.

<sup>17</sup> Krumholz HM, Wang Y, Mattera JA, et al. “An Administrative Claims Model Suitable for Profiling Hospital Performance Based on 30-day Mortality Rates Among Patients with an Acute Myocardial Infarction.” *Circulation*, 113(13): 1683-1692, 2006.

<sup>18</sup> *CAHPS Surveys and Guidance*. Rockville, MD: Agency for Healthcare Research and Quality, <http://cahps.ahrq.gov/surveysguidance.htm> (accessed April 2013).

<sup>19</sup> *Patient Reported Outcomes Measurement Information System (PROMIS)*. Silver Spring, MD: National Institutes of Health, [www.nihpromis.org](http://www.nihpromis.org) (accessed April 2013).

- <sup>20</sup> *Reform in Action: Can Publicly Reporting the Performance of Health Care Providers Spur Quality Improvement?* Princeton, NJ: Robert Wood Johnson Foundation, 2012, [www.rwjf.org/en/research-publications/find-rwjf-research/2012/08/reform-in-action--can-publicly-reporting-the-performance-of-heal.html](http://www.rwjf.org/en/research-publications/find-rwjf-research/2012/08/reform-in-action--can-publicly-reporting-the-performance-of-heal.html) (accessed April 2013).
- <sup>21</sup> Ferris RG and Torchiana DF. “Public Release of Clinical Outcomes Data—Online CABG Report Cards.” *New England Journal of Medicine* 363:1593-1595, 2010.
- <sup>22</sup> Totten AM, Wagner J, Tiwari A, et al. *Closing the Quality Gap: Revisiting the State of the Science (Vol. 5: Public Reporting as a Quality Improvement Strategy, Evidence Reports/Technology Assessments, No. 208.5)*. Rockville, MD: Agency for Health Care Research and Quality, 2012.
- <sup>23</sup> Totten et al.
- <sup>24</sup> Draper DA, Liebhaber A and Ginsburg PB. “High-Performance Health Plan Networks: Early Experiences.” Issue Brief No. 111. Washington, D.C.: Center for Studying Health System Change, 2007, [www.hschange.com/CONTENT/929/929.pdf](http://www.hschange.com/CONTENT/929/929.pdf) (accessed April 2013).
- <sup>25</sup> Ryan Ernst C, “Integrated Healthcare Association Launches Value Based Pay for Performance As It Marks 10 Years of P4P Leadership in Partnership with California Health Plans and Physician Organizations.” Oakland, CA: Integrated Healthcare Association, 2011, [www.ihc.org/pdfs\\_documents/home/VBPRReleaseFinal09162011.pdf](http://www.ihc.org/pdfs_documents/home/VBPRReleaseFinal09162011.pdf).
- <sup>26</sup> Available at <http://medicare.gov>.
- <sup>27</sup> Van Lare JM and Conway PH. “Value-Based Purchasing—National Programs to Move from Volume to Value.” *New England Journal of Medicine*, 367: 292-295, 2012.
- <sup>28</sup> Van Lare and Conway, 2012.
- <sup>29</sup> Rocco MV, Frankenfield DL, Hopson SD, et al. “Relationship between Clinical Performance Measures and Outcomes among Patients Receiving Long-Term Hemodialysis.” *Annals of Internal Medicine*, 145(7): 512-519, 2006.
- <sup>30</sup> Reid RO, Deb, P. Howell BL and Shrank WH. “Association Between Medicare Advantage Plan Star Ratings and Enrollment.” *Journal of the American Medical Association* 309(3):267-274, 2013.
- <sup>31</sup> Doyle M and Pegler E. “Medicare Advantage Star Ratings: Where Do We Go From Here?” Boston, MA: Health Dialogue, 2010, [http://www.healthdialog.com/Libraries/Research\\_Documents/Medicare\\_Advantage\\_Star\\_Ratings.sflb.ashx](http://www.healthdialog.com/Libraries/Research_Documents/Medicare_Advantage_Star_Ratings.sflb.ashx) (accessed April 2013).
- <sup>32</sup> Medicare Payment Advisory Commission (MedPAC). “Outlier Alternative.” In *Report to the Congress: Assessing Alternatives to the Sustainable Growth Rate System*. Washington: MedPAC, 2007, [www.medpac.gov/documents/Mar07\\_SGR\\_mandated\\_report.pdf](http://www.medpac.gov/documents/Mar07_SGR_mandated_report.pdf) (accessed April 2013).
- <sup>33</sup> Ryan, 2009; Jha AK, Joynt KE, Orav EJ, et al. “The Long-Term Effect of Premier Pay for Performance on Patient Outcomes.” *New England Journal of Medicine*, 366(17): 1606-1615, 2012.
- <sup>34</sup> Epstein AM. “Will Pay for Performance Improve the Quality of Care? The Answer Is in the Details.” *New England Journal of Medicine*, 367: 1852-1853, 2012.
- <sup>35</sup> Epstein, 2012.
- <sup>36</sup> Meyer MW. “Measuring Performance in Economic Organizations.” In *The Handbook of Economic Sociology*, Smelser NJ and Swedberg R (eds). Princeton, NJ: Princeton University Press, 1994.
- <sup>37</sup> Gates, B, “Bill Gates: A Fairer Way to Evaluate Teachers,” *Washington Post*, April 3, 2013, [www.washingtonpost.com/opinions/bill-gates-a-fairer-way-to-evaluate-teachers/2013/04/03/c99fd1bc-98c2-11e2-814b-063623d80a60\\_story.html](http://www.washingtonpost.com/opinions/bill-gates-a-fairer-way-to-evaluate-teachers/2013/04/03/c99fd1bc-98c2-11e2-814b-063623d80a60_story.html) (accessed April 2013).
- <sup>38</sup> Milgrom P and Roberts J. *Economics, Organization, and Management*. Englewood Cliffs, NJ: Prentice-Hall, 1992.
- <sup>39</sup> Casalino LP. “The Unintended Consequences of Measuring Quality on the Quality of Medical Care.” *New England Journal of Medicine*, 341(15): 1147-1150, 1999.
- <sup>40</sup> Krumholz HM, Bradley EH, Nallamothu BK, et al. “A Campaign to Improve the Timeliness of Primary Percutaneous Coronary Intervention.” *Journal of the American College of Cardiology: Cardiovascular Interventions*, 1(1): 97-104, 2008; Waters HR, Korn R Jr, Colantuoni E, et al., “The Business Case for Quality Economic Analysis of the Michigan Keystone Patient Safety Program in ICUs.” *American Journal of Medical Quality*, 26(5):333-339, 2011.

- <sup>41</sup> Woolhandler S and Ariely D. “Will Pay for Performance Backfire: Insights from Behavioral Economics.” *Health Affairs Blog*, October 11, 2011, <http://healthaffairs.org/blog/2012/10/11/will-pay-for-performance-backfire-insights-from-behavioral-economics/> (accessed April 2013).
- <sup>42</sup> Cassel CK and Jain SH. “Assessing Individual Physician Performance: Does Measurement Suppress Motivation.” *Journal of the American Medical Association* 307(24): 2595-2596, 2012.
- <sup>43</sup> Woolhandler and Ariely
- <sup>44</sup> Pink DH. *To Sell Is Human: The Surprising Truth About Moving Others*. New York: Riverhead Books, 2012.
- <sup>45</sup> Birkmeyer JB, Dimick JD and Staiger DO. “Operative Mortality and Procedure Volume as Predictors of Subsequent Hospital Performance.” *Annals of Surgery*, 243(3):411-417, 2006; Birkmeyer JB, Stukel TA, Siewers AE, et al. “Surgeon Volume and Operative Mortality in the United States.” *New England Journal of Medicine*, 349:2117-2127, 2003.
- <sup>46</sup> *Farewell to Fee For Service? A Real World Strategy for Health Care Payment Reform*. Working Paper. Minnetonka, MN: UnitedHealth Center for Health Reform and Modernization, 2012.
- <sup>47</sup> Chien AT, Wroblewski K, Damberg C, et al. “Do Physician Organizations Located in Lower Socioeconomic Status Areas Score Lower on Pay-for-Performance Measures.” *Journal of General Internal Medicine*, 27(5): 548-554, 2011; Franks P, Fiscella K, Beckett L, et al. “Effects of Patient and Physician Practice Socioeconomic Status on the Health Care of Privately Insured Managed Care Patients.” *Medical Care*, 41(7): 842-852, 2003; Pham HH, Schrag D, Hargraves JL, et al. “Delivery of Preventive Services to Older Adults by Primary Care Physicians.” *Journal of the American Medical Association*, 294(4): 473-481, 2005.
- <sup>48</sup> Bradley EH, Holmbee ES, Mattera JA, et al. “A Quantitative Study on Increasing B Blocker Use After Myocardial Infarction: Why Do Some Hospitals Succeed?” *Journal of the American Medical Association*, 285(20): 2604-2611, 2001; Bradley EH, Herrin J, Elbel B, et al. “Hospital Quality for Acute Myocardial Infarction: Correlation Among Process Measures and Relationship with Short-term Mortality.” *Journal of the American Medical Association*, 296(1): 72-78, 2006; Bradley EH, Curry LA, Spatz ES, et al. “Hospital Strategies for Reducing Risk Standardized Mortality Rates in Acute Myocardial Infarction.” *Annals of Internal Medicine*, 156(9):618-625, 2012.
- <sup>49</sup> Glance LG, Newman M, Martinez EA, et al. “Performance Measurement at a ‘Tipping Point.’” *Anesthesia and Analgesia*, 112(4):958-966, 2011.
- <sup>50</sup> Bishop T. “Pushing the Outpatient Quality Envelope.” *Journal of the American Medical Association*, 309(13):1353-1354, 2013.
- <sup>51</sup> Pronovost, Miller and Wachter, 2007.
- <sup>52</sup> Wachter RM. “Why Diagnostic Errors Don’t Get Any Respect—And What Can Be Done about Them.” *Health Affairs (Millwood)*, 29(9):1605–10, 2010; Newman-Toker DE and Pronovost PJ. “Diagnostic Errors—The Next Frontier for Patient Safety.” *Journal of the American Medical Association*, 301(10):1060–1062, 2009; Schiff GD and Bates DW. “Can Electronic Clinical Documentation Help Prevent Diagnostic Errors?” *New England Journal of Medicine*, 362(12): 1066–1069, 2012; Singh H, Davis Giardina T, Meyer A, et al. “Types and Origins of Diagnostic Errors in Primary Care Settings.” *JAMA Internal Medicine*, 173(6): 418-425, 2013.
- <sup>53</sup> Zeller JL and Fontanarosa PB. “Shared Accountability, Appropriateness, and Quality of Surgical Care.” *Journal of the American Medical Association*, 302(14):1590–1591, 2009; Korenstein D, Falk R, Howell E, et al. “Overuse of Health Care Services in the United States: An Understudied Problem.” *Archives of Internal Medicine*, 172(2): 171-178, 2012.
- <sup>54</sup> Wachter RM. “Expected and Unanticipated Consequences of the Quality and Information Technology Revolutions.” *Journal of the American Medical Association*, 295(23): 2780–2783, 2006; Boyd CM, Darer J, Boulton C, et al. “Clinical Practice Guidelines and Quality of Care for Older Patients With Multiple Comorbid Diseases: Implications for Pay for Performance.” *Journal of the American Medical Association*, 294(6):716-724, 2005.
- <sup>55</sup> Nicolas LH, Osborne NH, Birkmeyer JD, et al. “Hospital Process Compliance and Surgical Outcomes in Medicare Beneficiaries.” *Archives of Surgery*, 145(10): 999-1004, 2010.
- <sup>56</sup> Werner and Bradlow, 2006.
- <sup>57</sup> Bernheim SM, Wang Y, Bradley EH, et al. “Who is Missing from the Measures? Trends in the Proportion and Treatment of Patients Potentially Excluded from Publicly Reported Quality Measures.” *American Heart Journal*, 160(5): 943-950, 2010.
- <sup>58</sup> Krumholz HM, Normand SL, Spertus JA, et al. “Measuring Performance for Treating Heart Attacks and Heart Failure: The Case for Outcomes Measurement.” *Health Affairs*, 26(1): 75-85, 2007.

- <sup>59</sup> Bradley et al., 2012; Schwartz M, Cohen AB, Restuccia JD, et al. "How Well Can We Identify the High Performing Hospital?" *Medical Care Research and Review*, 68(3): 290-310, 2011; Werner RM, Bradlow ET and Asch DA. "Does Hospital Performance on Process Measures Directly Measure High Quality Care or Is It a Marker of Unmeasured Care?" *Health Services Research* 43(5 Pt 1): 1464-1484, 2008.
- <sup>60</sup> Chassin et al., 2010.
- <sup>61</sup> Pronovost and Lilford, 2011; Krumholz et al., 2007.
- <sup>62</sup> Krumholz et al., 2007.
- <sup>63</sup> Pronovost PJ and Freischlag JA. "Improving Teamwork to Reduce Surgical Mortality." *Journal of the American Medical Association*, 304(15): 1721-1722, 2010.
- <sup>64</sup> Glance et al., 2011.
- <sup>65</sup> Iezzoni LI. "Reasons for Risk Adjustment." In *Risk Adjustment for Measuring Health Care Outcomes*. Iezzoni LI (ed). Chicago: Health Administration Press, 2003.
- <sup>66</sup> Iezzoni LI. "The Risks of Risk Adjustment." *Journal of the American Medical Association*, 278(19):1600-1607, 1997.
- <sup>67</sup> Ash AS, Fienberg SE, Louis TA, et al. *Statistical Issues in Assessing Hospital Performance*. Baltimore: U.S. Centers for Medicare and Medicaid Services, 2012, [www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Downloads/Statistical-Issues-in-Assessing-Hospital-Performance.pdf](http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Downloads/Statistical-Issues-in-Assessing-Hospital-Performance.pdf) (accessed April 2013).
- <sup>68</sup> Ibid.
- <sup>69</sup> Pronovost and Lilford, 2011.
- <sup>70</sup> Horwitz L, Partovian C, Lin Z, et al. *Hospital-Wide (All-Condition) 30-Day Risk-Standardized Readmission Measure: DRAFT Measure Methodology Report*. Baltimore, MD: U.S. Centers for Medicare and Medicaid Services, 2011, [www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/downloads/MMSHospital-WideAll-ConditionReadmissionRate.pdf](http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/downloads/MMSHospital-WideAll-ConditionReadmissionRate.pdf) (accessed April 2013).
- <sup>71</sup> National Quality Forum. *Guidance for Measure Testing and Evaluating Scientific Acceptability of Measure Properties*. Washington, DC: National Quality Forum, January 2011, [www.qualityforum.org/Publications/2011/01/Measure\\_Testing\\_Task\\_Force.aspx](http://www.qualityforum.org/Publications/2011/01/Measure_Testing_Task_Force.aspx) (accessed April 2013).
- <sup>72</sup> Glance et al., 2011.
- <sup>73</sup> Farmer SA, Black B and Bonow RO. "Tension Between Quality Measurement, Public Quality Reporting, and Pay for Performance." *Journal of the American Medical Association*, 309(4):349-350, 2013; Lee GM, Kleinman K, Soumerai SB, et al. "Effect of Nonpayment for Preventable Infections in U.S. Hospitals." *New England Journal of Medicine*, 367(15):1428-1437, 2012.
- <sup>74</sup> Lawthers AG, McCarthy EP, Davis RB, et al. "Identification of In-Hospital Complications From Claims Data: Is It Valid?" *Medical Care*, 38(8):785-795, 2000.
- <sup>75</sup> Haut ER and Pronovost PJ. "Surveillance Bias in Outcomes Reporting." *Journal of the American Medical Association*, 305(23): 2462-2463, 2011.
- <sup>76</sup> Pierce CA, Haut ER, Kardooni S, et al. "Surveillance Bias and Deep Vein Thrombosis in the National Trauma Data Bank: The More We Look the More We Find." *Journal of Trauma*, 64(4): 932-937, 2008.
- <sup>77</sup> Pronovost, Miller and Wachter, 2007.
- <sup>78</sup> Farmer et al., 2013; *Introduction to Trigger Tools for Identifying Adverse Events*. Cambridge, MA: Institute for Healthcare Improvement, [www.ihf.org/knowledge/Pages/Tools/IntrotoTriggerToolsforIdentifyingAEs.aspx](http://www.ihf.org/knowledge/Pages/Tools/IntrotoTriggerToolsforIdentifyingAEs.aspx) (accessed April 2013).
- <sup>79</sup> Kern LM, Malhotra S, Barrón Y, et al. "Accuracy of Electronically Reported 'Meaningful Use' Clinical Quality Measures: A Cross-Sectional Study." *Annals of Internal Medicine*, 158(2):77-83, 2013.
- <sup>80</sup> Porter, 2010.

- <sup>81</sup> *Patient-Reported Outcomes in Performance Measurement*. Washington: National Quality Forum, [www.qualityforum.org/Projects/n-r/Patient-Reported\\_Outcomes/Patient-Reported\\_Outcomes.aspx](http://www.qualityforum.org/Projects/n-r/Patient-Reported_Outcomes/Patient-Reported_Outcomes.aspx) (accessed April 2013).
- <sup>82</sup> Jencks SF, Williams MV, Coleman EA. "Rehospitalizations Among patients in the Medicare Fee-for-Service Program." *New England Journal of Medicine*, 360:1418-1428, 2009 [Erratum, *New England Journal of Medicine*, 364:1582, 2011]; Chollet D, Barrett A and Lake T. *Reducing Hospital Readmissions in New York State: A Simulation Analysis of Alternative Payment Incentives*. Princeton, NJ: Mathematica Policy Research, 2011, <http://nyshealthfoundation.org/uploads/resources/reducing-hospital-readmissions-payment-incentives-september-2011.pdf> (accessed April 2013).
- <sup>83</sup> Bindman AB, Blum JD and Kronick R. "Medicare's Transitional Care Payment—A Step toward the Medical Home." *New England Journal of Medicine*, 368(8): 692-694, 2013.
- <sup>84</sup> Blum J. "Statement of Jonathan Blum, Acting Principal Deputy Administrator and Director, Center for Medicare, Centers for Medicare & Medicaid Services, On Delivery System Reform: Progress Report from CMS, Before the U.S. Senate Finance Committee." Feb. 28, 2013, [www.finance.senate.gov/imo/media/doc/CMS%20Delivery%20System%20Reform%20Testimony%202.28.13%20\(J.%20Blum\).pdf](http://www.finance.senate.gov/imo/media/doc/CMS%20Delivery%20System%20Reform%20Testimony%202.28.13%20(J.%20Blum).pdf) (accessed April 2013).
- <sup>85</sup> Abelson R, "Hospitals Question Medicare Rules on Readmissions," *New York Times*, March 29, 2013.
- <sup>86</sup> Brock J, Mitchell J, Irby K, et al. "Association Between Quality Improvement for Care Transitions in Communities and Rehospitalizations Among Medicare Beneficiaries." *Journal of the American Medical Association*, 309(4): 381-391, 2013.
- <sup>87</sup> Berenson RA, Paulus RA and Kalman NS. "Medicare's Readmissions-Reduction Program—A Positive Alternative." *New England Journal of Medicine*, 366: 1364-1366, 2012.
- <sup>88</sup> Abelson, 2013.
- <sup>89</sup> Pronovost PJ and Hudson DW. "Improving Healthcare Quality Through Organisational Peer-to-Peer Assessment: Lessons from the Nuclear Power Industry." *BMJ Quality and Safety*, 21(10):872-875, 2012.
- <sup>90</sup> *About the Partnership for Patients*. Baltimore, MD: Centers for Medicare and Medicaid Services, <http://partnershipforpatients.cms.gov/about-the-partnership/aboutthepartnershipforpatients.html> (accessed April 2013).
- <sup>91</sup> *Partnership for Patients*. Baltimore, MD: Centers for Medicare and Medicaid Services, <http://partnershipforpatients.cms.gov/about-the-partnership/what-is-the-partnership-about/what-the-partnership-is-about.html> (accessed April 2013).
- <sup>92</sup> *2011 National and State Healthcare-associated Infections Standardized Infection Ratio Report*. Atlanta, GA: Centers for Disease Control and Prevention, [www.cdc.gov/hai/national-annual-sir/index.html](http://www.cdc.gov/hai/national-annual-sir/index.html) (accessed April 2013).
- <sup>93</sup> Birkmeyer NJO and Birkmeyer JD. "Strategies for Improving Surgical Quality—Should Payers Reward Excellence or Effort?" *New England Journal of Medicine*, 354(8): 864-870, 2006.
- <sup>94</sup> *Roles of Regional Health Improvement Collaboratives: Measuring Healthcare Performance*. Pittsburgh, PA: Network for Regional Healthcare Improvement, [www.nrhi.org/performancemeasurement.html](http://www.nrhi.org/performancemeasurement.html) (accessed April 2013).
- <sup>95</sup> Ferris TG, Vogeli C, Marder J, et al. "Physician Specialty Societies And The Development Of Physician Performance Measures." *Health Affairs*, 26(6): 1712-1719, 2007.
- <sup>96</sup> Iglehart JK and Baron RB. "Ensuring Physicians' Competence — Is Maintenance of Certification the Answer?" *New England Journal of Medicine*, 367: 2543-2549, 2012.
- <sup>97</sup> Pronovost PJ, Jenckes MW and Dorman T. "Organizational Characteristics of Intensive Care Units Related to Outcomes of Abdominal Aortic Surgery." *Journal of the American Medical Association*, 281(14): 1310-1317, 1999.
- <sup>98</sup> Pham HH, Schrag D, O'Malley AS, et al. "Care Patterns in Medicare and Their Implications for Pay for Performance." *New England Journal of Medicine*, 356: 1130-1139, 2007.
- <sup>99</sup> Allina Health, Atrius Health, Banner Health Network, et al. "Letter to Rick Gilfillan and Hoangmai Pham, Center for Medicare and Medicaid Innovation." Feb. 25, 2013, [www.washingtonpost.com/blogs/wonkblog/files/2013/03/2013-Quality-Benchmarks.pdf](http://www.washingtonpost.com/blogs/wonkblog/files/2013/03/2013-Quality-Benchmarks.pdf) (accessed April 2013).
- <sup>100</sup> Manary MP, Boulding , Staelin R, et al., "The Patient Experience and Health Outcomes." *New England Journal of Medicine*, 368:201-203, 2013, [http://podcast.nejm.org/nejm\\_audio\\_interview.xml](http://podcast.nejm.org/nejm_audio_interview.xml) (accessed April 2013).

- 
- <sup>101</sup> Sorian R. *Measuring, Reporting, and Rewarding Performance in Health Care*. New York, NY: The Commonwealth Fund, 2006, [www.commonwealthfund.org/~media/Files/Publications/Fund%20Report/2006/Mar/Measuring%20%20Reporting%20%20and%20Rewarding%20Performance%20in%20Health%20Care/Sorian\\_measuringreporting\\_912%20pdf.pdf](http://www.commonwealthfund.org/~media/Files/Publications/Fund%20Report/2006/Mar/Measuring%20%20Reporting%20%20and%20Rewarding%20Performance%20in%20Health%20Care/Sorian_measuringreporting_912%20pdf.pdf) (accessed April 2013); Fenton JJ, Jerant AF, Bertakis KD, et al. “The Cost of Satisfaction: A National Study of Patient Satisfaction, Health Care Utilization, Expenditures, and Mortality.” *Archives of Internal Medicine*, 172(5): 405-411, 2012.
- <sup>102</sup> Clancy C, Brach C and Abrams M (eds). “Assessing Patient Experiences of Providers’ Cultural Competence and Health Literacy Practices: CAHPS® Item Sets.” *Medical Care*, 50(9), Supplement 2: i–ii, S1–S79, 2012.
- <sup>103</sup> Pronovost and Hudson, 2012.
- <sup>104</sup> Casalino 1999; Curtis JR and Levy M. “Improving the Science and Politics of Quality Improvement.” *Journal of the American Medical Association*, 305(4): 406-407, 2011; Baker DW and Qaseem A. “Evidence–Based Performance Measures: Preventing Unintended Consequences of Quality Measurement.” *Annals of Internal Medicine*, 155(9): 638-640, 2011; Pogach L and Aron DC. “Sudden Acceleration of Diabetes Quality Measures.” *Journal of the American Medical Association*, 305(7): 709-710, 2011; Wachter, 2006.
- <sup>105</sup> Casalino LP, Elster A, Eisenberg A, et al. “Will Pay-For-Performance and Quality Reporting Affect Health Care Disparities?” *Health Affairs*, 26(3): w405-w414, 2007; Chien, et al, 2011.
- <sup>106</sup> *Overview of the SGR Repeal and Reform Proposal*. Washington: Majority Staff of the U.S. House of Representatives’ Ways and Means Committee and Energy and Commerce Committee, [http://waysandmeans.house.gov/uploadedfiles/sgr\\_reform\\_short\\_summary\\_2013.pdf](http://waysandmeans.house.gov/uploadedfiles/sgr_reform_short_summary_2013.pdf) (accessed April 2013).