Summary

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

In this issue paper, we give an overview of performance measurement in U.S. health care, and then make seven policy recommendations aimed at improving the performance measurement enterprise. Specifically, we recommend how to develop better measures, when and how to use measures, and how to ensure the validity and comparability of publicly-reported performance measure data.

Introduction

There is a consensus that scientifically rigorous and valid measurement of performance can be instrumental in improving value in U.S. health care. 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. 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, substantial shortcomings in the quality of U.S. health care persist. Furthermore, the growth of performance measurement has been accompanied by increasing concerns about 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.

The Quality Measurement Enterprise

A seminal event that kicked off efforts to bring performance measurement to health care 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 using standardized quality measures aimed at understanding whether standard practice across the health care system could be improved. What was novel about this shift was the focus on explicit, objective criteria rather than implicit, subjective expert opinions, and the new goal of encouraging all health care providers to improve their performance, rather than just the poorest performers.

Today, 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 health care 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. The measures they develop are predominantly used by payers in public reporting and provider incentive programs. For example, Medicare initially produced and used measures to provide consumers with information to help them select health care providers and insurance plans, but has more recently begun using measures to financially reward providers who are able to deliver better-quality care to beneficiaries at a lower cost through a suite of new “value-based purchasing” programs.

Meanwhile, numerous for-profit companies, including Healthgrades and U.S. News and World Report, have developed their own proprietary measures, and use them to grade hospitals and doctors. These information brokers use measures not endorsed by the National Quality Forum (NQF, a multi-stakeholder organization that reviews and endorses quality...
measures), and do not always disclose the methods by which they rank providers. A number of researchers have questioned the validity and reliability of such proprietary “report cards”; in the absence of transparent measurement standards, the correlation among them is low. For example, recently none of the 17 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. Proprietary ranking systems likely confuse more than clarify.

**Types of measures.** Quality measures are typically thought of as falling into one of three types. *Structural* measures identify the presence of particular characteristics associated with a health care setting (e.g., whether a practice is made up of board-certified physicians); *process* measures assess what is done in the health care setting (e.g., the percentage of the time a recommended screening is performed); and *outcome* measures assess the ultimate results achieved for a patient after a given set of interventions (e.g., the percentage of patients who survived a particular surgery).

So-called “intermediate” or “surrogate” outcome measures are those that, while not true outcomes, are used as proxies for particular patient outcomes. For example, results of hemoglobin A1C blood tests are used as an indicator of whether diabetes is under control, because the test’s results correlate with the likelihood of experiencing diabetes complications.

Increasingly, quality experts also consider various aspects of patients’ reported experiences as important outcome measures, whether or not better experiences are associated with improved clinical outcomes.

**Data sources.** In general, the data needed to calculate performance measures are collected through three sources: administrative data, medical records, and patient surveys. Administrative data (e.g., insurance claims and enrollment files) are relatively easy and inexpensive to collect, but lack the clinical detail needed to generate many desired measures.

Meanwhile, medical record data provide substantially more detail about the care being provided, though the quality of this data is variable across different practices and organizations; medical record data are also substantially 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, though the widespread adoption of electronic health records (EHRs) should lower the cost of using this type of data in the future. Survey data are typically collected for the purpose of measuring patient experiences with care.

**How performance measure data are used.** In the U.S., performance measure data are predominantly used in public reporting and provider incentive programs, though health care organizations also use these data for internal quality improvement efforts. Commercial health plans often publicly report provider performance, and sometimes combine quality measurement data with price and cost information to attempt to categorize providers into different value tiers, such that plan members face lower cost-sharing when selecting providers in favored tiers.

Measures can also be used internally by a practice or facility for quality improvement purposes, and can be generated quickly by merely running a query in an EHR. When used for internal quality improvement purposes, measures need not be held to the same standards as publicly reported data. However, many internally-generated measures are starting to be reported on hospital websites and in marketing materials—often without sufficient information to determine their accuracy.

Until recently, Medicare used performance measure data primarily to provide information to consumers to help them select providers and health plans. However, CMS has begun offering a number of congressionally mandated pay-for-performance initiatives in recent years that provide direct financial rewards or penalties based on performance on quality measures. These include a suite of new programs that reward providers who deliver better performance for beneficiaries at lower cost, including the End-Stage Renal Disease (ESRD) Bundled-Payment and Quality Incentive Program, performance bonuses for Medicare Advantage plans based on star ratings, the Hospital Value-based Purchasing Program, and the Physician Value-based Payment Modifier.

Unfortunately, a number of studies suggest that pay-for-performance programs do not significantly improve quality, at least as implemented thus far. For example, the hospitals in the Medicare Premier Hospital Quality Incentive Demonstration did not actually produce better results than other hospitals. Similarly, tests of pay-for-performance for physicians in the United States and the United Kingdom show mixed results, at best showing small, sometimes temporary improvements in quality.

Behavioral economics offers some insights into why, despite intuitive appeal, pay-for-performance has had
a limited impact on improving quality of care, especially of physicians. At root, economic incentives seek to change behavior through extrinsic motivation—yet most clinicians already want to deliver high-quality care based on an intrinsic motivation to act in their patient’s best interests. A danger of poorly-designed financial incentives is that they can “crowd out” positive intrinsic motivation to do the right thing for patients. And even when motivation is lacking, performance bonuses can backfire—particularly for cognitively challenging activities performed by highly skilled persons needing to muster their skills to manage complexity and solve problems creatively.

The problem with process measures. Reliance mostly on process measures to assess health care quality—as is currently done—presents several problems. First, 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. In addition, it is virtually impossible to measure diagnosis errors, and difficult to measure whether services are inappropriately overused. Second, even evidence-based process measures do not always predict improved outcomes. Recent research suggests that even the longstanding CMS process measures for heart failure, heart attack, and pneumonia did not predict overall, short-term mortality in the Premier demonstration. Third, process measures can lead to negative impacts on patient outcomes if hospitals or medical practices focus on meeting certain process measures while ignoring problems in other areas of care not being measured.

A fourth practical limitation in using process measures relates to the high cost of data collection; that limitation produces a heavy reliance on measuring laboratory tests and prescription drug use, which are readily measured using administrative data, thereby limiting the care processes that are used. Finally, updating of process measures 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 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.

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.

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

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

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, policymakers, and providers all care about the end results of care—not the technical approaches that providers may adopt to achieve desired outcomes. 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, while actionable, often explain little of the variation in outcomes that exemplifies U.S. health care. A promising avenue for advancing outcome measurement is greater use of patient-reported outcomes, which are derived using surveys that ask patients what they are able to do and how they feel.

2. Use quality measures strategically, adopting other quality improvement approaches where measures fall short.
CMS’ current value-based purchasing efforts require reporting on a raft of measures of varying usefulness and validity, which may divert providers’ attention from efforts to make culture and work process improvements that could produce larger improvements in outcomes. In their place, Congress should direct CMS to use performance measures more strategically to improve specific quality deficiencies—promoting collaborative quality improvement activities and new payment approaches that incorporate performance measures instead of public reporting and pay-for-performance per se. For example, the current focus on reducing preventable rehospitalizations within 30 days of discharge represents a timely, strategic use of an important performance measure to address a problem with proven approaches available to achieve significant improvement. 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 to reduce the likelihood of readmission. Other opportunities for integrating performance measurement into targeted quality improvement efforts should be explored to replace the current attempt to provide report cards assessing the performance of all providers, whether accurate or not.

3. Measure quality at the level of the organization, not the clinician.

Historically, the physician has been viewed as the leader of medicine, with responsibility for the care and outcomes of patients. It is therefore not surprising that some information brokers, including U.S. News and World Report and city magazines like the Washingtonian, provide ratings of “top doctors,” often based mostly on reputation, warranted or not. Yet the notion that an individual health professional can be held accountable for the outcomes of patients in isolation from other health professionals is becoming an outdated perspective, as systems-based care is emerging as a key value and vital component of high quality care. Other issues with attributing performance measure data to individual clinicians are more technical: many individual clinicians lack sufficient volumes of certain types of patients to permit valid statistical inferences about their performance on a given measure. To circumvent these difficulties, performance can be measured at the organizational or departmental level, allowing measures to assess and promote team-based care, while addressing many of the technical issues that can undermine measurement efforts.

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

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 health care dimensions. With the growing array of scientifically rigorous surveys of patient experience with care, we now have the capacity to incorporate standardized assessments of that experience into the measurement enterprise. This is especially important since there is marked heterogeneity in patient experience ratings, and the quality of providers’ attention to patients’ needs can influence health outcomes.

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

The dissemination of quality measure data should be viewed as one prong in a multi-pronged strategy to improve health care quality. Accompanying strategies should include offering technical assistance to strengthen providers’ capacity to improve care and creating formal accountability systems. In addition, collaborative activities among institutions can produce insights that may elude them individually: measures can help identify top performers, and detailed analysis can then identify what distinguishes those who excel.

6. Invest in the “basic science” of measurement development.

There is no expert body with responsibility for advancing the science of performance measurement. NQF comes the 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. An infrastructure is needed to gain national consensus on: what to measure, how to collect the data needed to calculate measures, the accuracy of EHR data for use in performance measurement, how to measure the cost-effectiveness of particular measures, how to reduce the costs of data collection, what thresholds to use to ensure measure accuracy, and how to prioritize which measures to collect. Establishing general standards for performance measures could help move the policy discussion from whether measures are good enough to use despite their flaws to a more fundamental discussion of how to design good measures, how to
assess current measures, and whether the costs of producing better measures are worth the benefits.

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.

There is a plethora of types of health care quality data being disseminated to the public, yet no rules to assure the accuracy of what is being presented. 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 National Quality Forum 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. As a result of this lack of standards, payers, policy-makers, and providers often become embroiled in a tug-of-war—with payers and policymakers asserting that existing measures are good enough, and providers arguing that they are not. Often, neither side has data on how good the contested measures actually are. Indeed, most quality measurement efforts struggle to find measures that are scientifically sound yet feasible to implement with the limited resources available; too often, feasibility trumps sound science. The result is that the focus of health care organizations may change from improving care to looking good.

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, all supervised by an entity modeled after the SEC. Policy-makers 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 new government entity. Such a commission could promote standardization, transparency, and auditing of the reporting of quality and cost measures. We do not recommend giving this entity regulatory authority to require adherence to its standards; instead, we anticipate that organizations would voluntarily seek to comply with the applicable standards for reporting performance measures. 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 performance 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. Follow the Foundation on Twitter www.rwjf.org/twitter or Facebook www.rwjf.org/facebook.