

Fostering Transparency in Outcomes, Quality, Safety, and Costs

A Vital Direction for Health and Health Care

Peter J. Pronovost, Johns Hopkins Medicine; **J. Matthew Austin**, Johns Hopkins Medicine; **Christine K. Cassel**, Kaiser Permanente School of Medicine; **Suzanne F. Delbanco**, Catalyst for Payment Reform; **Ashish K. Jha**, Harvard T.H. Chan School of Public Health; **Bob Kocher**, Venrock; **Elizabeth A. McGlynn**, Kaiser Permanente; **Lewis G. Sandy**, UnitedHealth Group; **John Santa**, formerly of Consumer Reports

September 19, 2016

About the Vital Directions for Health and Health Care Series



This publication is part of the National Academy of Medicine's **Vital Directions for Health and Health Care Initiative**, which called on more than 150 leading researchers, scientists, and policy makers from across the United States to assess and provide expert guidance on 19 priority issues for U.S. health policy. The views presented in this publication and others in the series are those of the authors and do not represent formal consensus positions of the NAM, the National Academies of Sciences, Engineering, and Medicine, or the authors' organizations. Learn more: nam.edu/VitalDirections.

Introduction

Over the last 20 years, the United States has witnessed a shift from little readily available information about the performance of the health care system to the use of a wide variety of measures in different ways by multiple entities (Cronin et al., 2011). The explosion of performance measures and the public reporting of performance have served important functions in raising awareness of deficits in quality and stimulating efforts to close measured gaps (O'Neil et al., 2010). Despite the important gains, serious concerns have been raised about the value of performance measurement in its current state, including the validity and reliability of measures, the burden and complexity of measuring performance, substantial gaps in measuring important aspects of care, and limited evidence regarding the fundamental premise that measurement and

reporting drive improvement. The purposes of this paper are to identify the requirements of a valid and useful performance-measurement and performance-reporting system and to suggest a pathway to a better system. The timing of this paper is important inasmuch as the recent goal of moving away from rewarding volume to rewarding value depends on having valid and accurate measures so that the quality of care being delivered can be known and improved.

Transparent reporting of the performance of the health care system is often promoted as a key tool for improving the value of health care by improving quality and lowering costs, although the evidence of its effectiveness in achieving higher quality or lower costs is mixed (Austin and Pronovost, 2016; Hibbard et al., 2005; Totten et al., 2012; Whaley et al., 2014). Transparency can improve value by two key pathways: engaging

providers to improve their performance and informing consumer choice (Berwick et al., 2003). With respect to engaging providers, transparency can catalyze improvement efforts by appealing to the professionalism of physicians and nurses and by stimulating competition among them and their organizations (Lamb et al., 2013). With respect to informing consumer choice, public reporting can provide patients, payers, and purchasers with information about performance and enable preferential selection of higher-quality providers, lower-cost providers, or providers that demonstrate both characteristics. Although the potential for informing consumer choice exists, there is limited evidence to support the idea that consumers are using public reports in their current form to make better decisions (Faber et al., 2009; Shaller et al., 2014). We have pockets of success in public reporting that drive improved performance (Ketelaar et al., 2011), including the reporting of the Society of Thoracic Surgeons (STS) registries in cardiac surgery (Shahian et al., 2011a, b; Stey et al., 2015; STS, 2016); the Centers for Disease Control and Prevention's (CDC's) measures of health care-associated infections (CMS, 2016a; Pronovost et al., 2011); measures of diabetes-care processes, intermediate outcomes, and complications (Smith et al., 2012); and the Agency for Healthcare Research and Quality's Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) measure (CMS, 2016b; Elliott et al., 2010). Despite those successes, we have fallen short of the full potential of understanding the performance of the health care system; for example, only a minority of heart-surgery groups have voluntarily reported their performance from the STS registry, and cardiologists do not appear to refer patients to cardiac surgeons who have the best outcomes (Brown et al., 2013).

Health care organizations and providers frequently invest time and energy to improve their performance on reported measures and we should ensure that they are acting on valid information (Winters et al., 2016). That holds true for all types of measures—measures of outcomes (both clinically oriented and patient defined), quality, safety, and costs. The accurate measurement and reporting of health care system performance is important for all stakeholders. Patients, clinicians, payers, and purchasers need measures of absolute and relative performance to facilitate informed choice of providers, innovative benefit designs and provider

networks, and alternative payment methods that support quality improvement and greater affordability (Damberg et al., 2011). With transparency of performance results, markets are able to work more effectively; this enables higher-quality providers to attract greater market share, assuming that the incremental revenue to be gained from additional market share is financially beneficial to them. Physicians and hospitals need measures to make treatment decisions and to identify strengths and weaknesses so that they can focus their quality-improvement and performance-improvement activities and monitor progress (Benson and Rice, 2015). Transparency of performance facilitates identification of exemplary performers, who might in turn be emulated by others and encourage learning (Dixon-Woods et al., 2011).

Although transparency is beneficial, it poses risks if the results being shared are not valid (Adams et al., 2010; Austin, 2015; McGlynn and Adams, 2014). There is no standard for how reliable and valid a measure should be before it is publicly reported. Publicly reporting a measure whose reliability and validity are unknown poses risks, including disengaging clinicians from improvement work, and raises potential ethical concerns, such as imposing unjust financial and reputation harm on physicians and provider organizations, misinforming patients about the risks and benefits associated with a treatment option, and guiding patients to riskier rather than to safer care (Winters et al., 2016).

Medicine is based on science, but the science of health care delivery, its measurement, and how to improve it is immature (Marjoua and Bozic, 2012). There are insufficient studies, little research investment, and a lack of agreement on the best way to measure how well health care providers deliver their services (Dixon-Woods et al., 2012). The growth in measurement stems from a wide array of entities' development and use of measures and methods to assess performance, including accreditation organizations (such as the National Committee for Quality Assurance and The Joint Commission), the Centers for Medicare & Medicaid Services (CMS), state Medicaid programs, commercial health plans, consumer review platforms (such as Yelp), and independent parties, ranging from nonprofits to for-profit entities (such as HealthGrades and US News and World Report) (Jha, 2012). The variety of measures and methods and the lack of standards for measures and auditing of data have led to conflicting results in data

on quality, safety, patient experience, and cost (for example, a large proportion of hospitals are rated as top performers by at least one rating program), which potentially confuse those who want to use the data or encourage them to ignore the results altogether because they are incoherent or inconsistent (Austin et al., 2015).

The variety of measures and methods used to measure performance could be a product of different underlying hypotheses and biases (Shwartz et al., 2015). For example, Consumer Reports and the Leapfrog Group both issue patient-safety composites for hospitals (Consumer Reports, 2016; HSS, 2016). The two organizations have chosen to define safety differently: Leapfrog defines safety as “freedom from harm,” and Consumer Reports refers to “a hospital’s commitment to the safety of its patients.” The two organizations have chosen to include different measures in their composites to reflect their chosen definition of the construct (Austin et al., 2015). In this example, both organizations are fully transparent in their methods and underlying constructs, but most often the underlying hypotheses and biases are not transparent, and few are tested. When the data-collection and analytic processes are fully transparent, a robust scientific measurement process is possible. When the underlying hypotheses, assumptions, and biases of measurement methods are not transparent, confusion and misinformation can result.

Key Issues, Cost Implications, and Barriers to Progress

Key Issue 1: The Process of Measuring and Reporting on the Health Care System’s Performance is Error Prone and Lacks Standards

The variation in reports about the quality of care can be a function of true variation in quality, of the quality of the underlying data, of the mix of patients cared for by the provider, of bias in the performance measure, and of the amount of systemic or random error (Parker et al., 2012). Data used for performance measurement are often first developed for a different purpose, such as billing or meeting regulatory requirements. If the data were generated for a different purpose, it would not be surprising if they were problematic for “off-label” use (Lau et al., 2015).

There are four key steps in measuring and reporting health care system performance, with an opportunity for error in each step, different entities

involved in each step, and no entity entrusted with ensuring the validity of the entire process (Austin et al., 2014). The first step of the process is developing and specifying the performance measure. Developing the measure includes deciding what dimension of care is to be measured; when done well, it requires thinking about whether the dimension is a key aspect of care delivery, what evidence supports focusing on that dimension, and the likelihood that existing sources of data can be used to measure the dimension (McGlynn, 2003). Specifying the performance measure includes identifying the measure’s population of interest, the outcome or process of interest, and, if appropriate, the model for risk adjustment. Entities involved in measure development include measure developers and professional societies. The National Quality Forum (NQF) uses a multistakeholder consensus-development process to vet performance measures and endorses the ones that meet the criteria of importance, scientific acceptability, feasibility, and usability (NQF, 2016a). Although that process has helped to improve measures, the criteria are not evaluated in a strict quantitative sense. The NQF does not define specific validity tests for different types of measures, report a measure’s validity and reliability, or define specific thresholds for validity and reliability for endorsement. For example, the NQF endorsed the Patient Safety Indicator-90 (PSI-90) measure, for which the measure developer conducted construct validity testing by examining the association between the composite performance score and hospital structural characteristics potentially associated with quality of care (Owens, 2014). A complementary, and perhaps stronger, approach for demonstrating the construct validity of the score, which is based on administrative data, is to compare the positive predictive value (PPV) of the administrative data with the medical chart. A recent study that examined that approach found that none of 21 PSIs met a PPV threshold of 80%; the validity of most of the individual component measures that make up the PSI-90 composite was low or unknown (Winters et al., 2016). In addition, those who measure and report health care performance do not have to use NQF-endorsed measures. How do we ensure the validity and reliability of all performance measures used to hold the health care system accountable? How do we make transparent how “good” the measure is? Is the measure “fit for purpose”? That is, can it be applied as the user intends it to be?

The second step of the process is identifying and collecting the data used to populate measures and ensuring that the data are accurate for the intended purpose. Entities involved in obtaining data for measurement include physicians, hospitals, survey vendors, health systems, and payers. With the exception of data from the National Committee for Quality Assurance (NCQA)—which measures health plans, some clinical registries, and a small number of state health departments that validate health care–associated infection data—few of the data used for performance measurement are subjected to systematic quality-assurance procedures that are specific to the intended use for measurement. Such procedures can include assessment of the extent of missing data or out-of-range values. Challenges include incomplete or fragmented data and providers or sites that differ from one another in coding or recording of data. More is known about variations in claims data because of its longer history of use than about variations commonly occurring in data from electronic medical records (EMRs); EMRs might become a more frequently used data source for quality measurement in the future. The recommendation of systematic quality assurance aligns with the 1998 President’s Advisory Commission on Consumer Protection and Quality in the Health Care Industry recommendation that “information on quality that is released to the general public to facilitate comparisons among health care organizations, providers, or practitioners should be externally audited by an independent entity” (AHRQ, 1998). How do we ensure that the data used to populate measures are “good enough”? And how good is “good enough”? NCQA has developed and implemented an auditing process for its Healthcare Effectiveness Data and Information Set (HEDIS) performance measures that could serve as a model for others (NCQA, 2016).

The third step is applying the data to the specified measure. This step is a common source of variation. Multiple sources of data are used for measuring quality, safety, and cost, including claims, medical records, and surveys. Each has strengths and weaknesses that must be considered in the context of a particular measure. Entities that conduct measurement often state that they are using a “standard, endorsed” measure, but in the measurement program there may be minor, or even major, deviations from the endorsed measure, differing interpretations of what the measure

specifications mean, different sources of data, and “adjustments” of standards for convenience or administrative simplification. For example, NCQA’s HEDIS measure of breast-cancer screening attributes patients to clinicians by including patients who had any enrollment, claim, or encounter with a given clinician in the denominator population. A state-based quality collaborative chose to narrow the denominator of this measure to include only patients who had a primary care visit with the measured clinician (NQF, 2016b). Such variation in how the measure is implemented probably means that the validity of the results is unknown and the results are possibly not comparable.

For measures that are publicly reported, the fourth step is creating the public report. Errors at each step in the process cascade and compound, potentially imparting significant biases in published reports. In addition, variations in reporting templates, levels of detail, graphics, and many other factors are sources of variation in look, feel, interpretability, and usability of information. Entities involved in creating reports include government, consumer groups, consumer-oriented Web sites, news-media organizations, health plans, purchasers, and providers. Approaches to categorizing and communicating results have undergone little systematic evaluation (Totten et al., 2012). For example, several researchers raised methodologic concerns about how the results of a recent (2015) ProPublica measure of surgeon quality were constructed and reported, inasmuch as performance categories were determined by using the shape of the distribution of adjusted surgeon complication rates for each procedure and the thresholds chosen did not reflect statistically significant differences from the mean (Friedberg et al., 2015). Questions that still need to be answered about the best way to report results include, Should differences in categories be statistically significant? Should the differences be clinically or practically significant? Are users able to interpret the display accurately? Should current performance be displayed in the context of a trend over time?

Key Issue 2: The Health Care Measurement and Reporting Enterprise Could Benefit from Standards and a Standard-Setting Organization

In light of opportunities for error in each step of the measurement and reporting process and tensions regarding the release of performance measures of

uncertain validity (such as CMS's overall quality "star" ratings for hospitals), standard-setting could stimulate improvements in the integrity of the underlying data and methods used to generate performance measures. One possible opportunity is to learn from financial reporting standards and emulate the Financial Accounting Standards Board (FASB). FASB establishes financial accounting and reporting standards for public and private companies and not-for-profit organizations that follow Generally Accepted Accounting Principles (Pronovost et al., 2007). FASB's mission is to establish accounting and reporting standards whose faithful implementation results in financial reports that provide useful and standard information to investors, creditors, and other providers of capital. FASB develops and issues its standards through a transparent and inclusive process. FASB originated in the early 1970s, when capital-market participants began to recognize the importance of an independent standard-setting process separate and distinct from accounting professionals, so that the development of standards would be insulated from the self-interests of practicing accountants and their clients (FASB, 2016). The standards developed by an "FASB for Health Care" would need to be informed by and to inform a number of stakeholder audiences. The idea of an "FASB for Health Care" has been discussed in many circles for a number of years; now may be the right time for its development. We spent the better part of the last 2 decades in bringing health care stakeholders along to the ideas of performance measurement and transparency of data. We have reached a shift in the health care environment in which measurement and transparency are now considered the "norm," and this allows us to set priorities for improving the robustness of these systems. In addition, with the current focus on paying for value instead of for volume, an idea that depends on valid performance measures, the need for a robust measurement and reporting process is more important than ever. We may have to settle for imperfect measures in the short term, but having standards for health care performance measures would make it possible to set thresholds for minimum performance of a measure before the measure is used or at least to understand, and make transparent, the imperfectness of the measure.

Key Issue 3: Further Research and Development in Health Care Performance Measurement and Reporting Are Needed

The process of measuring and reporting health care system performance could be thought of as a "system-level" intervention that needs to be studied for efficacy, effectiveness, cost effectiveness, and impact. Questions that still need to be answered include, What are the benefits of and unintended risks posed by public reporting, and do these vary by type of measure? For which conditions and types of patients is public reporting useful or not useful? What do we know about the types of public-reporting tools that are useful for different stakeholders and about how and when they should be used? When is a measure so biased that it risks doing more harm than good? How do we improve our understanding of how consumers make decisions, given that many consumers already assume care to be of high quality and safe (Hibbard and Soafer, 2010)? Is the cost of measuring and reporting on quality for an area justifiable? What attributes of measures engage clinicians to improve?

We need to improve measures and reduce the burden and costs of measurement. We need to produce measures that are useful to patients, particularly measures for conditions that are important to patients; outcomes that matter to patients, such as functional status; and measures of the overall value of the care delivered. The Patient-Centered Outcomes Research Institute (PCORI) has funded some early work in developing measures that are important to patients, but many gaps remain. It is critical to ensure that measures are understandable, impactful, and actionable and that they align with skills and abilities of those who need to use the information. In addition to focusing on new measures, we need to retire measures of low validity, low utility, or low engagement so as to reduce measurement burden. Public and private stakeholders have made little investment in advancing the science of and innovation in performance measurement, and no single entity is responsible for coordinating this work. We lack adequate investment in the "basic science" of measurement development. We lack a safe space for innovation in improving measures (McGlynn and Kerr, 2016), including iterating upon measures between their endorsement by NQF and their use in public reporting and payment. We lack incentives to become a learning health care system that is informed by valid

and timely data and is focused on improving. We lack incentives for payers to share their price and quality data with external parties; such sharing might reduce the perceived value of their network discounts. And we lack incentives for consumers to use performance measures inasmuch as out-of-pocket maximums make cost data irrelevant for most care and consumers' inherent trust in the quality of care provided by their doctors and other health care providers may make quality data feel irrelevant (Hays and Ware, 1986). Payers and health care purchasers should continue their efforts to engage patients with these data because they benefit when patients seek higher-quality care, lower-cost care, or both. For a \$3 trillion health care system, the costs of investing in a more robust performance-measurement and performance-reporting system, to ensure that we accurately capture and report performance, would constitute a tiny fraction of the total expense and likely reduce costs in the long term.

Opportunities for Progress and Policy Implications

The debate about performance measures has not always been grounded in scientific evidence. Some argue that current measures are good enough, others argue that they are not, and neither side offers evidence on how valid the measures are, how we might make them better, what it might cost to do so, and how valid they need to be. If the health care system is to realize the potential of publicly reporting performance measures, the users and producers of such measures will need to collaborate and gain consensus on those and other key issues.

We can enhance the effectiveness of performance measurement in a number of ways. Transparency of both content and process are foundational for trust and understanding. We need a coordinated policy to fund, set standards for, and support research and innovation in performance measurement in health care just as in the reporting of financial data. As part of a continuously "learning" health care system, in which we constantly assess performance and learn from experience, we need to implement "feedback loops" to understand how to improve the usefulness of measures and to discern unintended adverse or weak effects so that we can create systems by which producers and consumers of measures collaborate, pilot-test, iterate,

and ensure the quality and continuous improvement of the entire measurement process (IOM, 2013). The feedback loops can occur at local, regional, and national levels. If lessons are systematically collected and shared, they can serve as a tool for improvement. We need "learning" or innovation laboratories with consumers and producers of measures to explore ways to make measures more useful and less burdensome (McGlynn and Kerr, 2016).

We need better communication with patients to raise their awareness of variation in the quality and costs of care. Few consumers are aware of the variation in quality and costs of care and how they can obtain information about them (O'Sullivan, 2015). For example, in one survey of patients with chronic conditions, only 16–25% of consumers were aware of hospital and physician comparisons on quality, respectively, and fewer (6% and 8%) had used such information for decision-making (Greene et al., 2015). Engaged consumers can drive health care systems and physicians to report valid measures. We need to engage patients in helping us to define value from their perspective and determining the appropriate selection of measures. We need to engage policy makers in making more data on the performance of our health care system publicly available. Such efforts should garner bipartisan support inasmuch as improving our health care delivery system is a public health issue and market solutions will play an important role in improving quality and reducing costs. We need to coordinate efforts to report the health care system's performance with efforts to improve performance, such as by expanding the use of implementation science, adopting financial incentives, and tapping into the professional motivations of our health-care providers (Berwick, 2008; Marshall et al., 2013).

Given those needs, several strategic, specific federal efforts could help. Policy makers could create an independent body to write standards for health care performance measures and for the data used to populate the measures and, when appropriate, could approve standards developed by others. The independent body could finance the work to develop better measures. It would initially be designed to apply to situations in which the performance of individual hospitals or providers is used for accountability, such as public reporting or pay for performance. The structure of the organization ideally would reflect the interests of all

stakeholders, it would operate openly and transparently, it would offer the public the opportunity to provide input, and it would evolve. One option would be to build on NQF, which operates in a similar manner. The entity charged with this work ideally would be a private, nongovernment self-regulating organization, to ensure independence from competing interests. Informed by lessons learned by FASB, the organization might be structured to have a two-level board structure: a board of “standard-writing experts,” who would serve terms of 5–7 years, be compensated to attract the brightest minds, and be required to sever ties with industry; and a “foundation board” that would include stakeholders of many types and would oversee the organization and be responsible for fundraising.

To achieve that vision, those leading the effort would need to get stakeholder buy-in and navigate multiple tensions, including tensions between stimulating (not stifling) innovation in measurement and reporting and reflecting the values and preferences of various stakeholders. One specific initial step that could be considered would be for the Department of Health and Human Services to fund a 1-year planning and convening project to engage stakeholders and develop an initial design of a standard-setting body. That would be consistent with the recommendations in the National Academy of Medicine’s report *Vital Signs* (IOM, 2015).

Other strategic federal initiatives include encouraging the Agency for Healthcare Research and Quality (AHRQ), CDC, and PCORI to fund research on the science and development of performance measures, encouraging CMS to continue its efforts in this regard, and encouraging the multiple federal agencies involved in performance measurement to collaborate. The federal agencies can also support innovation in setting up multistakeholder “learning laboratories,” creating feedback loops, and identifying data sources, expertise, and test beds to develop needed measures more quickly. The laboratories could pilot-test and improve measures in the interval between when a measure is endorsed by NQF and when it is publicly reported and used in pay-for-performance programs; this would avoid the current process in which measures are revised after they are implemented (QualityNet, 2016). One approach for a learning laboratory might be to pick a small number of measures and coordinate the reporting and improvement efforts around them.

Federal initiatives should encourage greater transparency and sharing of data. For example, policy makers could prohibit gag clauses that preclude health plans and providers from sharing their data with employers, making it clear that self-insured employers own their claims data and can choose to share them as they see fit and enacting time limits for hospitals to share patient data and payers’ claims data to ensure the timeliness of data. Policy makers could strengthen regulations to support the sharing of patients’ data with patients themselves; this would reflect the principle that patient data belong to the patient.

The potential effects of those collective efforts should be enhanced quality and safety, reduced costs, enhanced patient choice and satisfaction, enhanced measurement science, and enhanced usefulness and use of performance measures to drive improvements in our health care system. The effects of these efforts can be tracked by monitoring and reporting the degree of transparency in reporting efforts, the progress made nationally on quality and cost, and the shifts in market share toward higher-value providers. To realize those goals, health care needs leadership and trust.

Conclusion

Despite important steps toward public reporting of the performance of our health care system, health care performance measurement has not yet achieved the desired goal of a system with higher quality and lower costs. Transparency of performance is a key tool for improving the health care system; however, if transparency is to serve as a tool for improvement, we need to ensure that the information that results from it is both accurate and meaningful.

The measures outlined in the introduction that have been successful—measures of infection associated with health care developed by the Centers for Disease Control and Prevention, measures of diabetes care, and AHRQ’s HCAHPS measures—have several common attributes. All were developed with substantial financial investment; they underwent extensive validation, revisions, and improvement; they published information about their validity; and they have wide acceptance among their users. The time is right to evolve a better performance-reporting system. That requires a commitment to the science of performance measurement, which in turn requires imagination,

Steps in the Performance Measurement and Reporting Process	Example of Where Errors Can Occur in the Step
Step 1: Developing and specifying the performance measure.	ProPublica's Surgical Scorecard, an on-line report of surgical quality, uses a measure of surgical complications that does not include many common, in-hospital complications that may be important to patients. As a result, this measure has the potential to mischaracterize a surgeon's actual performance. The measure did not undergo the National Quality Forum's formal endorsement process, so those using the measure for improvement or decision making know little about its validity and accuracy.
Step 2: Identifying and collecting the data used to populate the measure.	US News and World Report uses a hospital's Medicare data to calculate the Patient Safety Indicators (PSIs), a component of its Best Hospitals recognition program. When a patient arrives at the hospital with a preventable harm, that harm is coded as "present on admission" (POA). Events with a POA code are excluded from the PSI calculations. Medicare started requiring POA coding by hospitals in 2007, but hospitals in Maryland, with its Medicare waiver, were not required to include POA codes until 2013. The lack of POA codes for Maryland hospitals in the Medicare data probably penalized those hospitals when Medicare data were used to measure a hospital's quality.
Step 3: Applying the collected data to the specified performance measure.	An examination of common methods for creating physician cost profiles found that the reliability of the measures varied widely by specialty, ranging from 0.05 for vascular surgery to 0.79 for gastroenterology and otolaryngology. Overall, 59% of physicians had cost-profile scores with reliabilities of less than 0.70, a commonly used marker of suboptimal reliability. The authors estimated that 22% of physicians would be misclassified in a two-tiered system (high cost vs low or average cost). Perhaps this example suggests the importance of having a mechanism to pilot-test measures between their endorsement and their use for public policy (Adams et al., 2010).
Step 4: Categorizing performance and communicating the results.	Research has demonstrated that many of the current public reports make it cognitively burdensome for the audience to understand the data. For example, a report might share quality information in one place and cost information in another place, rather than assisting the consumer in identifying the "best buy" (Damman et al., 2015; Vaiana & McGlynn, 2002).

investment, infrastructure, and implementation. Without such commitment, our opportunity to achieve the goal of higher-value care is limited by our inability to understand our own performance.

Recommended Vital Directions

The three following vital directions have been identified for improving the health care measurement and reporting systems:

1. Create a health measurement and data standard-setting body. Fund a 1-year planning and convening project to engage stakeholders and develop an initial design of a standard-setting body in a way that is consistent with the recommendations in the Institute of Medicine report *Vital Signs: Core Metrics for Health and Health Care Progress* (IOM, 2015). In 2018, on the basis of this initial design, launch an independent body to write standards for health care performance measures and the data used to populate the measures. This helps to ensure that the information we have on the performance of the health care system is valid and accurate. The potential success of payment reform

(payment for the value of care, rather than the volume) will be limited if we cannot accurately assess the quality of care being provided. Depending on whether this work is tagged onto existing entities or a new entity is created to accomplish the work, we anticipate that these steps would take 2–5 years to accomplish.

2. Build the science of performance measures. Fund research on the science of performance measures and on the best ways to develop them and to pilot-test and improve them and encourage the multiple federal agencies involved in performance measurement to collaborate. That will help to move forward the science of performance measurement and ensure that different entities involved in the work are aligned. We anticipate that significant progress on the funding of research and the alignment of efforts would take 1–3 years to accomplish, with ongoing work thereafter.

3. Improve the communication of data to patients. Fund research on how to improve communication with patients about variations in the quality and costs of care, including examining reporting formats and the

framework within which consumers make different types of health care choices. For the health care market to work efficiently, we need health care consumers who are knowledgeable about the quality and cost of

the services that they seek, including the variation in quality and costs among providers. We anticipate that significant progress on the funding of research could take 1-2 years, with ongoing work thereafter.

Summary Recommendations for Vital Directions

1. Create a health measurement and data standard-setting body.
2. Build the science of performance measures.
3. Improve the communication of data to patients.

References

- Adams, J., A. Mehrotra, J. Thomas, and E. McGlynn. 2010. Physician cost profiling reliability and risk of misclassification. *New England Journal of Medicine* 362(11):1014-1021.
- AHRQ (Agency for Healthcare Research and Quality and Research (AHRQ)). 1998. Summary of recommendations: Providing strong leadership and clear aims for improvement. Available at <http://archive.ahrq.gov/hcqual/prepub/recommen.htm> (accessed June 14, 2016).
- Austin, J. 2015. Why health care performance measures need their own grades. 2015. Voices for Safer Care, Available at <https://armstronginstitute.blogs.hopkinsmedicine.org/2015/05/27/why-health-care-performance-measures-need-their-own-grades/> (accessed May 24, 2016).
- Austin, J., A. Jha, P. Romano, S. Singer, T. Vogus, R. Wachter, and P. Pronovost. 2015. National hospital ratings systems share few common scores and may generate confusion instead of clarity. *Health Affairs* 34(3):423-30.
- Austin, J., and P. Pronovost. 2016. Improving performance on core processes of care. *Current Opinion in Allergy and Clinical Immunology* 16(3):224-230.
- Austin, J., G. Young, and P. Pronovost. 2014. Ensuring the integrity and transparency of public reports: How a possible oversight model could benefit healthcare. *American Journal of Accountable Care* 2(4):13--14.
- Austin, J., A. Jha, P. Romano, S. Singer, T. Vogus, R. Wachter, and P. Pronovost. 2015. National hospital ratings systems share few common scores and may generate confusion instead of clarity. *Health Affairs* 34(3):423-430.
- Berenson, R., and T. Rice. 2015. Beyond measurement and reward: Methods of motivating quality improvement and accountability. *Health Services Research* 50:2155--2186.
- Berwick, D., B. James, and M. Coyle. 2003. Connections between quality measurement and improvement. *Medical Care* 41(1 Suppl):I30-I38.
- Berwick, D. M. 2008. The science of improvement. *The Journal of the American Medical Association* 299(10):1182-1184.
- Brown, D., A. Epstein, and E. Schneider. 2013. Influence of cardiac surgeon report cards on patient referral by cardiologists in New York state after 20 years of public reporting. *Circulation: Cardiovascular Quality and Outcomes* 6(6):643-648.
- CMS (Centers for Medicare and Medicaid Services (CMS)). 2016a. Healthcare-associated infections. Available at <https://www.medicare.gov/hospitalcompare/Data/Healthcare-Associated-Infections.html> (accessed May 25, 2016).
- Centers for Medicare and Medicaid Services (CMS). 2016b. Survey of patients' experiences (HCAHPS). Available at <https://www.medicare.gov/hospitalcompare/Data/Over-view.html> (accessed May 26, 2016).

- Consumer Reports. 2016. How we rate hospitals. Available at <http://www.consumerreports.org/cro/2012/10/how-we-rate-hospitals/index.htm> (accessed May 26, 2016).
- Cronin, C., C. Damberg, A. Riedel, and J. France. 2011. State-of-the-art of hospital and physician/physician group public reports. Paper presented at the Agency for Healthcare Research and Quality National Summit on Public Reporting for Consumers, Washington, DC.
- Damberg, C., M. Sorbero, S. Lovejoy, K. Lauderdale, S. Wertheimer, A. Smith, D. Waxman, and C. Schnyer. 2011. *An Evaluation of the Use of Performance Measures in Health Care*. Santa Monica, CA: RAND Corporation.
- Damman, O., A. De Jong, J. Hibbard, and D. Timmermans. 2015. Making comparative performance information more comprehensible: an experimental evaluation of the impact of formats on consumer understanding. *BMJ Quality & Safety* 25:860-869.
- Dixon-Woods, M., C. Bosk, E. Aveling, C. Goeschel, and P. Pronovost. 2011. Explaining Michigan: Developing an ex post theory of a quality improvement program. *The Millbank Quarterly* 89(2): 167--205.
- Dixon-Woods, M., S. McNicol, and G. Martin. 2012. Ten challenges in improving quality in healthcare: lessons from the Health Foundation's programme evaluations and relevant literature. *BMJ Quality & Safety* bmjqs-2011.
- Elliott, M., W. Lehrman, E. Goldstein, L. Giordano, M. Beckett, C. Cohea, and P. Cleary. 2010. Hospital survey shows improvements in patient experience. *Health Affairs* 29(11):2061-2067.
- Faber, M., M. Bosch, H. Wollersheim, S. Leatherman, and R. Grol. 2009. Public reporting in health care: How do consumers use quality-of-care information? A systematic review. *Medical Care* 47(1):1-8.
- FASB(FinancialAccountingStandardsBoard(FASB)).2016. About the FASB. Available at <http://www.fasb.org/jsp/FASB/Page/LandingPage&cid=1175805317407> (accessed May 25, 2016).
- Friedberg, M., P. Pronovost, D. Shahian, D. Safran, K. Bilimoria, M. Elliott, C. Damberg, J. Dimick, and A. Zaslavsky. 2015. *A methodological critique of the ProPublica Surgeon Scorecard*. Santa Monica, CA: RAND Corporation.
- Greene, J., V. Fuentes-Caceres, N. Verevkina, and Y. Shi. 2015. Who's aware of and using public reports of provider quality? *Journal of Health Care for the Poor and Underserved* 26(3):873-888.
- Hays, R., and J. Ware, Jr. 1986. My medical care is better than yours: social desirability and patient satisfaction ratings. *Medical Care* 519-525.
- Hibbard, J., and S. Sofaer. 2010. *Best practices in public reporting no. 1: how to effectively present health care performance data to consumers*. Rockville, MD: Agency for Healthcare Research and QualityAHRQ.
- Hibbard, J., J. Stockard, and M. Tusler. 2005. Hospital performance reports: impact on quality, market share, and reputation. *Health Affairs* 24(4):1150-1160.
- HSS (Hospital Safety Score (HSS)). 2016. About the score. Available at <http://www.hospitalsafetyscore.org/your-hospitals-safety-score/about-the-score> (accessed May 25, 2016).
- IOM (Institute of Medicine (IOM)). 2013. *Best care at lower cost: The path to continuously learning health care in America*. Washington, DC: The National Academies Press.
- Institute of Medicine (IOM). 2015. *Vital signs: Core metrics for health and health care progress*. Washington, DC: The National Academies Press.
- Jha, A. 2012. Hospital rankings get serious. An Ounce of Evidence / Health Policy blog. Available at <http://blogs.sph.harvard.edu/ashish-jha/hospital-rankings-get-serious/> (accessed May 25, 2016).
- Ketelaar, N., M. Faber, S. Flottorp, L. Rygh, K. Deane, and M. Eccles. 2011. *Public release of performance data in changing the behaviour of healthcare consumers, professionals or organisations*. The Cochrane Library.
- Lamb, G., M. Smith, W. Weeks, and C. Queram. 2013. Publicly reported quality-of-care measures influenced Wisconsin physician groups to improve performance. *Health Affairs* 32(3):536-543.
- Lau, B., E. Haut, D. Hobson, P. Kraus, C. Maritim, J. Austin, K. Shermock, B. Maheshwari, P. Allen, A. Almarino, and M. Streiff. 2015. ICD-9 code-based venous thromboembolism performance targets fail to measure up. *American Journal of Medical Quality* 106286061558354731(5):448-453.

- Ketelaar, N., M. Faber, S. Flottorp, L. Rygh, K. Deane and M. Eccles. 2011. *Public release of performance data in changing the behaviour of healthcare consumers, professionals or organisations*. The Cochrane Library.
- Marjoua, Y., and K. Bozic. 2012. Brief history of quality movement in US healthcare. *Current Reviews in Musculoskeletal Medicine* 5(4):265-273.
- Marshall, M., P. Pronovost, and M. Dixon-Woods. 2013. Promotion of improvement as a science. *The Lancet* 381(9864):419-421.
- McGlynn, E. 2003. An evidence-based national quality measurement and reporting system. *Medical Care* 41(1):11-8.
- McGlynn, E., and J. Adams. 2014. What makes a good quality measure? *The Journal of the American Medical Association* 312(15):1517-1518.
- McGlynn, E., and E. Kerr. 2016. Creating safe harbors for quality measurement innovation and improvement. *The Journal of the American Medical Association* 315(2):129-130.
- NCQA (National Committee for Quality Assurance (NCQA). 2016. HEDIS Compliance Audit Program. Available at <http://www.ncqa.org/tabid/205/Default.aspx> (accessed June 21, 2016).
- NQF (National Quality Forum (NQF). 2016a. What NQF endorsement means. Available at http://www.qualityforum.org/Measuring_Performance/ABCs/What_NQF_Endorsement_Means.aspx (accessed May 25, 2016).
- National Quality Forum (NQF). 2016b. Variation in Measure Specifications Project 2015-2016: First draft report. Available at <http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=82318> (accessed May 25, 2016).
- O'Neil, S., J. Schurrer, and S. Simon. 2010. *Environmental scan of public reporting programs and analysis* (Final Report by Mathematica Policy Research). Washington, DC: National Quality Forum, Washington, DCNQF.
- O'Sullivan, M. 2015. *Safety in numbers: Cancer surgeries in California hospitals*. California Healthcare Foundation.
- Owens, P. 2014. Letter to NQF Patient Safety Steering Committee members and NQF staff, Subject: PSI 90 for maintenance endorsement. Available at <http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=77096> (accessed May 25, 2016).
- Parker, C., L. Schwamm, G. Fonarow, E. Smith, and M. Reeves. 2012. Stroke quality metrics systematic reviews of the relationships to patient-centered outcomes and impact of public reporting. *Stroke* 43(1):155-162.
- Provonost, P., M. Miller, and R. Wachter. The GAAP in Quality Measurement and Reporting. 2007. *The Journal of the American Medical Association* 298:1800-1802.
- Pronovost, P., J. Marsteller, and C. Goeschel. 2011. Preventing bloodstream infections: a A measurable national success story in quality improvement. *Health Affairs* 30(4):628-634.
- Pronovost, P, M. Miller, and R. Wachter. 2007. *The Journal of the American Medical Association* 298:1800-1802.
- QualityNet. 2016. Specifications manual, Version 5.1. Available at <https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier3&cid=1228775436944> (accessed July 19, 2016).
- Shahian, D., E. Edwards, J. Jacobs, R. Prager, S. Normand, C. Shewan, S. O'Brien, E. Peterson, and F. Grover. 2011a. Public reporting of cardiac surgery performance: part Part 1—history, rationale, consequences. *The Annals of Thoracic Surgery* 92(3):S2-S11.
- Shahian, D., E. Edwards, J. Jacobs, R. Prager, S. Normand, C. Shewan, S. O'Brien, E. Peterson, and F. Grover. 2011b. Public reporting of cardiac surgery performance: part Part 2—implementation. *The Annals of Thoracic Surgery* 92(3):S12-S23.
- Shaller, D., D. Kanouse, and M. Schlesinger. 2014. Context-based strategies for engaging consumers with public reports about health care providers. *Medical Care Research and Review* 71(5 supplSuppl):17S-37S.
- Shwartz, M., J. Restuccia, and A. Rosen. 2015. Composite measures of health care provider performance: A description of approaches. *The Milbank Quarterly* 93:788–825.
- Smith, M., A. Wright, C. Queram, and G. Lamb. 2012. Public reporting helped drive quality improvement in outpatient diabetes care among Wisconsin physician groups. *Health Affairs* 31(3):570-577.
- Stey, A., M. Russell, C. Ko, G. Sacks, A. Dawes, and M. Gibbons. 2015. Clinical registries and quality measurement in surgery: a A systematic review. *Surgery* 157(2):381-395.

STS (Society of Thoracic Surgeons). 2016. STS public reporting online. Available at <http://www.sts.org/quality-research-patient-safety/sts-public-reporting-online> (accessed May 25, 2016).

Totten, A., J. Wagner, A. Tiwari, C. O'Haire, J. Griffin, and M. Walker. 2012. Public reporting as a quality improvement strategy. Closing the quality gap: Revisiting the state of the science. Rockville, MD: Agency for Healthcare Research and Quality AHRQ.

Vaiana, M., and E. McGlynn. 2002. What cognitive science tells us about the design of reports for consumers. *Medical Care Research and Review* 59(1): 3-35.

Whaley, C., J. Chafen, S. Pinkard, G. Kellerman, D. Bravata, R. Kocher, and N. Sood. 2014. Association between availability of health service prices and payments for these services. *The Journal of the American Medical Association* 312(16):1670-1676.

Winters, B., A. Bharmal, R. Wilson, A. Zhang, L. Engineer, D. Defoe, E. Bass, S. Dy, and P. Pronovost. 2016. Validity of the Agency for Health Care Research and Quality patient safety indicators and the Centers for Medicare and Medicaid hospital-acquired conditions: A systematic review and meta-analysis. *Medical Care* 54(12):1105-1111.

Suggested Citation

Pronovost, P.J., J.M. Austin, C. K. Cassel, S.F. Delbanco, A.K. Jha, B. Kocher, E.A. McGlynn, L.G. Sandy, and J. Santa. 2016. *Fostering Transparency in Outcomes, Quality, Safety, and Costs*. Vital Directions for Health and Health Care Series. Discussion Paper, National Academy of Medicine, Washington, DC. <https://nam.edu/wp-content/uploads/2016/09/fostering-transparency-in-outcomes-quality-safety-and-costs.pdf>.

Author Information

Peter J. Pronovost, MD, PhD, FCCM, is Senior Vice President for Patient Safety and Quality, and Director, Armstrong Institute for Patient Safety and Quality, Johns Hopkins Medicine. **J. Matthew Austin, PhD, MS**, is Assistant Professor of Anesthesiology and Clinical Care Medicine, Johns Hopkins Medicine. **Christine K. Cassel, MD**, is Planning Dean, Kaiser Permanente School of Medicine. **Suzanne F. Delbanco, PhD**, is Executive Director, Catalyst for Payment Reform. **Ashish K. Jha, MD, MPH**,

is K.T. Li Professor of International Health and Director, Harvard Global Health Institute, Harvard T.H. Chan School of Public Health. **Bob Kocher, MD**, is a Partner, Venrock. **Elizabeth A. McGlynn, PhD**, is Director, Center for Effectiveness and Safety Research, Kaiser Permanente. **Lewis G. Sandy, MD, FACP**, is Executive Vice President, Clinical Advancement, UnitedHealth Group. **John Santa, MD, MPH**, is Former Director, Health Ratings Center, Consumer Reports.

Acknowledgments

Elizabeth Finkelman, director of the Vital Directions for Health and Health Care Initiative, provided valuable support for this paper.

Sponsors

The National Academy of Medicine's Vital Directions for Health and Health Care initiative is sponsored by the California Health Care Foundation, The Commonwealth Fund, the Gordon and Betty Moore Foundation, The John A. Hartford Foundation, the Josiah Macy Jr. Foundation, the Robert Wood Johnson Foundation, and the National Academy of Medicine's Harvey V. Fineberg Impact Fund.

Disclaimer

The views expressed in this Perspective are those of the authors and not necessarily of the authors' organizations, the National Academy of Medicine (NAM), or the National Academies of Sciences, Engineering, and Medicine (the National Academies). The Perspective is intended to help inform and stimulate discussion. It has not been subjected to the review procedures of, nor is it a report of, the NAM or the National Academies. Copyright by the National Academy of Sciences. All rights reserved.