Viewing Health Care Delivery as Science: Challenges, Benefits, and Policy Implications

Peter J. Pronovost and Christine A. Goeschel

The need for health services research is likely to rise rapidly as the population ages, health care costs soar, and therapeutic and diagnostic choices proliferate. Building an effective and efficient health care delivery system is a national priority. Yet the national health care quality report concludes that we lack the ability to monitor progress toward even basic quality and patient safety goals effectively.

The gap between the need to improve and our ability to do so exists in part because we fail to view the delivery of health care as science, we lack national improvement priorities, and we lack a national infrastructure to achieve our stated goals.

We discuss key challenges implicit in correcting these failures and recommend actions to expedite progress.

Key Words. Quality of care/patient safety (measurement), health policy/politics/law/regulation, health care organizations and systems

BACKGROUND

The U.S. health care system is expensive, consuming 18 percent of our gross domestic product in 2009, and frequently fails to deliver high-quality care. Over 100,000 patients die annually from hospital-acquired infections (HAIs), (Klevens et al. 2007) and patients on average receive 50 percent of recommended therapies (McGlynn et al. 2003). In 2001, the Institute of Medicine report, *Crossing the Quality Chasm*, estimated up to a 17-year lapse before some evidence-based therapies are commonly used (Institute of Medicine 2001).

The poor performance in health care delivery contrasts sharply with the tremendous success in basic and clinical biomedical research in the United States. For example, researchers sequenced all three billion letters of the human genome with 99.99 percent accuracy, and life-saving drugs and devices emerge weekly. This difference is understandable; the United States invests little in understanding and improving the science of health care delivery. Until

now, for every dollar Congress allocated to develop breakthrough treatments, it allocated a penny to ensure patients received those treatments. Consequently, the science to improve health care delivery is immature. Examples of large-scale quality improvements are rare, methods to evaluate progress in quality are virtually nonexistent, and most importantly, patients remain at risk of harm. We believe a public-private partnership and research investment similar to the human genome project could improve quality and reduce the costs of health care.

The lack of data to analyze, understand, and ultimately improve health care is a complex local and national problem, leaving consumers in the dark, and ensuring patients continue to suffer preventable harm and costs. The 2008 National Health Care Quality Report concluded that most areas of quality lack valid measures to evaluate progress (Agency for Health Care Research and Quality [AHRQ] 2008).

To improve quality, the country needs clear goals and a coordinated strategy to achieve them; health care clinicians and leaders with the knowledge and skills to lead and evolve the strategy; transparent and robust measures; clinicians incented to improve performance; and policy makers committed to the science and to public accountability. This will require far greater collaboration and interdependence than currently exists.

In this essay, we discuss some of the challenges, potential benefits, and policy implications of measuring progress in quality. We also make recommendations to rapidly mature the science and public accountability for the quality of care provided to patients.

Challenges to Viewing Health Care Delivery as Science

Quality of care is a nascent science (Wachter 2004; Pronovost et al. 2009), and health policy to support science should involve measuring performance,

Address correspondence to Peter J. Pronovost, M.D., Ph.D., Sc.D., (hon) F.C.C.M., Departments of Anesthesiology and Critical Care Medicine and Surgery and The Quality and Safety Research Group, School of Medicine, The Johns Hopkins University, 1909 Thames Street, Baltimore, MD 21231; e-mail: ppronovo@jhmi.edu. Peter J. Pronovost, M.D., Ph.D., Sc.D., (hon) F.C.C.M., is also with the Department of Health Policy and Management, Bloomberg School of Public Health, and School of Nursing, The Johns Hopkins University, Baltimore, MD. Christine A. Goeschel, M.P.A., M.P.S., Sc.D., R.N., is with the Department of Anesthesiology and Critical Care Medicine and The Quality and Safety Research Group, School of Medicine, Johns Hopkins University, Baltimore, MD. Christine A. Goeschel, M.P.A., M.P.S., Sc.D., R.N., is also with the Department of Health Policy and Management, Bloomberg School of Public Health, and Department of Health Systems Outcomes, School of Nursing, The Johns Hopkins University, Baltimore, MD.

making it transparent, and buying smarter. Though such support is necessary, current health care quality measures are not up to this task (Petersen et al. 2006; Rosenthal and Dudley 2007). National policy that effectively improves quality value will require investments in measuring and reporting the quality of care.

Valid, Reliable Measures are Needed

A major obstacle in developing rigorous measures to evaluate the quality of care has been the difficulty of distinguishing indicators that can be validly measured as rates, from those that cannot (Pronovost et al. 2006a, b, c, d; Pronovost, Miller, and Wachter 2006f). Many parameters of quality are inaccurately measured and inappropriately presented as rates for several reasons: (1) events are uncommon (medication errors associated with significant harm or death) or rare (wrong site surgical procedures); (2) few have standardized definitions for events or those at risk for events; (3) surveillance systems typically rely on self-reporting; (4) denominators (the populations at risk) are largely unknown; and (5) the time period for exposure (patient day or device day) is unspecified. Creating a measurement system free of the biases introduced by these limitations will be complex and costly. With broad input, health care has to decide what to measure and where the costs of data collection are worth the benefits. Perhaps analytic tools such as the value of information can inform this process.

For events that are not presented as rates, the measure of improvement is whether we learned from the event and reduced the risk of future patient harm. No measure is perfect, and all measures will have some bias. Generally, less biased measures are more expensive than more biased measures. Yet the biases and costs of data collection should be known and made transparent. For most measures of quality, this information is unknown, limiting the usefulness of the measures, squandering scarce resources, and misinforming the public.

Difficulty Discerning Preventable versus Inevitable Harm

Pay-for-performance policies that focus on patient outcomes build on the premise that errors in the delivery of care are singularly responsible for specific patient harms (e.g., infections or decubitus ulcers). Yet some patients will inevitably suffer harm despite receiving evidence-based therapies; disentangling preventable from inevitable harm is a thorny issue. For example, if the harm (e.g., mortality from acute myocardial infarction) is only partially preventable (as most are), we must distinguish inevitable from preventable

harm (Bradley et al. 2006; Werner and Bradlow 2006). The methods to make this distinction are immature. Over time, therapeutic and technical advances will no doubt change our knowledge of what is preventable and what is inevitable. An important policy question is what proportion of harm should be avoidable before we characterize it as preventable? How accurate must a measure of quality be before it guides payment or is publically reported?

Valid measures of preventable harm require clear definitions of the event (numerator) and those at risk for the event (denominator), and a standardized surveillance method to identify both indicators. Clinicians have generally labeled almost all harm as inevitable, seeking to create highly specific (truly inevitable cases are accurately labeled) though not very sensitive (truly preventable cases are not always labeled as preventable) measures. This approach misses many patients who experience preventable harm, although those who are identified provide valuable information for improvement efforts.

Recently, the Centers for Medicare and Medicaid Services and private insurers stopped paying the marginal costs for harms they believe are entirely preventable, creating highly sensitive though not very specific measures, capturing all preventable harm, and mislabeling inevitable harm as preventable. Both approaches have risks and benefits. Policy makers should carefully consider the following three strategies and understand the relative risks and benefits of each. The strategy selected to guide policy should be diagnosis or event specific.

1. Assume All Harm Is Preventable: A High-Sensitivity—Low-Specificity Strategy

This approach (assume all harm is preventable and harm rates can be monitored directly) is appropriate when evidence suggests that nearly all harm is preventable. Central line-associated bloodstream infections (CLABSI) can be validly measured and are largely preventable.

2. Adjust for Preventability: A Low-Sensitivity—Low-Specificity Strategy

A second strategy could use risk-adjustment models to account for preventable versus inevitable harm. Intensive care unit mortality, overall hospital mortality, or mortality after specific conditions (e.g., acute myocardial infarction) are often calculated using risk-adjustment models. These models typically adjust for patient variables trading off random versus systematic error. By selecting a specific condition compared with overall hospital mortality,

risk-adjustment models may have less systematic error but (due to smaller sample sizes) will have larger random error. In addition, surveillance bias and incomplete risk adjustment often influences the performance of the risk-adjustment models (Shojania et al. 2002). Accruing evidence suggests that case mix adjustment may at times be more misleading than crude comparisons because of “the constant risk fallacy,” which assumes that the relationship between case mix variables and outcomes are the same in all comparison groups (such as among different hospitals) when most often they are not (Mohammed et al. 2009). Thus, methods that use observed-to-expected ratios to estimate hospital mortality are likely inaccurate (Hofer, Kerr, and Hayward 2000; Shojania et al. 2002).

Because risk-adjustment models do not account for differences in the quality of care delivered, and because patients on average receive half of the recommended therapies, risk-adjustment models may anchor performance to the status quo, underestimating the degree of preventable harm (McGlynn et al. 2003). Finally, measures of overall hospital mortality could encourage overly aggressive care that may not be consistent with patients’ wishes (Holloway and Quill 2007). Given that patients often receive aggressive therapies they do not desire, that substantial health care spending occurs in the last 6 months of life, and the biases in overall hospital mortality as a measure of quality, investment to create better measures is needed (Fisher 2006).

3. Link Care Received to Outcome: A High-Specificity—Low-Sensitivity Strategy

The third strategy could link care processes to the adverse outcome measure (Pronovost and Colantuoni 2009). For example, the extent to which surgical-site infections are preventable is unknown. Yet we know that administering antibiotics before surgery can reduce the risk of developing an infection. If the patient does not receive an antibiotic and develops an infection, the infection could be labeled as preventable. Although this process-outcome model has face validity, it has shortcomings; evidence-based therapies are often absent or limited, omitting harm that results from errors in teamwork and communication.

Given the risks and benefits of each strategy, science should guide policy. Clinicians and researchers should assume that all harm is preventable, determining the extent to which this is true. If most harm is preventable, payment policy could follow. If some harm is preventable, policy makers could link the process to the outcome or use risk adjustment to identify preventable harm, and support research to develop new knowledge regarding

how to mitigate that harm. Health policy should support the valid and transparent measurement of outcomes and encourage provider innovation to improve those outcomes.

Lack of a Meaningful Scorecard

In spite of existing limitations, a model to measure quality of care is necessary and can be informative. We developed a scorecard that builds upon Donabedian's quality measurement model—structure (how we organize care), process (what we do), outcome (the results achieved)—by adding a culture of safety (Pronovost et al. 2006b, d). Thus, the scorecard contains four measures (Pronovost, Miller, and Wachter 2006f). The first two measures are intended to be diagnosis specific and measured as rates with minimal bias: (1) How often are patients harmed? (2) How often do clinicians provide evidence-based interventions? The second two measures are agnostic to diagnosis and generally not feasible or inappropriate for rate-based reporting: (3) How do we know clinicians learned from mistakes (defects)? (4) How successful are clinicians and health care organizations at creating and improving a culture of safety?

While we recognize that this score must mature and that it focuses on patient safety rather than the broader construct of quality, it provides a useful construct to measure progress in improving patient safety. Admittedly, there are precious few valid measures of preventable harm—a limit based more on our lack of investment than the boundaries of science.

The second type of rate-based measure is whether patients received evidence-based interventions (Hofer and Hayward 2002; Pronovost et al. 2004). Medicare publicly reports compliance on a variety of process measures (Werner and Bradlow 2006). Although few process measures can be reliably linked to outcomes, they are often informative especially if linked with outcome measures. The failure of process measures to correlate with outcomes may have more to do with measurement error than the lack of an empiric relationship between process and outcome (Bradley et al. 2006). When evaluating process measures, it is important to consider their validity on two levels: is the intervention valid (is the intervention associated with improved outcomes), and is the method of measuring the intervention valid (is the intervention accurately measured)? For example, educating acute myocardial infarction patients about smoking cessation is frequently reported as a quality indicator but not empirically associated with reduced mortality. Even if smoking cessation education were associated with mortality, the method by which it is measured (checking a box to indicate it was done)

likely correlates poorly with a patient's comprehension or behavior change. The number of process measures is growing exponentially; a value of information analysis may help inform which measures to collect. Unless these measures are rooted in science, they will not likely lead to improved patient outcomes.

The third scorecard measure addresses learning from defects. Although most harmful events cannot feasibly be validly measured as rates, it is possible to evaluate organizational learning from these mistakes. For example, after a harmful medication error and corrective action, it is important to know if this risk is reduced for future patients by assessing four things (organized from least to most valid and resource intensive): does an appropriate policy currently exist, or was a new policy or procedure developed, do frontline clinicians know about it, are they using it as intended (this generally requires an audit of behaviors), and do frontline clinicians perceive risks were reduced (Pronovost et al. 2006a, c)?

The final scorecard metric focuses on organizational safety culture, a measure that is linked to most types of sentinel events, is now required by The Joint Commission and can be improved with focused interventions (Gaba et al. 2003; Colla et al. 2005; Pronovost et al. 2006e; Sexton et al. 2006; Singla et al. 2006).

Tension remains between broad national measures (low cost with substantial bias) and narrowly focused measures (high cost and more robust) (Holloway and Quill 2007). We believe quality scorecards must be meaningful to the clinicians who are expected to improve care, useful to consumers who purchase care, and scalable from the unit level up to the national level. Such a scorecard was applied in over 100 intensive care units in Michigan; teams continue to use it to track culture scores and CLABSI rates, and some have added other rate-based measures like surgical-site infections, and defects such as wrong-sided surgeries and mislabeled specimens.

Benefits of Having Robust and Transparent Measures of Quality

Improved Capacity to Establish National Improvement Priorities. Imagine that the United States is faced with a deadly disease that claimed nearly 100,000 victims a year or 2.5 million lives over the last quarter century. Researchers developed a new therapy to nearly eliminate most of these deaths. They implement the therapy throughout the state of Michigan, saving an estimated 2,000 lives and U.S.\$200 million annually (Pronovost et al. 2006a, b). If the therapy were a drug or device, the private sector would quickly produce and sell it. Costs would drop; quality would rise; and the therapy would rapidly

spread throughout the United States, saving more lives than virtually any other medical discovery in the last quarter century.

The disease is real, deadly, and expensive. It is HAI. The equally real therapy is not a drug or device, but a quality improvement program. This program used tools to improve teamwork and safety culture, summarized clinical evidence into a checklist, measured infection rates, and reported results at the unit, hospital, and state levels. Yet when results of this program were publicized, it did not quickly spread. While every state said they were using the “checklist,” only 11 measured their rates of infections and none were as low as Michigan. With support from the AHRQ, the MI program is now being implemented in all 50 states (<http://www.safercare.net>).

Why did the market fail to move swiftly to reduce these infections? First, without valid and transparent reporting of these infections, these deaths are largely invisible. Second, public investment in the science of health care delivery is woefully inadequate and we mistakenly view infections as inevitable rather than preventable. Third, in many respects it is easier to produce a specific technology, drug, or medical device than it is to implement an organization-wide change (Heifetz 1994; Pronovost et al. 2006b,d; Shortell and Singer 2008).

Ability to Ensure Data Integrity. Quality-of-care measures will only be useful if they are reasonably accurate. Our current system for reporting this data is neither sufficiently standardized nor accurate. The U.S. Securities and Exchange Commission (SEC) and the Financial Accounting Standards Board (FASB) offer data integrity models that we can adapt in health care.

The Securities Exchange Act of 1934 created the SEC to ensure accurate reporting of financial data. The mission of the SEC is to protect investors; maintain fair, orderly, and efficient markets; and facilitate capital formation (SEC 2008). To achieve this, the SEC requires public companies to disclose meaningful financial and other information to the public. Importantly, the development and reporting of standardized, timely, and comprehensive financial data did not happen voluntarily; it required wise regulation.

Current challenges in health care are similar to those faced in the 1930's financial markets. Public confidence in health care is waning; data are often inaccurate or misleading, of limited use to clinicians and consumers, and seemingly intended more for marketing than improving quality or ensuring accountability. Few of the professionals measuring quality at hospitals have the training to understand the biases in quality measures, report quality-of-care data, and perform audits to ensure the reports are accurate. Nor are there

meaningful penalties for those who digress. The reporting of health care quality is reminiscent of pre-SEC financial reporting. The federal government could create a health care version of the SEC, transparently reporting outcomes and costs, creating standards and auditing in the private sector, and enforcing penalties.

State and federal government agencies are issuing “performance” reports on individual providers and hospitals. Hospitals and other health care providers are using scorecards to market their services on websites, billboards and television, and in glossy brochures. Nonetheless, the accuracy of claims is often questionable given that most quality measurement in health care is neither standardized nor consistently reliable. Examples of misleading health care information are easy to find (Pronovost, Miller, and Wachter 2007). Hospitals, health systems, and private sector companies report quality measures without oversight guaranteeing the accuracy of their claims. Not surprisingly, most U.S. hospitals boast about being part of at least one “top” list, evoking echoes of author Garrison Keillor’s Lake Wobegone, a fictional town in Minnesota where all the children are “above average.”

There are tensions associated with determining the quality of data necessary to publicly report or to justify payment based on performance. For example, data from the federal government are generally thought to be more robust than other sources. The CDC collects robust clinical data regarding CLABSI, yet CMS and many states measure these infections using notoriously inaccurate hospital billing data (U.S. Government Accounting Office 2006; Centers for Medicare & Medicaid Services 2009). The cost-effectiveness of various data sources is unknown, and the use of health information technology (HIT) will likely reduce the costs of collecting data. Nevertheless, the trade-off between the two should be made explicit.

SUMMARY

Health care has few measures of quality, especially outcome measures, generally relying on administrative data collected for financial purposes rather than more accurate clinical data. The electronic medical record offers great hope to increase the accuracy and reduce the costs of collecting valid outcome measures, yet research is needed to make this hope a reality. Health policy makers have yet to determine the extent to which they are committed to robust quality measures. Such a commitment would require investments to develop measures, create a data infrastructure, and train professionals to administer programs with clearly articulated specifications, audits, and data accuracy

plans. By creating the equivalent of the SEC, policy makers could accomplish all of these, encourage innovation in measurement development and data collection, and likely reduce preventable harm.

The investment in HIT presents a unique opportunity to augment the backbone for measuring quality of care. If measures and their specifications are built into HIT systems, the United States could evolve a robust and efficient mechanism to assess quality of care. Then, as the science of quality evolves, the ability to measure progress in a scalable and actionable manner will become a reality. Policy makers will be able to make decisions based on the best evidence rather than on hunches, incomplete information, or special interests.

Recommendations

1. Invest in Measurement

- a. Support the development and implementation of robust, scalable quality-of-care measures aligned with national quality and quality improvement goals.
- b. For example, the National Patient Safety Goal for improved anticoagulation (NPSG.03.05.01) could be a national pilot study. The goal currently exists without standardized measures, data collection methods, reporting structures, or auditing functions. Valid tracking of national progress is impossible without such specifications.
- c. Make measure performance transparent; errors should be explicit, quantified, and transparent.
- d. Focus on outcomes and provider performance. Policy goals should seek to ensure patients receive the best-known science using standardized measures of performance, and encourage providers to be innovative in improving performance.

2. Prioritize National Quality Improvement Goals

The federal government must incorporate the following characteristics when developing methods to prioritize quality improvement efforts: transparency and accountability of the work, including disclosure of ethical and technical rationales for decisions taken; an ethics framework that identifies and explicates the specific moral principles that should guide priority setting in quality; and public engagement and participation in patient safety priorities (Pronovost and Faden 2009). They should also consider using more formal

methods, such as the value of information, to assess the costs and benefits of collecting measures of quality.

3. *Create the Equivalent of the SEC and FASB in Health Care*

The federal government should institute SEC-like regulations for reporting financial data when reporting quality data to ensure its accuracy (Prohovost, Miller, and Wachter 2007).

- a. Develop national standards for the reporting of health care quality data (equivalent of generally accepted accounting principles).
 - i. Link HIT efforts with quality monitoring to develop structured data elements. Data mining and hand-collected data are inefficient, often inaccurate, and costly.
 - ii. Build measure specifications into HIT systems to establish a robust and efficient mechanism to assess quality of care based on clinical rather than administrative data.
- b. Create a private sector standards setting and monitoring body to develop measures, audit, and enforce the accurate reporting of health care quality (equivalent of FASB) (FASB 2007).
 - i. Decisions about allocation of resources and choice of provider rely heavily on concise, transparent, credible, and understandable information about quality of care.
- c. Focus on outcome measures or linked outcome-process measures rather than just process measures.
- d. Require disclosure by provider organizations, with public reporting and private sector secondary analysis. Providers should be required to make this information publicly available. Yet innovation is needed for more meaningful reporting of performance to the public and clinicians. The private sector could compete to accomplish this.
- e. Support the development of a health care quality profession, where individuals trained in standardized quality monitoring rules, auditing functions, and accountability and reporting requirements assume formal and credible roles within organizations.

President Obama suggested the new administration would restore science to its rightful place . . . raise health care quality . . . and lower its costs and that our “business should be conducted in the light of day.” The suggestions described above can help achieve these goals.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: Drs. Pronovost and Goeschel each report receiving honoraria for speaking about health care leadership, quality and patient safety measurement and improvement, health policy to improve the science of health delivery, and related topics. The authors would like to acknowledge the unrelenting passion and tireless efforts of the entire Quality and Safety Research Group. They are changing the world with their thirst for new ideas, their willingness to collaborate, and their unflinching commitment to patients as our north star.

Disclosures: None.

Disclaimers: None.

REFERENCES

- Agency for Health Care Research and Quality (AHRQ). 2008. *National Healthcare Quality Report 2007*. Rockville, MD: Agency for Healthcare Research and Quality.
- Bradley, E. H., J. Herrin, B. Elbel, R. L. McNamara, D. J. Magid, B. K. Nallamothu, Y. Wang, S. L. Normand, J. A. Spertus, and H. M. Krumholz. 2006. "Hospital Quality for Acute Myocardial Infarction: Correlation among Process Measures and Relationship with Short-Term Mortality." *Journal of the American Medical Association* 296 (1): 72–8.
- Centers for Medicare & Medicaid Services. 2009. "Reporting Hospital Quality Data for Annual Payment Update" [accessed on July 15, 2009]. Available at http://www.cms.hhs.gov/HospitalQualityInits/08_HospitalRHQDAPU.asp
- Colla, J., A. Bracken, L. Kinney, and W. Weeks. 2005. "Measuring Patient Safety Climate: A Review of Surveys." *Quality and Safety in Health Care* 14: 364–6.
- Financial Accounting Standards Board (FASB). 2007. *Facts about FASB*. Norwalk, CT: FASB.
- Fisher, E. S. 2006. "Paying for Performance—Risks and Recommendations." *New England Journal of Medicine* 355 (18): 1845–7.
- Gaba, D. M., S. J. Singer, A. D. Sinaiko, J. D. Bowen, and A. P. Ciavarella. 2003. "Differences in Safety Climate between Hospital Personnel and Naval Aviators." *Human Factors* 45 (2): 173–85.
- Heifetz, R. A. 1994. *Leadership without Easy Answers*. Cambridge, MA: The Belknap Press of Harvard Press.
- Hofer, T. P., and R. A. Hayward. 2002. "Are Bad Outcomes from Questionable Clinical Decisions Preventable Medical Errors? A Case of Cascade Iatrogenesis." *Annals of Internal Medicine* 137 (5, part 1): 327–33.

- Hofer, T. P., E. A. Kerr, and R. A. Hayward. 2000. "What Is an Error?" *Effective Clinical Practice* 3 (6): 261–9.
- Holloway, R. G., and T. E. Quill. 2007. "Mortality as a Measure of Quality: Implications for Palliative and End-of-Life Care." *Journal of the American Medical Association* 298 (7): 802–4.
- Institute of Medicine. 2001. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academies Press.
- Klevens, R. M., J. R. Edwards, C. L. Richards Jr, T. C. Horan, R. P. Gaynes, D. A. Pollock, and D. M. Cardo. 2007. "Estimating Health Care-Associated Infections and Deaths in U.S. Hospitals, 2002." *Public Health Reports* 122 (2): 160–6.
- McGlynn, E. A., S. M. Asch, J. Adams, J. Keesey, J. Hicks, A. DeCristofaro, and E. A. Kerr. 2003. "The Quality of Health Care Delivered to Adults in the United States." *New England Journal of Medicine* 348 (26): 2635–45.
- Mohammed, M. A., J. J. Deeks, A. Girling, G. Rudge, M. Carmalt, A. J. Stevens, and R. J. Lilford. 2009. "Evidence of Methodological Bias in Hospital Standardised Mortality Ratios: Retrospective Database Study of English Hospitals." *British Medical Journal* 338: b780.
- Petersen, L. A., L. D. Woodard, T. Urech, C. Daw, and S. Sookanan. 2006. "Does Pay-for-Performance Improve the Quality of Health Care?" *Annals of Internal Medicine* 145 (4): 265–72.
- Pronovost, P., D. Needham, S. Berenholtz, D. Sinopoli, H. Chu, S. Cosgrove, B. Sexton, R. Hyzy, R. Welsh, G. Roth, J. Bander, J. Kepros, and C. Goeschel. 2006a. "An Intervention to Decrease Catheter-Related Bloodstream Infections in the ICU." *New England Journal of Medicine* 355 (26): 2725–32.
- Pronovost, P. J., S. M. Berenholtz, C. A. Goeschel, D. M. Needham, J. B. Sexton, D. A. Thompson, L. H. Lubomski, J. A. Marsteller, and M. A. Makary. 2006b. "Creating High Reliability in Healthcare Organizations." *Health Services Research* 41 (4): 1599–617.
- Pronovost, P. J., and E. Colantuoni. 2009. "Measuring Preventable Harm: Helping Science Keep Pace with Policy." *Journal of the American Medical Association* 301 (12): 1273–5.
- Pronovost, P. J., and R. R. Faden. 2009. "Setting Priorities for Patient Safety: Ethics, Accountability, and Public Engagement." *Journal of the American Medical Association* 302 (8): 890–1.
- Pronovost, P. J., C. A. Goeschel, J. A. Marsteller, J. B. Sexton, J. C. Pham, and S. M. Berenholtz. 2009. "A Framework for Patient Safety Research and Improvement." *Circulation* 119 (2): 330–7.
- Pronovost, P. J., C. G. Holzmueller, E. Martinez, C. L. Cafeo, D. Hunt, C. Dickson, M. Awad, and M. A. Makary. 2006c. "A Practical Tool to Learn from Defects in Patient Care." *Joint Commission Journal on Quality and Patient Safety* 32 (2): 102–8.
- Pronovost, P. J., C. G. Holzmueller, J. B. Sexton, M. Miller, S. Berenholtz, A. W. Wu, T. Perl, R. Davis, D. Baker, L. Winner, and L. Morlock. 2006d. "How Will We Know Patients Are Safer? An Organization-Wide Approach to Measuring and Improving Patient Safety." *Critical Care Medicine* 34 (7): 1988–95.

- Pronovost, P. J., J. King, C. G. Holzmüller, M. Sawyer, S. Bivens, M. Michael, K. Haig, L. Paine, D. Moore, and M. Miller. 2006e. "A Web-Based Tool for the Comprehensive Unit-Based Safety Program (CUSP)." *Joint Commission Journal on Quality and Patient Safety* 32 (3): 119–29.
- Pronovost, P. J., M. Miller, and R. M. Wachter. 2007. "The GAAP in Quality Measurement and Reporting." *Journal of the American Medical Association* 298 (15): 1800–2.
- Pronovost, P. J., M. R. Miller, and R. M. Wachter. 2006f. "Tracking Progress in Patient Safety: An Elusive Target." *Journal of the American Medical Association* 296 (6): 696–9.
- Pronovost, P. J., M. L. Rinke, K. Emery, C. Dennison, C. Blackledge, and S. M. Berenholtz. 2004. "Interventions to Reduce Mortality among Patients Treated in Intensive Care Units." *Journal of Critical Care* 19 (3): 158–64.
- Rosenthal, M. B., and R. A. Dudley. 2007. "Pay-for-Performance: Will the Latest Payment Trend Improve Care?" *Journal of the American Medical Association* 297 (7): 740–4.
- Securities and Exchange Commission (SEC). 2008. "The Investor's Advocate: How the SEC Protects Investors, Maintains Market Integrity, and Facilitates Capital Formation" [accessed on April 17, 2009]. Available at <http://www.sec.gov/about/whatwedo.shtml>
- Sexton, J. B., R. L. Helmreich, T. B. Neilands, K. Rowan, K. Vella, J. Boyden, P. R. Roberts, and E. J. Thomas. 2006. "The Safety Attitudes Questionnaire: Psychometric Properties Benchmarking Data, and Emerging Research." *BMC Health Services Research* 6: 44, doi:10.1186/1472-6963-6-44.
- Shojania, K. G., B. W. Duncan, K. M. McDonald, and R. M. Wachter. 2002. "Safe But Sound: Patient Safety Meets Evidence-Based Medicine." *Journal of the American Medical Association* 288 (4): 508–13.
- Shortell, S. M., and S. J. Singer. 2008. "Improving Patient Safety by Taking Systems Seriously." *Journal of the American Medical Association* 299 (4): 445–7.
- Singla, A. K., B. T. Kitch, J. S. Weissman, and E. G. Campbell. 2006. "Assessing Patient Safety Culture: A Review and Synthesis of the Measurement Tools." *Journal of Patient Safety* 2 (3): 105–15.
- U.S. Government Accounting Office. 2006. *Hospital Quality Data. CMS Needs More Rigorous Methods to Ensure Reliability of Publicly Released Data*. Washington, DC: United States Government Accountability Office.
- Wachter, R. M. 2004. "The End of the Beginning: Patient Safety Five Years after 'To Err Is Human.'" *Health Affairs* (W4): 534–45.
- Werner, R. M., and E. T. Bradlow. 2006. "Relationship between Medicare's Hospital Compare Performance Measures and Mortality Rates." *Journal of the American Medical Association* 296 (22): 2694–702.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.