
Editorial

The Performance of Performance Measurement

Continued increases in health care expenditures, coupled with incontrovertible evidence of a substantial gap between the best possible care and that which is routinely delivered, have motivated a growing interest in the use of performance measurement to drive clinical improvements and inform choices made by purchasers and consumers. In the late 1980s, New York State led the way by publishing reports of cardiac procedure outcomes of hospitals (initially) and then individual physicians. Since then, the National Committee for Quality Assurance, a private sector organization that accredits health plans, launched public reporting of health plan performance in the early 1990s and developed a broad consensus regarding performance measurement through HEDIS, an effort that has continued to expand and mature. The Agency for Healthcare Research and Quality (AHRQ) supported the development and implementation of the CAHPS family of surveys, which have become a *de facto* standard for assessing patient experience of care, and the Joint Commission developed and implemented a performance measure set for hospitals.

Purchasers have also played a pivotal role. The Medicare program provided essential leadership and support for performance measurement, and leading private sector purchasers such as the Leapfrog Group have continued to challenge providers to make performance assessment a core part of doing business. Today, the public can consult websites to compare the performance of hospitals, nursing homes and home health care—and information on physician performance is on the horizon.

These efforts can and should be viewed as a success for health services researchers, whose contributions have shaped the intellectual landscape. As a result of health services research, we have learned that public reporting is associated with improvements in care, that better performance is inversely related to mortality, and that we lack rigorous metrics for some dimensions of performance, such as efficiency (Epstein 1998; Marshall et al. 2000; Hibbard, Stockard, and Tusler 2005; Jha et al. 2007). AHRQ's mandate to submit annual reports to the Congress on quality and disparities in health care is just

one manifestation of the recognition that quality is now seen as the organizing principle for serious health system reform. Performance measures are an essential tool to assess whether and how health care delivery is improving, and whether or not the fruits of biomedical science and health services research are translated as improved health care and health.

This brief and highly selective overview reflects increased interest and activity in performance measurement for selection of providers and quality improvement, as well as the power of incremental strategies to propel improvement efforts. Although the number and types of measures and measurement systems have proliferated, there have been few efforts to compare the validity of different approaches or to gauge how well different approaches match one or more of the desired goals of measurement. In short, how well do performance measures perform with respect to capturing the underlying construct?

Thus, celebration of tangible progress should be tempered by humility regarding how much we have yet to learn and clearly recognize that current measures reflect a strong emphasis on feasibility (e.g., low costs of collection and ease of availability of data) and produce limited information on performance. The subsequent limitations of the measures have made it difficult to get people to pay attention to, and use, performance measures.

In this issue of the journal Kerr and colleagues report on an unprecedented comparison of different approaches to performance measurement in VA hospitals. Their study compared the results of three approaches to performance measurement: focused explicit or condition-specific measures which measure adherence to a normative professional standard (e.g., the percentage of patients with a heart attack who receive beta blockers); global explicit measures, a composite derived from summarizing multiple focused explicit metrics; and implicit measures, which take advantage of professional judgments of overall quality of care.

Each approach has unique strengths and limitations. Focused explicit approaches resonate with clinicians and can inspire focused improvement within a specific area, but they are not designed to result in improvements in broader areas. Global explicit approaches, used in the oft-cited study reporting that Americans receive recommended care only 54.9 percent of the time, offer the possibility of selecting among providers, but cannot provide a clear roadmap for improvement (McGlynn et al. 2003). Implicit approaches take advantage of professional knowledge of important clinical nuances but frequently exhibit poor inter-rater reliability.

The results of the comparison presented by Kerr and colleagues are mostly encouraging. They found a high degree of convergence across

summary measurement systems for summary measures of quality as well as substantial agreement across three approaches for diabetes and preventive care. However, this level of agreement was not found in all areas, such as hypertension. The authors note that this finding cannot be entirely explained by differences in the availability of evidence or professional standards and speculate that the combination of few measures per person and ceiling effects may have reduced the amount of useful information obtained. Clarity regarding what to measure and a robust understanding of the statistical properties of the measures used are both essential. Their results are even more impressive because they used a novel strategy to adjust for measurement error attributable to differences in the number of items, which has received very little attention previously.

The findings presented of the study by Kerr and colleagues are important for several reasons. First, we cannot take full advantage of performance assessment without a clear-eyed appreciation of strengths and limitations. As Kerr's analysis cogently demonstrates, the details matter. Second, the use of performance measures as a tool to make informed choices and to drive much-needed improvements depends not only on their validity but also on their replicable implementation. Notwithstanding widespread use of report cards, if reports are not reliable, the promise of performance measurement as a tool to increase value will fall short of expectations. As health services researchers know well, the data source and quality of those data influence both the accuracy and the precision of performance measures. Biased or very "noisy" measures can have significant consequences through an impact on providers' reputations (public reporting) or by misallocation of resources for improvement.

An important and sometimes unspoken rate-limiting step in performance measurement to date has been the feasibility of identifying available data sources. As more providers adopt electronic health records and policies to support interoperability that facilitates sharing of health information across settings of care, there is a unique opportunity to make sure that these technologies support and enhance the quality enterprise. Health information technology (HIT) is likely to decrease the cost of data collection significantly and to enhance the specificity and completeness of clinical data available for performance assessments.

These features alone have enormous promise but are insufficient to achieve the full potential of performance assessment. The very real possibility that information can follow patients as they traverse multiple parts of the health care system could increase the accuracy of measurements by including

relevant information not always accessible even through computerized records (e.g., the information in an outpatient office that an individual should never receive a beta blocker). But this capability also requires that performance assessment be conducted using data from multiple sources, with implications for governance, privacy, and reliability of methods. As efforts to align reimbursement with quality such as pay for performance continue to proliferate and evolve, the need for clear audit trails that exceed current efforts will also be required.

More importantly, the enhanced capacity to report on performance will not of itself add value to clinical care or result in substantial improvements in that care. HIT also brings the promise of using the same data required to assess performance to create decision support tools that help facilitate delivery of just-in-time information to ensure better care as well as better reports. Translating this promise into reality will require serious study of work flow, business processes, and human and other factors in order to redesign care processes, which are not to be found in the box with the HIT's computer. This work will require the best efforts of health services researchers, working closely with clinicians, patients and health care leaders.

Performance measurement today, despite its long history, is in a relatively early stage of development and implementation. Viewed as a transformative tool that can address the current shortcomings in health care delivery, the imperative for action is unlikely to abate. Kerr and colleagues have made an important contribution by demonstrating the relative advantages of different approaches to assessment. There is an enormous need for the health services researchers to help health care move beyond 'managing what is (easily) measured' to developing the intellectual and conceptual architecture that will help us close the yawning gap between the promise of health care today and current performance.

Carolyn Clancy, M.D.

REFERENCES

- Epstein, A. M. 1998. "Rolling Down the Runway: The Challenges Ahead for Quality Report Cards." *Journal of the American Medical Association* 279 (21): 1691-6.
- Hibbard, J. H., J. Stockard, and M. Tusler. 2005. "Hospital Performance Reports: Impact on Quality, Market Share, and Reputation." *Health Affairs (Millwood)* 24 (4): 1150-60.

- Jha, A. K., E. J. Orav, Z. Li, and A. M. Epstein. 2007. "The Inverse Relationship Between Mortality Rates and Performance in the Hospital Quality Alliance Measures." *Health Affairs (Millwood)* 26 (4): 1104–10.
- Marshall, M. N., P. G. Shekelle, S. Leatherman, and R. H. Brook. 2000. "The Public Release of Performance Data: What Do We Expect to Gain? A Review of the Evidence." *Journal of the American Medical Association* 283 (14): 1866–74.
- McGlynn, E. A., S. M. Asch, J. Adams, J. Keeseey, 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.