

The current state of performance measurement in health care

JEROD M. LOEB

Division of Research, Joint Commission on Accreditation of Healthcare Organizations, Oakbrook Terrace, IL, USA

Abstract

Health care quality measurement is at least 250 years old. While the names and faces of the measures and measurees have changed, the intent of such measurement, i.e. obtaining data and information bearing clinical outcomes, has not changed over the years, and nor have the challenges associated with the measurement of quality in health care.

Measurement is not a neutral activity. It evokes considerable anxiety and frustration among all concerned: those who are being measured, those who are doing the measuring, and those who are seeking the data for a variety of purposes. There is little agreement on the philosophy of measurement, on what to measure, on whether or how to adjust for what the patient brings to the clinical encounter, on how data should be analyzed, or on how to report the data; and of course the ultimate questions relate to the value of measurement. Measurement adds new costs to the health care delivery system.

Tangible improvements in care, directly attributable to performance measurement, are repeatedly being documented. However, the jury has not yet returned a verdict on whether performance measurement data are being used by stakeholders for better decision-making.

Keywords: measurement, performance measurement, quality

The measurement of health care quality is a seemingly simple endeavor, beginning with a decision on what to measure, identifying the proper measures along with their respective data sources, and culminating in the analysis, aggregation, understanding, and dissemination of the results. However, as with most deceptively simple activities or tasks, the difficulty lies in the details.

The various perspectives

A significant part of the challenge in measuring performance in health care derives from the disparate nature and variable perspectives represented among the key stakeholders. For example, the professional perspective—held by some physicians, nurses, and scientists who are trained to think critically and analytically—is that performance measurement is simply too fraught with problems to be of much practical use. This viewpoint is based on the seeming myriad of unresolved issues associated with the measurement process itself [1], including ambiguous data element definitions, complex algorithms, risk adjustment techniques, and uncertainties about the psychometric properties of the measures themselves. In addition, performance measurement is seen, particularly by those focused on the operational and financial aspects of health care, as a costly endeavor that, in the absence of the availability of electronic data capture, does not provide

sufficient cost benefit. Another confounding factor often articulated is that the easiest (i.e. cheapest) aspects of care to measure are often those least important in a quality improvement context.

On the other hand the public policy perspective, often held by regulatory officials and large corporate purchasers, is that performance measurement should be an integral part of the everyday business of health care. These stakeholders are willing to let imperfect measures be used and the desire to move ahead with measurement, even though accuracy may be somewhat questionable, is the key driving force. In many respects, demands by purchasers and regulators for demonstrable evidence of quality, and demands for accountability, have become a major driver (if not the major driver) responsible for the burgeoning work in performance measurement over the past decade or so. No longer can health care organizations afford to remain complacent and assume that stakeholders understand that quality care is being provided; rather, evidence is required.

Finally, the consumer perspective is to desire clear answers to vexing questions, even though sufficiently enlightened answers may not be available in many cases. Whether consumers are asking the right questions is an open issue in itself. Consumers resonate with traditional measures of patient satisfaction (i.e. appointment waiting times, communication among caregivers); however, such measures are not standardized, are rarely in the public domain, can be significantly influenced by

Address reprint requests to J. M. Loeb, Division of Research, Joint Commission on Accreditation of Healthcare Organizations, 1 Renaissance Boulevard, Oakbrook Terrace, IL 60181, USA. E-mail: jloeb@jcaho.org

sample size and response rate characteristics, and sometimes cannot be linked directly with the delivery of high quality health care services. These varying perspectives are often at cross-purposes with each other and, when taken together, demonstrate how difficult it can be to achieve consensus about performance measurement.

Some essential history

Performance measurement in health care is not a new endeavor. In many ways, such measurement has been in existence for generations. While there is evidence that patient outcome data were being collected at the hospital of the University of Pennsylvania as early as the middle of the 18th century, McIntyre *et al.* [2] and Colton [3] describe several distinct periods in the evolution of quality improvement in health care:

1. 1850–1915: the Industrial Revolution and scientific management
2. 1915–1935: bureaucracies and organizations
3. 1935–1960: human resources, statistical process control, and expansion of health care
4. 1960–1980: status quo
5. 1980 to present: the quality health care organization.

By the middle of the 19th century, Florence Nightingale was collecting mortality data and infection rates for the principal hospitals in England during the Crimean War. However, it was not until Dr Ernest A. Codman began to conceive his ‘end results hypothesis’ at the Massachusetts General Hospital in the early part of the 20th century that performance measurement began to emerge as a viable tool for assessing health care quality. Codman actually proposed a detailed system of patient records, including post-discharge follow-up, identification of the ‘best’ and ‘worst’ surgeons based on the actual results of their care, and patient access to the results of various treatments, including interhospital comparisons [4].

Codman’s ideas about performance measurement were clearly revolutionary in the early 1900s. In fact, Codman was considered quite radical by the ever-conservative Boston, Massachusetts medical establishment. When Codman proposed his ‘end results hypothesis’ at a meeting of the local medical society in 1915, he was audacious enough to present a now famous cartoon in which he depicted the trustees of the Massachusetts General Hospital as concerned only with the financial remuneration associated with surgery, not the ultimate outcomes of those surgical procedures. Codman’s behavior was so radical that at one point he resigned as an attending surgeon and reapplied to become Department Chair, with the notion that only he was capable of demonstrating the results of care provided. Needless to say, he was turned down. Codman also opened his own hospital (the ‘End Results Hospital’) in Boston, but the facility was forced to close within 2 years because of a paucity of patients [4]. In many ways, Codman’s unique ideas and principles still underpin performance measurement activities a full century later.

In the realm of more recent history, perhaps no individual has done more to advance modern thinking about the measurement of quality in health care than Avedis Donabedian. Donabedian proposed a three-element model associated with quality measurement, including structure (the characteristics associated with a health care setting), process (what is done in the health care setting), and outcome (the ultimate status of the patient after a given set of health care interventions [5]). In fact, it is this paradigm that essentially established the modern field of outcomes research [6].

The challenge of differing philosophies

In today’s complex health care environment, there are a variety of salient issues and challenges associated with the measurement of health care quality. Foremost among these issues is the notion that while most people are in favor of measurement, few are comfortable being measured. Measurement provokes considerable angst, frustration, and worry among those being measured, and often also among those doing the measuring. This is especially true since not all decision making in medicine is grounded in scientific fact and clinical evidence (i.e. opinion plays a significant role in medical decision-making). While evidence-based clinical practice guidelines exist in a variety of specialties and subspecialties in medicine, consistent evidence suggests that adherence to guidelines is poor. There are well documented and large geographic disparities in care for particular services in medicine that are likely the result of the predominance of opinion over evidence [7].

Unfortunately there is little consistency in the various philosophies and approaches to performance measurement in the contemporary health care environment. One school of thought seeks to ascertain which data may be available within a given type of organization (i.e. hospital, nursing home) followed by questions about what might be measured using the available data. This approach tries to minimize the collection of new data by concentrating on data already available. While this is, of course, a cost-effective application of performance measurement, it fails to fulfill at least one basic tenet of good organizational management, namely the upfront development of strategic measurement goals as a guiding principle associated with a robust quality improvement strategy. Asking the questions after the data collection has already occurred also has the disadvantage of eliminating the opportunity for the gathered data to be useful in a concurrent manner.

A second approach to performance measurement incorporates asking first what needs to be measured, and then ascertaining whether sufficient internal or external data are available to answer the question posed. For example, a hospital may be interested in determining its average time to implementing percutaneous coronary interventions in patients with acute coronary syndromes. In this case, the key issues include the accuracy and timeliness of the diagnosis and the rapidity of intervention. The hospital may already be collecting some of the data required to answer these questions, but it is likely that additional data elements will be needed. Thus, once appropriate, clinically relevant questions have been developed, it is

a relatively simple task to determine whether any of the data required can be derived from extant data sources. If not, additional data elements can be effectively added and key questions answered.

However, the most rational approach would be first to define strategic measurement goals for the organization (i.e. what needs to be measured, why, and how), and then determine whether reliable data exist to support such measurement [8]. As noted earlier, additional data elements can always be added but, since new data collection can be a costly undertaking, it is important to determine whether the benefits of the additional data collection outweigh the costs.

Standardizing what to measure

The central issue in performance measurement remains the absence of agreement with respect to what should be measured. Not everything in health care can or should be measured. Certainly within a few specific clinical focus areas (i.e. acute myocardial infarction, heart failure, community-acquired pneumonia), there has been considerable work pertaining to key processes of care that are inextricably linked to salutary outcomes. In these few instances, there is substantial evidence to support the use of clusters of standardized process measures as key indicators of health care performance. However, beyond these 'low hanging fruit' there is significant controversy about what to measure. Single measures are, in most cases, too limited in scope to be of much use to stakeholders, while at the other extreme, large numbers of measures are likely to be too cost-prohibitive to those collecting the data or to those seeking clear and simple answers to a variety of questions. While it is apparent that multiple measures are needed to 'weave a portrait of care', few measures (even those in common use) have been adequately tested for reliability, validity, and data collection effort. Moreover, the existence of multiple measures that appear at face value to ostensibly assess the same thing, but are defined differently, is a significant issue. There are many examples of extant measures that purport to measure the same thing (e.g. beta blockers prescribed at discharge for patients with acute myocardial infarction, cesarean section rate, etc.) but that are defined differently. These differences relate to variations among the populations included or excluded, risk adjustment algorithms, and so on.

Measures are often needed for a variety of related, but saliently different clinical situations. For example, when attempting to assess the quality of the medication use process, measures may be required for each of the following situations: (i) ideal candidates for a particular drug who *did not* receive the drug (i.e. indicative of poor quality care); (ii) ideal candidates for a particular drug who *did* receive the drug (i.e. indicative of good quality care); (iii) poor candidates for a particular drug who *did not* receive the drug (i.e. indicative of good quality care); and (iv) poor candidates for a particular drug who *did* receive the drug (i.e. indicative of poor quality care). It is clear that each of these potential scenarios is certainly possible, but the availability of strongly linked measures for each is often

an insurmountable issue. Also problematic is the issue of how to statistically adjust for what the patient brings to the clinical encounter (see below). Moreover, in the absence of an information technology infrastructure (including interoperability) that can support the collection of these types of data as a byproduct of clinical care rather than as an intensive data collection exercise in itself, requiring considerable chart abstraction, it is unlikely that much progress in the field of performance measurement will ensue. This is perhaps the most critical impediment to the diffusion of performance measurement in routine use.

Data quality and data integrity

Data quality and data integrity are integral to the judicious use of data in any health care context. Data quality can be defined as the extent to which data are valid, reliable, complete, and timely. Although there is undoubtedly agreement that data quality is an essential element of performance measurement, the quality of data collected is often quite variable, as is the interpretation of the data. For example, data quality can be affected by a variety of extraneous influences, including ambiguous data element definitions, misinterpretation or misapplication of data element definitions, incomplete or missing data (a result of the fact that specific data elements are often difficult to locate in medical records or that medical records are poorly documented), reluctance of organizations to change extant processes, variations in coding practices within and across organizations, lack of ongoing support by measure developers, and a lack of external audits of collected data. The science behind the measures is also important with respect to data quality, but, as noted earlier, few measures have been tested adequately with respect to their psychometric properties. Often, measures have been developed by performance measurement experts with little clinical experience. On the other hand, clinicians have developed a variety of measures that may make sense clinically, but are not strong psychometrically.

Risk adjustment

Accounting for patient factors is another controversial topic in performance measurement. It is important for outcome-based performance measures to account for potential comorbidities that the patient may bring to the clinical encounter so as not to unfairly or inappropriately depict a health care facility as an outlier. It is clear that a variety of factors affect clinical outcomes. These factors may be clinician-related (i.e. clinician knowledge, technical and interpersonal skills, judgment), organization-related (i.e. availability and management of human resources, capital resources, information resources, financial resources, systems and processes for governance, management, clinical care, and support services), community-related [i.e. structure of community health care system, amount and allocation of health care resources in the community, other factors potentially affecting health (environmental pollution,

workplace safety)], or patient-related (i.e. normal biological variation, severity of illness, comorbidities, patient's preference for treatment or lack of treatment, patient's involvement in treatment).

Risk adjustment or stratification techniques can be used to account for what a patient might bring to a clinical encounter, but significant questions remain with respect to which factors should be used in the calculation, which technique(s) are most appropriate to account for these factors (i.e. logistic regression), and whether such adjustment may provide a false sense of security. Moreover, the potential role of adjusting for socioeconomic risk factors is a difficult and controversial issue in health care today.

Data analysis and reporting of health care data

Another salient and often overlooked issue relates to the analysis of the performance measurement data. In most cases, it is important to track data longitudinally within the same health care organization to understand whether key processes of care are stable and in control. On the other hand, it is also important to understand whether performance in a given focus area in one institution is equal to, better than, or worse than performance across a cohort of similar organizations—in the same community or across the world. Such data analysis requires a modicum of sophistication and, often, the use of a suite of tools (i.e. control and comparison charts) to turn data into useful and actionable information. While statistical process control techniques have been extant in the manufacturing sector for decades, the use of such tools in health care is only beginning to become mainstreamed [9]. Further confounding the analysis of performance measurement data is the question of sample size. While statistical significance is of course important, clinical significance is equally important. Thus, it is possible that certain data may be clinically but not statistically significant or vice versa. Caution must be used in interpreting data where sample size considerations may be an issue. At the same time, it is important to note that misuse of data—even good and well intentioned data—can have profound economic and social implications, irrespective of whether the data refer to individual performance or organizational performance.

And finally, how should health care data be reported? There is general agreement about the need for public reporting, but there is little agreement on how such data should be reported, whose health care 'report card' is best, and whether different audiences require different reporting formats [10]. Report cards have existed for many generations in education. However, even today within the educational world, there is considerable disagreement on what to report, when to report, and how to report. Some school districts utilize relatively traditional grading schemes, whereby students receive letter grades designating performance (A–F), while others use numerical designations (0–100) or even a variety of adjectives to describe performance associated with particular skill sets ('emerging', 'progressing', 'developed'). While such report cards purport to provide the student and parent with useful,

actionable information, it is still unclear whether this goal is being achieved. In health care, performance reporting is still in its infancy. Further, even when detailed information is made available to consumers, it is rare that such data are being actively used to guide health care decision-making. In the final analysis, it may be found that stakeholders are simply not receiving reports in viable formats that are suitable for their needs.

In the United States, the National Committee for Quality Assurance, which accredits managed care plans, has been publicly reporting HEDIS (Health Plan Employer Data and Information Set) data since 2000. The HEDIS data set incorporates standardized performance measures pertinent to a variety of public health issues, including cancer, heart disease, smoking, asthma, and diabetes. The data set also includes a standardized survey of consumers' experiences associated with care, including service, access, and claims processing (<http://www.ncqa.org/programs/HEDIS>). Similarly, data on the quality of US nursing home care are publicly available [11].

A variety of data reporting formats exists today. For example, in the United States, the Centers for Medicare and Medicaid Services (<http://www.medicare.gov>) provides access to an online database on nursing home care that reflects performance on a variety of specific measures, including percentage of residents with loss of ability in basic daily tasks, percentage of residents with pressure sores (risk adjusted and non-risk adjusted), percentage of residents with pain, percentage of residents in physical restraints, and percentage of residents with infections. The data are displayed in a visually appealing format that includes data on individual nursing home performance, with comparisons at the state and national level. Users can configure searches by state, county, postal code, etc. Similarly, in the United States, a joint voluntary hospital performance reporting initiative was recently announced that represents a collaboration of the Centers for Medicare and Medicaid Services, American Hospital Association, Federation of American Hospitals, Association of American Medical Colleges, Joint Commission on Accreditation of Healthcare Organizations, National Quality Forum (NQF), and the Agency for Healthcare Research and Quality. This initiative provides for the public release of hospital-specific data derived from a total of 10 measures pertinent to the care of patients with acute myocardial infarction, heart failure, and community-acquired pneumonia. A pilot project funded by the federal government is now underway in three states to determine the reporting formats that will resonate most with various stakeholders, including consumers. The first data are expected to be available later this year. Ultimately, this initiative will also include other clinical measures and survey data dealing with patient experience of care. A standardized, public domain 'patient experience of care' instrument is currently being developed and tested.

While many efforts are underway to analyze and report organizational-level data, there is great interest in further drill-down to the level of individual physician performance. Although publicly available data pertinent to individual physician performance is just beginning to become available, it is still fraught with difficulty, mainly as a consequence of issues

associated with defining adequate sample size and the need for standardized measures. Such reports are now becoming available in many communities [12]. Blumenthal and Epstein [13] summarize some of the efforts related to reporting performance at the level of individual physicians. More recently, Landon *et al.* [14] discussed the technical barriers inherent in assessing physician performance and concluded that some of these barriers (including lack of evidence-based measures for many specialties, definition of thresholds for acceptable care, sample size, confounders in process and outcome measurement, representativeness, and feasibility and cost) may be insurmountable, at least at present.

The future

Performance measurement is increasingly becoming an integral part of the fabric of health care. Demands for data and information for multiple purposes, including performance improvement, accountability, and stakeholder decision making, are likely to increase in the future. It is imperative that performance measures are standardized so that data collection efforts can be minimized. In the United States, the NQF is the key entity responsible for developing the toolbox of standardized, evidence-based measures. The NQF was created in 1999 as a public sector–private sector partnership, to develop and implement a national strategy on health care quality measurement and reporting. To date, 39 standardized measures have been approved by the NQF in several priority areas, including acute coronary syndrome, heart failure, pneumonia, patient safety, pregnancy/childbirth/neonatal conditions, surgical complications, pediatric conditions, and smoking cessation. However, while the NQF can approve and promote standardized performance measures, the ultimate leverage for use of the measures must come from elsewhere, most notably regulatory bodies, accrediting bodies, and purchasers of care.

In the international arena, it is clear that the availability of standardized performance measures that are applicable across countries remains more of a hope than a reality [15]. At the same time, the Maryland Hospital Association Quality Indicator Project currently has >200 hospitals in nine countries participating in data collection on standardized measures focusing on in-patient mortality, perioperative mortality, management of labor, unscheduled readmissions, and unscheduled returns to the intensive care unit, operating room, and emergency department [15]. Similarly, beginning in 2004, hospitals accredited by Joint Commission International will be collecting data—on a voluntary basis to start with—on indicators covering a variety of similar clinical focus areas. Clearly, the key challenge in designing indicators that transcend country borders is applicability and value to local users. While internationally standardized performance measures are highly desirable, significant differences across countries in terms of health care cultures, data availability, practice guidelines, etc. make the design issues highly problematic.

Ultimately, it must be accepted that performance measurement adds new costs to the health care equation. The true value associated with performance measurement will only be

realized when it can be shown conclusively that tangible improvements in care are being engendered as a byproduct of the measurement, and that key stakeholders are making data-driven decisions. That day has almost arrived.

References

1. Eddy DM. Performance measurement in health care. *Health Aff* 1998; **17**: 7–25.
2. McIntyre D, Rogers L, Heier EJ. Overview, history, and objectives of performance measurement. *Health Care Financ Rev* 2001; **22**: 7–21.
3. Colton D. Quality improvement in health care: conceptual and historical foundations. *Eval Health Prof* 2000; **23**: 7–43.
4. Mallon WJ. *Ernest Amory Codman: The End Result of a Life in Medicine*. Philadelphia, PA: W.B. Saunders Company, 2000.
5. Donabedian A. Evaluating the quality of medical care. *Milbank Mem Fund Q* 1966; **44**: 166–206.
6. Perrin EB. Some thoughts on outcomes research, quality improvement, and performance measurement. *Med Care* 2002; **40**: III89–III91.
7. Pilote L, Calif RM, Sapp S *et al.* Regional variation across the United States in the management of acute myocardial infarction. *N Engl J Med* 1995; **333**: 565–572.
8. Houser J. Using the data management process to improve performance. In *Managing Performance Measurement Data in Health Care*. Oakbrook Terrace, IL: Joint Commission Resources, 2001.
9. Carey RG, Lloyd RC. *Measuring Quality Improvement in Health Care: A Guide to Statistical Process Control Applications*. Milwaukee, WI: ASQ Press, 2001.
10. Krumholz HM, Rathore SS, Chen J, Wang Y, Radford MJ. Evaluation of a consumer-oriented internet health care report card. The risk of quality ratings based on mortality data. *J Am Med Assoc* 2002; **287**: 1277–1287.
11. Mor V, Berg K, Angelelli J, Gifford D, Morris J, Moore T. The quality of quality measurement in U.S. nursing homes. *Gerontologist* 2003; **43**: 37–46.
12. Hofer TP, Hayward RA, Greenfield S, Wagner EH, Kaplan SH, Manning WG. The unreliability of individual physician ‘report cards’ for assessing the costs and quality of care of a chronic disease. *J Am Med Assoc* 1999; **281**: 2098–2105.
13. Blumenthal D, Epstein AM. Quality of health care. Part 6: the role of physicians in the future of quality management. *N Engl J Med* 1996; **335**: 1328–1331.
14. Landon BE, Normand ST, Blumenthal D, Daley J. Physician clinical performance assessment: prospects and barriers. *J Am Med Assoc* 2003; **290**: 1183–1189.
15. Kazandjian VA, Matthes N, Wicker KG. Are performance measures generic? The international experience of the Quality Indicator Project. *J Eval Clin Pract* 2003; **9**: 265–276.

Accepted for publication 24 September 2003

