Assessing Hospital Quality: A Review for Clinicians

Physicians frequently react to “quality reports” with suspicion, regarding attempts to measure “quality” as thinly disguised efforts at cost reduction or marketing salvos engineered by managed care organizations. Skepticism about quality reports is reinforced by the confusion in many report cards, which often present a handful of relatively crude clinical outcomes (e.g., mortality and readmission rates) mixed with measures highlighting resource use (e.g., lengths of stay).

Nevertheless, physicians must acknowledge the need for quality measurement in American health care, given the widespread gaps between recommended and actual practice and unimpressive performance in terms of basic health care outcomes. While quality reports to date have demonstrated limited impact, interest in them remains high, especially in light of the recent publicity regarding medical errors. Widely used quality reports (e.g., the Health Employer Data Information Set [HEDIS]) have focused largely on preventive strategies most relevant to ambulatory patients in health plans. In contrast, the attention given to the challenges and opportunities involved in measuring the quality of inpatient care has been far less comprehensive.

In this article, we present clinicians with a conceptual scheme for assessing hospital quality reports. We illustrate the dominant paradigm of measuring quality in terms of structure, process, and outcome by using examples from the clinical and health services research literature to illustrate advantages and disadvantages for each category. We emphasize comparative quality measurement (as with hospital quality report cards), rather than the development of indicators designed to fuel internal quality improvement. In addition, because productive quality measurement clearly must stimulate quality improvement, we touch on the interrelationship between these two activities.

A Definition of Quality

One often-cited definition describes quality as the “degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge.” In practice, distinctions between quality and cost have often been blurred. Cost-cutting efforts and “quality assurance” activities have often been regarded as identical. Thus, a growing trend recognizes the distinction between quality and economy, assessing “value” (quality divided by cost) as the preferred metric for informing health care purchasing and policy decisions.

Structural Measures of Quality

Structure is the setting in which care occurs and the capacity of that setting to produce quality. Table 1 presents traditional examples of structural measures related to hospital quality. More recent candidate measures include the adoption of organizational models for inpatient care (e.g., closed intensive care units and dedicated stroke units) and the presence of sophisticated clinical information systems.
The growing use of patient volume as a quality indicator reflects an extensive literature that documents superior outcomes for hospitals and physicians with higher patient volumes. The literature suggests that substantial reductions in mortality rate might result from regionalizing treatment for certain high-risk conditions. For example, recent analysis suggests that transferring the care of patients with acute myocardial infarction (MI) from hospitals in the lowest-volume quartile to those in the highest-volume quartile would save 2.3 lives per 100 patients. The corresponding "number needed to treat"—50—falls within the range of many accepted therapeutic interventions.

When volume–outcomes relationships reflect a "practice-makes-perfect" effect, using volume to assess quality may be reasonable. However, such relationships could also reflect "selective referral," which may occur when the community has informal knowledge of the superior quality of these high-volume centers. Moreover, the degree to which volume-based performance measurement would promote quality improvement remains unclear. One study of cardiac surgery indicated improved outcomes as a result of increased patient volume, but a study of pediatric trauma centers suggested that, beyond a certain threshold, increases in volume place a strain on provider resources and worsen outcomes. Diverting patients to large referral centers also has important health policy implications and may decrease patient satisfaction.

Similar arguments apply to specialist care as a quality measure. For example, although cardiologists achieve better outcomes than generalists in caring for patients with acute MI, much of this benefit is derived from the performance of established care processes. The case for measuring these processes directly is strengthened by the finding that even cardiologists fail to provide proven therapies to many eligible patients. In contrast, a comparable set of proven interventions does not exist for inpatient management of the complications of HIV infection. Thus, management by high-volume HIV "specialists" may provide the best available marker for superior inpatient HIV care.

Accreditation listings and billing data permit easy ascertainment of teaching status and patient volumes, making such structural measures extremely efficient quality indicators. However, existing databases do not characterize the organizational models and information systems that hospitals use. Measures targeting these hospital characteristics may thus prove difficult to implement.

### Process Measures of Quality

Process measures depict the interior of the hospital "black box," allowing measurement of the care patients actually receive. The example of MI care best illustrates the potential power of process-based quality measurement. Clinical research has firmly established the benefits of acute reperfusion, therapy with aspirin and \( \beta \)-blockers, and certain other processes of care. Numerous studies have documented delays in translating this research into practice and the adverse conse-

<table>
<thead>
<tr>
<th>TABLE 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Structural Measures of Quality</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Examples</th>
<th>Hospital teaching status, ownership, availability of sophisticated technologies, and qualifications of hospital personnel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient volume</td>
<td>Efficient measurement</td>
</tr>
<tr>
<td>Advantages</td>
<td>Potential for capturing aspects of care not easily described with other measures, especially in complex situations involving multiple processes of care</td>
</tr>
<tr>
<td>Disadvantages</td>
<td>Very &quot;blunt&quot;; more detailed measures often exist</td>
</tr>
<tr>
<td></td>
<td>Promising indicators (e.g., organizational models of care and the presence of sophisticated clinical information systems) may not be discernible from existing databases</td>
</tr>
<tr>
<td></td>
<td>Many are not &quot;actionable&quot; on a practical basis (e.g., hospital teaching status and ownership)</td>
</tr>
</tbody>
</table>

The example of MI care best illustrates the potential power of process-based quality measurement. Clinical research has firmly established the benefits of acute reperfusion, therapy with aspirin and \( \beta \)-blockers, and certain other processes of care. Numerous studies have documented delays in translating this research into practice and the adverse conse-
quences of these delays. The Cooperative Cardiovascular Project confirmed that variations in therapeutic and preventive processes explain differences in acute MI mortality rates between hospitals; the project also documented improved outcomes following improvements in process performance.52

Unfortunately, the literature provides surprisingly few examples of process–outcome links as compelling as those associated with acute MI care. For example, despite the extensive literature documenting the benefits of postoperative thromboembolism prophylaxis, the relative advantages of various pharmacologic and mechanical methods remain unclear. A process measure that captured “some form of thromboembolism prophylaxis” would probably not reveal significant underuse by providers. Nonetheless, any variation noted would be significant, and negative consequences of promoting this process are difficult to imagine. The same cannot be said for perioperative antibiotic prophylaxis. While the literature demonstrates improved patient outcomes for many surgical procedures, the evidence supporting a particular antibiotic regimen is not as strong. In fact, a recent systematic review concluded that appropriately timed single-dose prophylaxis is as effective as multidose regimens. Thus, to develop a measure that captures perioperative antibiotic use, the extent of two potential quality problems—postoperative infections related to underuse of prophylaxis and increased adverse drug events and microbial resistance from antibiotic overuse—would need to be compared.

This example highlights the point that overuse and misuse of available therapies represent quality problems that are at least as important as those associated with underuse. However, overuse and misuse have generally proved more difficult to target. Despite marked variations in the use of many common procedures, it is still difficult to distinguish between overuse and inappropriate use. In fact, repeated cesarean section represents the only widely employed measure of overuse, and growing concerns about the safety of vaginal birth after cesarean section may lead to its abandonment. Endarterectomy for asymptomatic carotid stenosis presents a more promising potential target for a measure of overuse. Population-level analysis indicates that all providers, including hospitals enrolled in the Asymptomatic Carotid Atherosclerosis Study, have significantly higher mortality rates than those reported in the trial. Inappropriate patient selection seems to be the most plausible explanation. At the individual patient level, judgments of appropriateness for endarterectomy remain highly sensitive to estimates of surgical risk. When such estimates can be generated more precisely, appropriate patient selection for endarterectomy may provide an accurate quality measure.

Despite these complexities, process measures remain attractive quality indicators because of their sensitivity in detecting quality problems and their direct connection to quality improvement strategies. Process-based measurement is particularly well-suited to medical conditions, for which the literature supplies more candidate processes and in which technical skill in process performance plays a minimal role.

### Outcome Measures of Quality

Although such outcomes as mortality, morbidity, and patient satisfaction represent the bottom line in defining quality, these measures have problems (Table 3). Even for common conditions, it may take years to detect differences in outcomes between hospitals. Moreover, such differences may reflect patient factors, variations in admission practices, or chance rather than differences in quality of care. On the other hand,
appropriate structure- or process-based measures do not exist for many conditions, and certain conditions remain particularly suited to outcome-based performance measurement.

Two immediate questions to ask about any candidate outcome measure are: Is the measure really an outcome? Does it relate to medical care delivered in the setting in question? Length of stay is a widely used quality measure despite the fact that it primarily involves resource use. Rates of early readmission more plausibly represent a combined clinical and financial outcome, although analyses of their relationship to quality have produced conflicting results.

More complicated questions regarding candidate outcome measures relate to perverse incentives and opportunities for “gaming.” Perverse incentives may arise from the criteria used to define target patient populations. For example, restricting inpatient mortality to deaths that literally occur in the hospital allows hospitals to lower their mortality rates simply by discharging patients to die at home or in other institutions.

Complication rates that capture only in-hospital events will similarly reward hospitals that merely reduce length of stay.

“Upcoding” of diagnoses related to severity of illness and comorbid conditions represents another opportunity for gaming. Systematic biases in diagnostic coding were observed after the introduction of the prospective payment system and may have contributed to some of the early reductions in adjusted mortality attributed to the Cardiac Surgery Reporting System in New York State. Early concerns that surgeons in New York avoided operating on high-risk patients seem unfounded, but the incentive for physicians or hospitals to avoid caring for sicker patients remains a substantial concern for outcome-based performance measurement.

For measures of patient morbidity, accurate outcome assessment will have to overcome a reverse reporting bias—that is, a bias toward underreporting of diagnoses and events. Most hospitals rely on voluntary incident reporting of adverse drug events, a method known to be severely flawed. Thus, measurement of such events may penalize hospitals that use sophisticated detection methods and reward hospitals that do not. A similar reporting bias probably explains the observation of high surgical complication rates in hospitals with low mortality rates and other plausible indicators of superior quality.

The questions on outcome-based quality measures that have received the most attention concern the variations in outcomes that are attributable to factors other than quality, such as chance, selection effects, and patient factors. Most of the work in this area has focused on risk adjustment for hospital mortality rates. Early efforts were hampered by the use of models originally designed to predict financial rather than clinical outcomes, reliance on administrative data with known systematic errors and biases, and methodologic problems relating to model generation and performance analysis. Progress has been made by focusing on specific subgroups of patients instead of overall hospital mortality. Sophisticated risk adjustment can now be used to interpret mortality rates for patients undergoing cardiac surgery and interventional cardiology procedures; critically ill patients; and patients with acute MI, community-acquired pneumonia, and upper gastrointestinal hemorrhage.

In addition to adopting better models, sophisticated outcome measures tend to adjust for patient risk by using clinical rather than administrative data. The inaccuracy of the diagnostic codes contained in administrative data undermined early hospital mortality reports.

<table>
<thead>
<tr>
<th>TABLE 3</th>
<th>Outcome Measures of Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Examples</strong></td>
<td>Mortality after cardiac bypass surgery</td>
</tr>
<tr>
<td><strong>Advantages</strong></td>
<td>Generally meaningful from all points of view: patients, payers, providers</td>
</tr>
<tr>
<td></td>
<td>Useful when clinical outcomes of interest are determined by multiple aspects of structure or when relevant processes involve an important component of technical skill</td>
</tr>
<tr>
<td><strong>Disadvantages</strong></td>
<td>Sophisticated techniques are required to adjust for patient differences, random variation, and selection effects</td>
</tr>
<tr>
<td></td>
<td>Long periods of observation and large numbers of patients are required to detect significant differences</td>
</tr>
<tr>
<td></td>
<td>Sometimes difficult to attribute outcomes to specific episodes of care or components of the health care system</td>
</tr>
<tr>
<td></td>
<td>Often unclear how to “fix” quality problems identified by poor outcome results</td>
</tr>
</tbody>
</table>

For measures of patient morbidity, accurate outcome assessment will have to overcome a reverse reporting bias—that is, a bias toward underreporting of diagnoses and events. Most hospitals rely on voluntary incident reporting of adverse drug events, a method known to be severely flawed. Thus, measurement of such events may penalize hospitals that use sophisticated detection methods and reward hospitals that do not. A similar reporting bias probably explains the observation of high surgical complication rates in hospitals with low mortality rates and other plausible indicators of superior quality.

The questions on outcome-based quality measures that have received the most attention concern the variations in outcomes that are attributable to factors other than quality, such as chance, selection effects, and patient factors. Most of the work in this area has focused on risk adjustment for hospital mortality rates. Early efforts were hampered by the use of models originally designed to predict financial rather than clinical outcomes, reliance on administrative data with known systematic errors and biases, and methodologic problems relating to model generation and performance analysis. Progress has been made by focusing on specific subgroups of patients instead of overall hospital mortality. Sophisticated risk adjustment can now be used to interpret mortality rates for patients undergoing cardiac surgery and interventional cardiology procedures; critically ill patients; and patients with acute MI, community-acquired pneumonia, and upper gastrointestinal hemorrhage.

In addition to adopting better models, sophisticated outcome measures tend to adjust for patient risk by using clinical rather than administrative data. The inaccuracy of the diagnostic codes contained in administrative data undermined early hospital mortality reports.
a “sixth digit” to International Classification of Diseases, ninth revision, codes to distinguish secondary diagnoses present at admission from those that develop during hospitalization and may represent complications of care. This refinement may increase the validity of risk-adjusted mortality rates that use administrative data, but the expectation of such improvement remains speculative.

Even with ideal risk adjustment, the effects of chance and selection are considerable and probably cause more variation than do provider differences. Furthermore, even providers and hospital executives who have helped develop outcome measures tend to distrust performance reports that are based solely on outcomes, presumably for many of the previously cited reasons. Nonetheless, given the current limitations of other quality measures and some notable successes of outcome-based reports, outcome assessment will continue to play a role in quality measurement. Valid application will depend on recognition of situations in which the concerns described here can be adequately addressed.

Quality Measurement versus Quality Improvement

We have focused on quality measurement as it relates to report card–style comparisons between hospitals, rather than the development of indicators designed to fuel quality improvement activities. Providers may believe that quality measurement has value only if it informs quality improvement. On the other hand, patients have a legitimate interest in knowing how the providers who are available to them vary in quality. Moreover, although developers of report cards must ensure that overemphasis on measurement does not undermine quality (e.g., by randomly rotating subsets of standard measures to eliminate “playing for the test”), such publicly reported quality measurement has already produced at least one notable success. By contrast, the dissemination of clinical practice guidelines and other models of quality improvement has generally failed to match this accomplishment.

Conclusions

We have illustrated the use of the conceptual scheme of structure, process, and outcome to develop quality indicators related to hospital care. In discussing the advantages and disadvantages of these types of indicators, we have also touched on the issues involved in developing indicators that are appropriate for clinical and measurement contexts and the importance of targeting the correct quality problem: underuse, overuse, or misuse.

Beyond the technical issues discussed, the major barrier to effective quality measurement and improvement remains a health care culture in which most hospitals muster the resources to measure quality only in response to immediate financial or organizational pressures. The recent spotlight on medical error may provide an opportunity to secure funding for broader initiatives related to hospital quality measurement and improvement. A substantial increase in the dollars spent on quality-related activities would still consume only a tiny fraction of the resources presently expended on developing novel inpatient diagnostic strategies and therapeutic agents.

References

9. Pear R. Clinton to order steps to reduce medical mistakes; a mandate for hospitals; states, in turn, will be asked to add systems to require the reporting of errors. New York Times. 2000: A1(N), A1(L).


Frick KD, Lantz PM. How well do we understand the relationship between prenatal care and birth weight? Health Serv Res. 1999;34:1063-73; discussion 1074-82.


Iezzoni LI. Risk adjustment for measuring healthcare outcomes. 2nd ed. Chicago, Ill.: Health Administration Pr; 1997.


Acknowledgment
The authors thank Amy J. Markowitz and Drs. Harold Luft, Arnold Milstein, and Jeffrey Rideout for their suggestions and comments on earlier drafts of this manuscript.

Grant Support
In part by Josiah Macy Jr. Foundation (Dr. Shojania).

Correspondence
Robert M. Wachter, MD, Department of Medicine, Box 0120, University of California, San Francisco, San Francisco, CA 94143-0120; telephone: 415-476-5632; fax: 415-502-5869; e-mail: bobw@medicine.ucsf.edu.