

Methodology Matters

From a process of care to a measure: the development and testing of a quality indicator

HAYA R. RUBIN¹, PETER PRONOVOST² AND GREGORY B. DIETTE³

¹Departments of Medicine, Health Policy & Management, and Epidemiology, and Director, Quality of Care Research, The Johns Hopkins Medical Institutions, Baltimore, MD, ²Departments of Anesthesiology & Critical Care Medicine, Surgery, and Health Policy & Management; The Johns Hopkins Medical Institutions, Baltimore, MD and ³Departments of Medicine and Epidemiology, The Johns Hopkins Medical Institutions, Baltimore, MD, USA

Abstract

This paper outlines the steps in developing and implementing process measures of quality. Developing a process measure includes defining the purpose of and audiences for the measures, choosing the clinical area to evaluate, organizing the assessment team, choosing the component of the process to measure, writing the indicator specifications, performing preliminary tests of feasibility, reliability and validity, and determining scoring and analytical specifications. Given the growing evidence in the literature regarding the impact of care, and an evolving understanding of how to develop and implement process of care measures as outlined here, the future should bring the development and implementation of quality indicators that are rigorously developed and that will provide insights into opportunities to improve the quality of care.

Keywords: outcomes, performance improvement, quality

In a companion paper in this issue, we have discussed the need for health care quality measures for the various audiences who may wish to use them for health care purchasing, utilization, regulatory accreditation and monitoring, or performance improvement [1]. For all these purposes it is imperative that the measures are meaningful, scientifically sound, generalizable, and interpretable [2]. To achieve this, quality measures must be designed and implemented with scientific rigor. In this paper, we discuss the steps and issues in developing and testing process-based measures of health care quality.

An organization, group or team intending to develop process-based measures of health care quality should have a thorough understanding of their strengths and limitations *vis-à-vis* other types of quality measures such as outcome measures [3–9]. This will assist in understanding whether process-based measures are the best choice for the team's aims. If, after consideration, process measures are deemed most useful, the initial steps required to develop them will include: (1) define the audience and the purpose of measurement; (2) choose the clinical area to evaluate; (3) organize the measurement team; (4) select the process criterion; (5) write the measure

specifications, including the unit of analysis, the indicator, the target population, any needed risk adjustment strategy, the data sources, and the data collection specifications; (6) perform preliminary tests including pilot testing of the measures and data collection methods and testing the scientific strength (reliability and validity) of the indicator; and (7) develop scoring and analytical specifications. We discuss these seven steps in detail below.

Define the audience and purpose of measurement

Because quality is a multidimensional construct, multiple indicators, each providing insights into a different domain of quality, are required to fully evaluate quality of care [10]. Data about quality of care have different uses and different audiences. These diverse audiences may emphasize different domains of the process, and may be interested in different units of analysis. Therefore, before developing quality measures, it is important to clarify the goals and purpose of the

Address reprint requests to H. R. Rubin, Quality of Care Research, The Johns Hopkins University, 1830 East Monument Street #8015, Baltimore, MD 21205, USA. E-mail: hrubin@jhmi.edu

quality measurement effort including the intended audience or consumer of the information [11,12]. For example, will the quality measure be used for local improvement efforts, for benchmarking externally against other organizations, or by an accreditation agency? One should explicitly state what these uses are and who will use the measure. The uses and audiences for the measure are important to define, as they will dictate the focus on particular clinical areas and elements of care, discussed more extensively below. For example, patients using quality measures to select a physician may be more interested in indicators of physicians' communication and interpersonal interactions than would a provider organization selecting a physician. Purchasers interested in choosing providers may be more interested in high cost clinical areas than would clinicians undertaking quality improvement efforts. Regardless of the goal, the consumers of the information must believe that the metric is valid and represents an important domain of quality.

Different audiences and uses also determine which unit of analysis is of interest. For example, in the US, the unit of analysis may include the patient, the individual doctor or clinician, the office practice group, the larger provider organization such as the hospital or practice network, the health plan, the state, the region, or the nation. Various audiences, such as patients, employers, accreditation agencies, provider organizations and clinicians are interested in different units of analysis for different purposes. For example, comparison of a health care plan's performance with that of other plans may have no meaning to consumers who are choosing physicians within the same plan, nor to the health plan choosing or evaluating its participating physicians. On the other hand, such data at the health plan level will be critical to a health plan for marketing or accreditation. As the desired unit of analysis moves closer to individual doctors or clinicians, concern is raised regarding the adequacy of sample size. Individuals may have too few patients with the disease of interest to provide precise estimates of the quality measure.

Choose the clinical area to evaluate

In choosing clinical areas to evaluate, several criteria must be considered (Table 1). Firstly, the clinical problem under study must be important to the intended audience for the desired purpose. For example, clinicians, provider organizations, accreditation agencies, and purchasers of care are interested in clinical areas that have a significant impact on morbidity, mortality, and/or costs of care. In some cases, clinician or purchaser audiences may wish to focus on clinical areas that are highly important to patients. The clinical area or process domain can be of local importance, for example, a hospital may have a particular problem with care in a particular unit, or a user may focus on a problem of national significance such as congestive heart failure, which has a large impact on a national population [13,14]. Indicators are designed as a means to improve clinical, service, and economic performance. When selecting measures of quality, some users

will wish to consider not only the clinical rationale for collecting the data but also the economic rationale. This will necessitate knowledge of the costs of data collection and potential cost savings from improving the aspect of care being evaluated. Some users may be reluctant to measure aspects of care for which the cost of collecting the data exceeds the potential economic benefit from improvement of that process. Clinical audiences and public representatives may be interested in quality improvement efforts even if they increase health care costs, but it will be difficult to obtain funding from provider organization administrators, purchasers, or care management organizations for efforts that increase costs because these individuals and organizations are held accountable for reducing health care expenses.

Organize the assessment team

Once the purpose and use have been defined, and the clinical area and process domain selected, one can assemble an appropriate assessment team and appropriate advisors for the measurement or improvement effort. The team's advisors should include representatives of the desired audience for the measure, as well as the clinicians that will be evaluated and the administrators whose resources will be used. To ensure that valid content is included, clinical experts and representatives of national professional and specialty societies should be represented. In addition, quality of care researchers with training in epidemiology who know how to evaluate measurement reliability and validity should be included in the assessment team. Process measures of health care quality will change regularly as scientists and clinicians develop better treatment methods. Because intensive resources are required to maintain and update these measures, joint efforts by professional societies, provider organizations, and regulators or accreditors will lead to the most efficient use of resources. When possible, an organization wishing to develop process indicators would use its own resources most efficiently if it teams with others that also require these measures, or uses measures that have been developed by national organizations such as, in the USA, the National Committee on Quality Assurance [15], the Center for Medicare and Medicaid Services (CMS), formerly the Healthcare Financing Administration (HCFA) [16], or the Institute for Healthcare Improvement (IHI) [17].

Select the process domain or criterion

Evaluating the quality of a process of care requires determination of whether clinicians adhere to practices that are important to achieve the best outcomes for similar patients. The linkage of practice to outcomes must either have been demonstrated scientifically or must be widely accepted by peers. Process evaluation can be done by asking clinical experts in the field to judge the care based on their own knowledge of the standard of care without delineating specific

Table 1 Steps in developing and implementing a clinical performance measure

1. Define audience and use for measurement	Identify audience: clinicians, administrators, purchasers, regulators, patients Identify purpose: quality improvement, regulation, purchasing, selection of providers
2. Choose clinical area to evaluate	Impact on morbidity, mortality, costs Volume high enough to measure
3. Organize assessment team	Expertise in measurement Expertise in clinical area Expertise in project administration User representatives
4. Select aspect of care or process criteria to be measured	Strong scientific evidence that this aspect of process affects important outcomes Feasible to measure Previously tested measures of demonstrated reliability and validity exist or new measures are relatively easily developed Providers or managers can influence this aspect of process Variability in performance or substandard care exists
5. Write measure specifications	Specify unit of analysis Define indicator Identify intended sample and exclusions Specify risk-adjustment strategy Decide upon data source and data collection method Write data collection specifications
6. Perform preliminary tests	Pilot test feasibility of measures and data collection method Test reliability and validity of measure
7. Write scoring, analytical specifications	Specify how measures will be scored, how to define acceptable performance Determine analytical plan

criteria. This procedure is known as implicit peer review [18, 19]. Such subjective evaluation may be helpful in areas such as critical care where the patient's condition is too complex to fit into a category for which a guideline exists. However, due to poor inter-rater reliability with implicit review, five to seven evaluators are generally required to get an average rating that has high reliability. In addition, this method obscures feedback to providers about specific elements of care that can be improved and puts the evaluation in a 'black box'.

Explicit process review establishes criteria for specific practices that should be done for a carefully defined population of patients, and determines in what proportion of patients this practice has been applied. Explicit review is more useful to providers because it provides feedback about exactly what has or has not been done. In addition, once explicit criteria have been developed or chosen by clinical experts, an individual assessor without clinical training can reliably determine whether they have been met. In explicit quality review, inconsistency in opinions about which elements of care are important is adjudicated using a group process or consensus method to develop the criteria. Most of the following discussion focuses on choosing specific components and measures for explicit review, although some parts are applicable to implicit review measurement methods as well.

In order to determine whether to measure a specific component of the care process, it is helpful to review the strength of scientific evidence that inclusion of this component in the process measure will affect outcomes, to consider whether it can be influenced by clinicians or managers, whether it is feasible to measure this component, whether there are existing measurement methods for it with demonstrated reliability and validity, and whether there is demonstrated variability among clinicians in how or whether it is performed.

Evaluate the strength of scientific evidence for the process criterion

The scientific evidence for the relationship of the chosen component of the process (what providers do) and outcome measures (what happens to patients) should be established by empirical data. Ideally, process components measured would be those demonstrated to cause a higher probability of achieving the desired outcome [13,20]. Once the link is established, providers have confidence that improvement in a process will translate into improvement in an outcome. Because evidence establishing these types of links is rare, developers of process measures often use consensus among providers to establish the validity of process-based quality measures. The belief among providers that the process measures evaluate an important domain of quality of care is critical

to engaging providers in an evaluation and improvement process. Most non-clinical audiences for quality measures will be satisfied that a component of the process and the specific method used to measure it are important for patient outcomes if providers with clinical expertise in the area agree that the component and the method are important and if clinical researchers with knowledge of the area confirm that the evidence supports their use.

In considering the strength of the evidence for a particular explicit criterion, the assessment team should consider the type and number of studies that are related to the indicator. Randomized clinical trials provide the strongest evidence, while observational studies and reports of consensus groups provide significantly weaker evidence [21]. Multiple studies or meta-analyses of a topic provide stronger evidence than single studies. The compendium *Clinical Evidence* [22] and the Cochrane Collaboration website [23] are good resources for determining the strength of evidence available for various elements of the process of health care. The strength of the evidence for an indicator will determine its scientific soundness or the likelihood that improvement in the quality indicator will produce consistent and credible improvements in quality of care.

One widely used method for developing or choosing criteria for explicit review of the process of care is the RAND appropriateness method. This combines a systematic review of the literature with ratings by an expert panel based on the literature and, where there are gaps in the literature, on the panel's own experience. The panel is asked to define care for all possible subgroups of patients with a particular condition or 'indication' as appropriate, uncertain, or inappropriate [24, 25]. Criteria developed by expert panel ratings have been demonstrated to predict improved outcomes in patients undergoing angioplasty or coronary artery bypass surgery for coronary artery disease, and for patients treated in the hospital for myocardial infarction, pneumonia, hip fracture, and stroke [20,21,26]. Expert panel ratings may provide a reliable and valid method to evaluate quality for other types of patients and medical conditions where there is little evidence from randomized clinical trials from which to create guidelines.

Determine if the component of process is feasible to measure

An important criterion for choosing the component of the process is that it must be feasible to measure. For example, measuring the use of barrier precautions while inserting central venous catheters requires direct observation, and the resources for such observation may not be available. On the other hand, the frequency of prescription of beta-blockers for patients with myocardial infarction may be determined very easily if there is a computerized order entry system that allows flexible search inquiries.

Determine if there are existing measures that are reliable and valid

Measures may already have been developed for certain aspects of the clinical process, whereas others that may be equally

important may not yet have had much attention. For example, in the area of diabetes, the American Diabetes Association has developed guidelines for care. Some of the process components specified in these guidelines have had extensive measure development. For example, for evaluation of glycosylated hemoglobin and referral for retinal screening, the Center for Medicare and Medicaid Services [16] and the National Committee on Quality Assurance [15] developed precisely defined and specified indicators that have been widely applied and reported [27]. In the case of other aspects of the process of care that may be demonstrated to be equally or even more important to patient outcomes, such as whether or not dietary counseling has been provided, the measurement methods may not have had such thorough testing. If an organization lacks extensive resources for measure development and testing, it may be worthwhile to choose a process area for which measures and data collection methods that have been demonstrated to be reliable and valid already exist.

Decide if providers and managers can influence the process component

For a measure of the process of care to be helpful in quality improvement, providers or managers must be able to influence that process. The process component measured should be within the control of the group that is the unit of analysis. For example, although actual use of inhaled corticosteroids is what improves patient outcomes, a measure of whether inhaled corticosteroids have been prescribed to patients may be more satisfying to doctors than a measure of actual use, as doctors may not feel that they can markedly influence patient adherence.

Determine if there is variability among clinicians or substandard quality

The sixth criterion in selecting clinical areas to evaluate is whether there is reason to believe that variability or substandard quality exists in that aspect of the process. The variability and substandard quality, usually unintended, represent the potential opportunity for quality improvement [12, 13]. This criterion is particularly important for internal quality improvement efforts. Regulators and accreditors may use minimal standards upon which there is nearly uniform good performance, although the trend in these organizations is to provide a multi-level evaluation that highlights excellent performance as well as merely meeting minimal standards.

Write the measure specifications

After the quality measure or component of the process has been selected, the specifications for that measure must be stated. These specifications are the methods by which the target population is identified and the data are actually collected and should be developed with the same scientific rigor with which we would develop research samples and data

collection instruments. There are six steps in defining the measure specifications.

Select the unit of analysis

The quality assessment team must specify the unit of analysis for the measure of process. As discussed earlier, this often stems from identifying the desired purpose and audience for the measure. For example, for US inpatient care, one must specify whether the unit of analysis is the patient, the provider, the provider group, the hospital service or unit, the hospital, the health system, or the insurer. Explicitly stating the unit of analysis will help with interpretation of the measure. If multiple measures are used, it is helpful to choose a common unit of analysis. The choice of the unit of analysis may depend on other factors such as the purpose of the measure and the audience as has been described earlier.

Define the indicator

The quality assessment team must define the indicator to be measured. For dichotomous measures (where the answer is either 'yes' or 'no'), this is presented as a proportion of a population where a numerator and a denominator are defined, for example, the proportion of patients staying at least 24 hours in the intensive care unit who received stress ulcer prophylaxis [28]. Other measures may be continuous measures that can be averaged, such as time to an event. Examples of such indicators include time to antibiotic administration for patients presenting to an emergency department for community-acquired pneumonia, or minutes to thrombolytic therapy for patients presenting to hospitals with myocardial infarction or stroke. Alternatively, an indicator may be a rate, defined as a proportion within a given time frame, such as the proportion of patients for whom restraints were used in a 24-hour period. Still other measures may be scores on a scale such as a pain or patient satisfaction scale.

Identify the intended sample and exclusions

The intended sample is the group in whom the measure will be assessed who represent the population at risk. Developers of a measure must decide upon the inclusion criteria, exclusion criteria, and must choose prevalent (existing) versus incident (new) cases. Explicit selection criteria reduce potential measurement bias. For example, if a quality measure is the percent of intensive care unit patients who receive stress ulcer prophylaxis, the assessor must define the denominator for this rate. This denominator may include patients who are mechanically ventilated, older patients, patients with raised intracranial pressure, or patients who are in intensive care for more than 24 hours. For the use of inhaled corticosteroids in patients with asthma, patients included should clearly require the use of this medication. National guidelines state that patients who have a severity level of moderate persistent asthma, for example, should be using inhaled corticosteroids. Identifying such patients, however, could require data collection of symptoms or physiology from patients or from the medical records.

Define the risk-adjustment strategy

Risk adjustment plays a smaller role when the indicator is a process measure than when it is an outcome measure. However, for certain measures of process, risk adjustment may reveal that patient factors are influencing a measure and that the measure does not actually reflect the process accurately. For example, use of inhaled corticosteroids by asthma patients might be most easily measured by prescriptions filled, assessed from pharmacy data. However, as described earlier, most clinicians, rather than learning what their patients took compared with other clinicians' patients, would rather know what they prescribed as compared with other clinicians. This evaluation might require more expensive medical record review rather than the proxy indicator obtained from pharmacy data. Such proxy indicators of the process of care may be more informative if presented with risk adjustment. For example, prescriptions filled for inhaled steroids could be presented for different types of patients such as low-income patients or those with fewer years of education, or adjusted-use rates could be developed using a model with many patient characteristics. The more closely an indicator measures the actual process of care delivered rather than patient adherence or other factors, the less risk adjustment will be needed.

Identify data sources and data collection procedures

Once all the required data are known, the assessment team must state where and how the data will be obtained [29]. There are four likely sources of data: existing administrative databases, existing medical records, clinical data collected prospectively for quality assessment purposes, and survey data collected prospectively. Each data source has its limitations and benefits (Table 2). Regardless of the audience and use for the data, it is most desirable to incorporate data collection into routine patient care, that is, to standardize clinicians' and administrators' documentation of needed information about patient characteristics and care delivery that is already being recorded for clinical purposes. Such a strategy will reduce missing data, will reduce the additional cost of data collection, and will educate clinicians about important elements of the process of care. If it eliminates other clinician data collection rather than being duplicative, this data collection strategy will increase clinician commitment to quality assessment and improvement initiatives. As electronic medical records become more widespread, standard data specifications useful for quality assessment can be incorporated into these systems. Making sure that routinely collected administrative and clinical data meet such specifications has the potential to provide data for quality assessment while streamlining clinical data collection rather than creating additional duplicative data collection efforts. The ability to achieve efficient data collection integrated with the clinical and administrative health care routine poses the single most important consideration affecting the data collection strategy.

Table 2 Advantages and disadvantages of types of data for measuring quality of care

Type of data	Advantage	Disadvantage
Secondary data (administrative)	Readily available Inexpensive	Lacks specificity and detail
Medical record data	Available Richer in detail than administrative data If standardized in an electronic medical record, reduces data collection burden	Expensive to obtain May have insufficient detail
Prospectively collected clinical data	Most specific; can define exactly what data are required Quality control of data collection	Not readily available Expensive to obtain unless already incorporated into electronic medical record
Survey data	Can collect what is important to patients Collects data not otherwise available	Not readily available Expensive to collect Valid instrument required

Produce data collection specifications

Detailed and explicit data specifications ensure quality control of data collection and reduce the risk of error and bias. The team should define explicitly both the population at risk and the method for evaluating patients’ exposure to the specific element of the process that represents the quality indicator. For example, for prophylaxis for deep venous thrombosis, we would need to define high quality as any treatment provided, or treatment above a specified threshold. To define a high-quality process of care to which the patient might or might not be exposed, issues arise similar to those encountered when evaluating exposure to a risk factor in a cohort study [30]. Without the same methodological rigor as would be applied in that situation, it will be difficult to determine whether the differences between groups as measured by a quality indicator are true differences in quality or merely differences in data collection methods. After the specifications are written, pilot tests of the measure ensure that the data collection is feasible and accurate.

If the assessment team uses a sampling strategy, such as random or systematic sampling, the sample size and power calculations require justification [25,26]. In addition to a sampling strategy, the specifications should describe the time interval in which the data are to be collected, how patients who are ‘censored’ (for example, those who die, or are transferred to other services or locations) will be handled, and, when applicable, which of the multiple measures obtained on a single patient will be included. Finally, the specifications need to include some method for data quality control.

Preliminary testing

Pilot testing the measure and the data collection method

Although the design specifications are explicit, pilot testing the quality measure generally identifies areas requiring further

modification. For example, the measure specifications may need to be more detailed, or the data to obtain the measure may not be available. Pilot testing can generally be performed on a small sample. Nonetheless, skipping this step is perilous. As an example, a common process measure for congestive heart failure is whether left ventricular function has been assessed. In the USA, this assessment is often performed by an outpatient physician and transmitted to an attending physician in the hospital, and therefore is often missing from the inpatient medical record. Pilot testing can identify that a review of hospital records alone cannot be used as the data source for this process indicator.

Test the scientific strength of the measure

The next step in developing a quality measure is testing the scientific strength of the measure, i.e. evaluating its reliability and validity. Reliability can be thought of as reproducibility: if I repeat the measure, will I obtain the same result? Reliability is important when comparing the quality indicator among groups and within groups over time. The commonly used measures of reliability include inter-rater reliability (comparing differences among evaluators), internal consistency (comparing variation among items that should provide similar results), and test-retest reliability (comparing differences when the same person repeats the measure at two different time points).

In process measurement, one can test inter-rater reliability, test-retest reliability and internal consistency in order to determine whether selection and training of data abstractors, interviewers, or programmers, the indicator definition, and the data collection methods are precise enough to provide reproducible results. For example, if a measure is well specified, and data collectors receive good training, different data collectors should come up with the same results and there should be good inter-rater reliability. In addition, if the measure and data collection procedures are well specified, the same data abstractor should come up with the same

results on two different occasions when presented with the same data, defined as good test-retest reliability. Finally, internal consistency testing can also provide a check on the reliability of the overall method when there are two or more related items. For example, one would hypothesize that if appropriately defined, certain sub-indicators would be related to each other. If data collection methods are reliable, then, for example, the proportion of patients who have orders for hemoglobin A1C testing should be equal to or greater than the proportion of patients for whom a result of glycosylated hemoglobin testing is present in the medical record or the information system, and there should be a good correlation between scores on these two items.

Validity is the extent to which the measure accurately evaluates the domain of quality being assessed. The reliability of a measure is necessary but not sufficient for validity; that is, a measure can be consistent but there may be other reasons why it is not valid. Validity can be tested by: (1) confirming that scores on the measure are linked to measures of important outcomes; (2) demonstrating that the scores that are obtained from using the measure are able to differentiate between good and bad quality assessed using another previously validated method known to be related to important outcomes; or (3) evaluating whether the measure represents the process domain of interest as judged by the audience of users [14, 19].

Write specifications for scoring and analysis

Define acceptable performance and scoring

To determine what is acceptable performance, the assessment team must develop a protocol for scoring the measure [19]. With some dichotomous variables, the raw data are sufficient. For example, for process errors that always represent poor quality care, such as administering the wrong dose of a medication, the proportion of patients experiencing the event is a sufficient specification. Ordinal, or continuous data such as ‘time-to-events’ or rates may be converted into dichotomous variables by specifying the proportion of patients above a particular threshold. To aid in their interpretation, continuous variables can also be presented as means for a population in a comparison with a standard or benchmarks. For survey data, specifications and rationale for converting item responses into scale scores must be included. The specifications must also include a plan for handling missing data or data that are outside of a logical range. All of the stakeholders involved in developing the quality measure should agree on what constitutes acceptable performance.

Develop the analytical plan

The specifications must include a detailed plan for how the measure is to be analyzed and how statistical and clinical significance will be determined. Part of determining the analytical plan must include a description of the population,

an evaluation of the distribution of the data, how missing data are handled, a description of the unit of analysis, a description of the comparison group, and the statistical analyses and tests to be used. It is important for the investigators to consider what would be clinically significant differences in quality among groups rather than simply statistically significant.

One of the considerations in choosing the statistical test is the interpretation of the test results. Even if the measure is scientifically strong, it must be possible for the audience for whom it is intended to interpret it. As a result, the way the measure is presented may vary by the audience. Pilot testing is often required to ensure that the desired audience is interpreting the measure appropriately. Data must be presented in a meaningful way and efforts should be taken to focus on efficient communication of information. If the audience for the measures consists of purchasers or patients who seek the information for health care purchasing decisions, analysis plans that include summary variables or aggregations of individual process measures may be more helpful than the individual process measures. Comparing levels of statistical significance between groups becomes increasingly important; patients and enrollees want to know ‘which is better.’ Further research is needed to design and present quality measures that will enable patients to make informed decisions regarding quality of care.

Summary

In this paper, we have outlined how to develop or choose explicit process indicators. Process measures are highly acceptable to providers because they demonstrate clearly how providers can improve their outcomes. Clinicians are also more accountable for the process of care than outcomes, which are affected by many other factors. As electronic medical records become more common, process measures can be unobtrusively tracked as part of routine clinical care, which will aid in their implementation. Process measures that are incorporated into routine clinical data collection provide a constant educational reminder to clinicians about the correct process, and eliminate duplicative data collection for quality assessment. On the other hand, implementing process measures can be difficult because they require constant updating as the science of medicine advances. Joint efforts among providers, professional societies, and external government or payer agencies to develop and maintain process measures have made them more feasible. To be valid, process measures should have links to important outcomes, or should at least be determined by consensus methods and be judged by clinical experts to be important to patient outcomes. The past decade has brought a greater emphasis on synthesizing the evidence basis for how process of care affects outcomes and has made this information more easily available to the provider community as well as to the public. In the future this will provide the ultimate base for the development of process measures of quality.

References

1. Rubin HR, Pronovost P, Diette GB. The advantages and disadvantages of process-based measures of healthcare quality. *Int J Qual Health Care* 2001; **13**: 469–474.
2. McGlynn EA. Choosing and evaluating clinical performance measures. *Jt Comm J Qual Improv* 1998; **24**: 470–479.
3. Palmer RH, Miller MR. Methodologic challenges in developing and implementing measures of quality of child health care. *Ambul Pediatr* 2001; **1**: 39–52.
4. Palmer RH. Using health outcomes data to compare plans, networks, and providers. *Int J Qual Health Care* 1998; **10**: 477–483.
5. Palmer RH. Process based measures of quality: the need for detailed clinical data in large health care databases. *Ann Intern Med* 1997; **127**: 733–738.
6. Brook RH, McGlynn EA, Cleary PD. Measuring quality of care. *N Engl J Med* 1996; **335**: 966–970.
7. Jollis JG, Romano PS. Pennsylvania's focus on heart attack—grading the scorecard. *N Engl J Med* 1998; **338**: 983–987.
8. Green J, Wintfeld N. Report cards on cardiac surgeons; assessing New York State's approach. *N Engl J Med* 1995; **332**: 1229–1232.
9. Rainwater JA, Romano PS, Antonius DM. The California Hospital Outcomes Project: How useful is California's report card for quality improvement? *Jt Comm J Qual Improv* 1998; **24**: 31–39.
10. Blumenthal D. Quality of care—what is it? *N Engl J Med* 1996; **335**: 891–894.
11. McGlynn EA. Choosing and evaluating clinical performance measures. *Jt Comm J Qual Improv* 1998; **24**: 470–479.
12. McGlynn EA, Asch SM. Developing a clinical performance measure. *Am J Prev Med* 1998; **14**: 14–21.
13. Angus DC, Kelly MA, Schmitz RJ, White A, Popovich J. Current and projected workforce requirements for care of the critically ill and patients with pulmonary disease: Can we meet the requirements of an aging population? *J Am Med Assoc* 2000; **284**: 2762–2770.
14. Halpern NA, Bettes L, Greenstein R. Federal and nationwide intensive care units and healthcare costs, 1986–1992. *Crit Care Med* 1994; **22**: 2001–2007.
15. National Committee on Quality Assurance, USA: <http://www.ncqa.org>. Note: website address at time of printing.
16. Center for Medicare and Medicaid Services (CMS) [formerly Healthcare Financing Administration (HCFA)]: <http://www.cms.gov>. Note: website address at time of printing.
17. Institute for Healthcare Improvement (IHI): <http://www.ihl.org>. Note: website address at time of printing.
18. Rubenstein LV, Kahn KL, Reinisch EJ et al. Changes in quality of care for five diseases measured by implicit review, 1981 to 1986. *J Am Med Assoc* 1990; **264**: 1974–1979.
19. Rubin HR, Rogers WH, Kahn KL, Rubenstein LV, Brook RH. Watching the doctor-watchers. How well do peer review organization methods detect hospital care quality problems? *J Am Med Assoc* 1992; **267**: 2349–2354.
20. Brook RH, McGlynn EA, Cleary PD. Measuring quality of care. *N Engl J Med* 1996; **335**: 960–970.
21. Centre for Evidence-Based Medicine, UK: <http://cebm.jr2.ox.ac.uk>. Note: website address at time of printing.
22. Clinical Evidence, BMJ Publishing Group, UK: <http://www.clinicalevidence.org>. Note: website address at time of printing.
23. The Cochrane Collaboration: <http://www.cochrane.org>. Note: website address at time of printing.
24. Winslow CM, Kosecoff JB, Chassin M, Kanouse DE, Brook RH. The appropriateness of performing coronary artery bypass surgery. *J Am Med Assoc* 1988; **260**: 505–509.
25. Hilborne LH, Leape LL, Bernstein SJ et al. The appropriateness of use of percutaneous transluminal angioplasty in New York State. *J Am Med Assoc* 1993; **269**: 761–765.
26. Kahn KL, Rogers WH, Rubenstein LV et al. Measuring quality of care with explicit process criteria before and after implementation of the DRG-based prospective payment system. *J Am Med Assoc* 1990; **264**: 1969–1973.
27. Jencks SF, Cuerdon T, Burwen DR et al. Quality of medical care delivered to Medicare beneficiaries: A profile at state and national levels. *J Am Med Assoc* 2000; **84**: 1670–1676.
28. Cook D, Guyatt G, Marshall J et al. for the Canadian Clinical Care Trials Group. A comparison of sucralfate and ranitidine for the prevention of upper gastrointestinal bleeding in patients requiring mechanical ventilation. *N Engl J Med* 1998; **338**: 791–797.
29. McGlynn EA, Asch SM. Developing a clinical performance measure. *Am J Prev Med* 1998; **14**: 14–21.
30. Hulley SB, Cummings SR. Designing a New Study: I. Cohort Studies. In Hulley SB, Cummings SR (eds), *Designing Clinical Research*. Baltimore, MD: Williams and Wilkins, 1988.

Accepted for publication 28 September 2001