What Makes a Good Quality Measure?
Elizabeth A. McGlynn, PhD; John L. Adams, PhD

In this issue of JAMA, reports on the relationship between clinical outcomes and quality measures by Neuman and colleagues1 and Howell and colleagues2 facilitate reflection on the attributes of good quality measures. These studies raise issues that are familiar to experts in quality measurement3 but may not be understood by physicians and others whose performance is being assessed. Although the reports present information on a variety of types of measures, this Editorial focuses on the clinical quality measures.

Neuman et al1 examined the relationship between several publicly reported measures of performance in skilled nursing facilities (SNFs) and hospital readmission or death among Medicare beneficiaries receiving postacute care in these facilities. Based on their analysis of more than 1.5 million Medicare beneficiaries discharged from 3537 hospitals to 14,251 SNFs from September 2009 through August 2010, the authors found mixed relationships between the 3 intermediate clinical outcome measures and readmission or death. Patients in SNFs with lower rates of new or worsening pressure ulcers were less likely to be readmitted or die (23.2% vs 23.4%, \( P = .001 \)). There was no relationship between the prevalence of delirium and readmission or death (since the time of the study, the delirium measure has been dropped from public reporting). Lower prevalence of moderate to severe pain was associated with slightly higher rates of readmission or death (23.5% vs 23.3%, \( P = .01 \)).

Howell et al2 examined the relationship between 2 measures of hospital perinatal process quality (elective delivery between 37 and 39 weeks of gestation; cesarean deliveries among low-risk women) and severe maternal and neonatal morbidity in hospitals. The study included more than 103,000 newborns delivered at 41 hospitals in New York state in 2010. The authors found no correlation between the 2 process measures and the composite outcome measures.

Quality measures are typically evaluated on 4 criteria: importance, scientific soundness, usability, and feasibility.4 These 2 studies focus attention on scientific soundness, including both clinical soundness (the process-outcome relationship) and measurement soundness (reliability and validity). Readers skeptical of quality measurement and reporting may misinterpret these articles as evidence that the evaluated measures are inadequate for assessing quality. Given the frequency with which this question arises, reviewing the steps required to answer the question is warranted.

The first step is to determine whether a particular process should be related to a particular outcome. The relationship can be established conceptually (similar to the idea of biologic plausibility), clinically (typically through expert consensus), or empirically (through clinical trials or other studies). The complex bundle of both care processes and potential outcomes, as illustrated in these studies, adds to the challenge of discerning a relationship. In the study by Howell et al, the question was whether differences in performance on one element (early elective delivery or nonmedically indicated cesarean delivery) in a bundle of prenatal and perinatal services provided over a period of time were correlated with a composite childbirth outcome. The authors point out that changes in patient, clinician, and system factors are likely all required to decrease adverse maternal and neonatal outcomes; this includes care quality prior to delivery. Similarly, in the SNF setting evaluated in the study by Neuman et al,1 a variety of factors contributed to decisions about readmission or the outcome of death beyond clinical factors such as uncontrolled pain, worsening pressure ulcers, or delirium.5 Studies examining process-outcome (or outcome-outcome) links should clearly establish the empirical or logical basis for the relationships they are evaluating. In many regards this is the key step in the development and testing of valid quality measures.

The second step, if evidence exists to establish an expected process-outcome relationship, is to determine the magnitude of that relationship (similar to an effect size). The study by Howell et al reported a range in observed adjusted severe maternal morbidity from 0.9% to 5.7%, suggesting the upper bound opportunity is an absolute reduction of 4.8%. The range for adjusted severe neonatal morbidity was much larger (3.1%-21.3%). The study by Neuman et al reported an absolute difference in readmission rates of 0.7% (and did not report an overall mortality rate). Because those composite outcomes result from a large number of factors, the key question for this step is how much of a difference can reasonably be attributed to the process (or intermediate outcome) examined. Although the results of clinical trials may offer the best estimates of expected marginal contributions of a process to an outcome, such evidence may not be available. Benchmarking (identifying best-in-class performance) is a reasonable alternative and is one of the benefits of publicly reported performance. Little information is provided in either of these articles about this question.

The third step is to consider the difference in outcomes that might be reasonable to expect with optimal performance on the measure being evaluated.6 The proportion of the population to whom a process is relevant is frequently overlooked in studies of process-outcome relationships. For example, in the study by Neuman et al the interquartile range for the pressure ulcer measure was 8.2% to 16.7%;
moving every SNF from the 75th percentile of performance (poor) to the 25th percentile (better) could be associated with a reduction in the prevalence of pressure ulcers by 8.5 percentage points. However, only those patients at risk of developing pressure ulcers can contribute to changes in the overall outcome by this mechanism. Neuman et al reported a 2.4 percentage point difference in readmission or death rates (24.3% vs 21.9%) between SNFs in the top and bottom quartiles of pressure ulcer rates. Assuming a rehospitalization or death rate of 30% for these cases, the difference in pressure ulcer rates is consistent with the reported difference in readmissions and death. Because the empirical results match the expectations, the conclusion is that performance on this measure is consistent with the results. This underscores how a potentially useful process or intermediate outcome measure may be overlooked if it applies only to a subset of the population. A similar analysis should be undertaken for the findings reported by Howell et al. For example, about 20% of women in New York state in 2010 had an early elective delivery, and it would be important to know the difference in their outcomes compared with women who did not have an early delivery.

In addition, this analysis could be used to evaluate the conclusions about these measures. Neuman et al concluded that improving performance on these 3 intermediate outcome measures would not result in large improvements in the probability of death or readmission within 30 days. While true, this does not diminish the importance of these intermediate outcome measures; if patients had a choice they would prefer to be admitted to SNFs that were more effective at preventing pressure ulcers and delirium and managing pain. The measures have face validity. The challenge with many outcome measures is that evidence about how to improve those outcomes is lacking and by the time serious outcomes occur (eg, death), the opportunity to improve care for those patients has been lost. The implication of the study by Neuman et al is that clinicians need to accurately identify more determinants of early death and readmission so those additional processes or intermediate outcomes could be improved. Similarly, Howell et al suggest that a larger number of quality measures are necessary to inform consumer choice among hospitals based on perinatal quality. While reasonable, this does not negate the utility of the existing measures. Quality measurement systems always face the challenge of parsimony—ie, trying to minimize the number of measures required to adequately represent performance in a particular area.

So what do these studies suggest about the future? Neither study looked at other processes as potential predictors of the outcomes of interest (readmission, death, childbirth outcomes) and therefore missed an opportunity to identify areas for future measure development. Further, careful consideration of the ultimate goal of a quality measure must be made. The information required for consumers to choose among nursing homes or hospitals may be different than the information required to improve clinical outcomes. Measures that work for one purpose and not another are still valuable. Future studies of quality measures should establish a clear framework and expectations for the intended goals of quality measures. Both reports make it clear that a great deal of additional work is needed to achieve the quality measures necessary for a more complete characterization of system performance and potential improvement opportunities.

ARTICLE INFORMATION

Author Affiliations: Kaiser Permanente Center for Effectiveness and Safety Research, Pasadena, California.

Corresponding Author: Elizabeth A. McGlynn, PhD, Kaiser Permanente Center for Effectiveness and Safety Research, 100 S Los Robles Ave, Pasadena, CA 91101 (elizabeth.a.mcglynn@kp.org).

Conflict of Interest Disclosures: The authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest and none were reported.

REFERENCES


