

All-or-None Measurement Raises the Bar on Performance

Thomas Nolan, PhD

Donald M. Berwick, MD, MPP

THE PURSUIT OF EVIDENCE-BASED MEDICINE IS NOW AT the core of the agenda for improving health care in the United States. All major quality measurement systems use science-based indicators of proper processes of care, such as the ORYX measures of the Joint Commission on Accreditation of Healthcare Organizations,¹ the Health Employer Data and Information Sets measures of the National Committee on Quality Assurance,² the measures used by the Quality Improvement Organizations under contract with the Centers for Medicare & Medicaid Services,³ and at least 70 of the 179 measures in the 2004 National Health Care Quality Report from the Agency for Healthcare Research and Quality.⁴

Often, several individual performance measures are used to assess care of the same condition. For example, a recent summary of data on the Joint Commission on Accreditation of Healthcare Organizations' standardized performance measures included 9 measures for acute myocardial infarction, 4 measures for congestive heart failure, and 5 measures for pneumonia. The Joint Commission on Accreditation of Healthcare Organizations' standard set for pneumonia measures performance with respect to oxygenation assessment, pneumococcal vaccination, blood cultures, smoking cessation counseling, and mean time to initiation of antibiotics. The first 4 of these are discrete measures that indicate the presence or absence of the item in the medical record. The time to initiation of antibiotics is a continuous variable measured in minutes. There are at least 3 different options for calculating performance on multiple, discrete measures for the same condition.

Option 1: Item-by-Item Measurement

Performance on the provision of each element of care is reported separately as a percentage, in which the denominator is the total number of patients in the sample and the numerator is the total number of patients for whom the item was present in the medical record. The 2004 National Health Care Quality Report,⁴ for example, uses item-by-item measurement when it states, "The national rate for HbA_{1c} testing at least once annually for adults with diabetes age 18 and over was nearly 90% in both 2000 and 2001"^{4(p24)} and "The percentage of pneumonia discharges for patients 65 and older who had blood cultures before antibiotics was 81%. . . ."^{4(p59)}

Option 2: Composite Measurement

Performance on the provision of several elements of care is reported by computing a percentage across all patients and criterion indicators. For example, for the 4 elements of pneumonia care (excluding the continuous variable of time to treatment), a composite measure of performance can be computed by summing the numerators for each measure across the population of interest to create a composite numerator (all the care that was given), summing the denominators for each measure to form a composite denominator (all the care that should have been given), and reporting the ratio (the percentage of all the needed care that was given). This approach to measurement gives partial credit for incomplete care of an individual patient. If a patient receives 3 of the 4 recommended care elements, a hospital whose performance is being assessed with such a composite measure gets credit for delivering 3 elements. The Centers for Medicare & Medicaid Services uses composite measurement of this type in its Hospital Quality Incentive Demonstration Project.⁵

Option 3: All-or-None Measurement

A percentage is determined by applying an all-or-none rule at the patient level. For example, in pneumonia care the denominator could be the number of patients eligible to receive at least 1 of the 4 discrete elements of care and the numerator could be the number of patients who actually received all of the care for which the specific patient was eligible. No partial credit is given. The Centers for Medicare & Medicaid Services has moved to such an all-or-none approach in defining the appropriate care measure in its 8th Scope of Work.⁶

In measurement terms, the all-or-none approach of process quality yields a picture quite different from either the item-by-item approach or the composite approach. According to the 2004 National Healthcare Quality Report, for example, item-by-item measurement shows the following performance for standards in diabetes care: HbA_{1c} testing: 90.0%; lipid profiling: 93.8%; retinal examinations: 69.7%; foot examinations: 66.3%; and influenza vaccination: 56.5%.^{4(p24)} However, all 5 of these interventions reached only 32.1% of patients. HealthPartners, an integrated care system based in Minneapolis, Minn, and the first major health care organization we encountered that reported on all-or-none mea-

Author Affiliations: Institute for Healthcare Improvement, Cambridge, Mass.
Corresponding Author: Donald M. Berwick, MD, MPP, Institute for Healthcare Improvement, 20 University Ave, Seventh Floor, Cambridge, MA 02138 (dberwick@ihi.org).

tures, found typical performance ranging from 37.4% to 92.4% for each of the 7 process measures of diabetes care but only 7.6% of patients received all 7 interventions.⁷

Advantages of All-or-None Measurement. The movement to all-or-none performance assessment is an important milestone on the journey to high-quality health care. The all-or-none approach to measuring performance offers several important advantages over either item-by-item measurement or composite measurement.

First, all-or-none measurement more closely reflects the interests and likely desires of patients. This is especially true when process components interact with each other synergistically or when partial execution of a series of steps is insufficient to achieve a result. For example, violation of a single step in the sterile technique in surgery may vitiate the benefits of proper execution of all the other steps; credit for proper hand washing makes little sense if the surgeon breaks the sterile technique while putting the gown on. Choosing the correct prophylactic antibiotic in surgery may not help the patient if the medication is not given at the proper time. At a deeper psychological level, patients may expect and deserve reliability in their care even when components do not interact clinically. Patients might fairly judge a hospital that correctly performs 10 steps in myocardial infarction care reliably every single time to be much more than 10% better than one that is 90% reliable on average. Quality often is an all-or-none property.

Second, all-or-none measurement fosters a system perspective. It encourages concern with the design of the whole sequence of care and handoffs, not only parts. For example, imagine a measurement set addressing the goal of reducing hospital readmissions that depends on 5 component processes, 4 of which (for example, having to do with prescribing the right medications) are relatively easy to achieve and have reliabilities of 98%, but one of which (for example, ensuring complete communication between inpatient and outpatient care settings) is more difficult, with a reliability of 60%. For every 100 patients, there are 500 opportunities to deliver care in compliance with the measures. On average, 452 steps (98 + 98 + 98 + 98 + 60) would be carried out correctly for the 100 patients. The composite score is 442/500 or 90.4%. However, all-or-none measurement reveals that fewer than 60% of patients received all of the intended care. Item-by-item or composite approaches more easily allow caregivers and organizations to defer difficult system changes, such as improving cooperation across boundaries, while still achieving relatively high marks. All-or-none measurement emphasizes the importance of tackling the toughest problems in care.

Third, all-or-none measurement offers a more sensitive scale for assessing improvements. According to item-level process measures, many hospitals are achieving more than 90% compliance with individual elements of care.⁸ This leaves little room on the scale for setting goals and measuring progress. High scores can dampen enthusiasm for further improvement even if overall reliability for complete pro-

cesses remains low. If 5 items in a measurement set are statistically independent and each shows 90% reliability, only 59% of patients will receive all the care they need. The all-or-none measurement, unmasking a result in the middle of the scale, provides ample room for improvement and may stimulate more motivation to make changes.

Issues and Opportunities With All-or-None Measurement. Widespread public measuring and reporting of performance is in the early stages of development. There is still much to learn technically and politically about measurement use. Scholars and practitioners will face challenges in improving on early versions of all-or-none measurement. In the further development and refinement of all-or-none measurement, the following issues, among others, deserve consideration.

Selection of Measures Included in an All-or-None Measurement Set. The number of measures in the set should be small (perhaps 4 to 8) and each one should measure performance with respect to the specified elements of good care. Each of the care elements should be supported by evidence that links it to 1 or more desirable end points. The best candidate sets for all-or-none scoring are probably the ones thought of as the indisputable basics of care for a given condition. The all-or-none approach will amplify errors of measurement (one unreliable component measure will contaminate the whole score) so that it is essential that each of the component measures be well designed. One useful set of criteria for such measures is that sought by the Joint Commission on Accreditation of Healthcare Organizations for its core measures: well defined and specified, reliable, and valid.⁹ Not surprisingly, organizations using the all-or-none approach have found it important to improve processes of documentation for the specified elements of care to ensure that the medical record reflects the processes of care as accurately as possible.

A clear distinction should be made between process and outcome measures. When outcomes can be measured directly, they should be. All-or-none measurement seems best suited to process measures. Once an organization reaches a high score on the all-or-none measure, much of the variation in the delivery of care will have been removed, and the relationships between outcomes and other causal factors can be studied with much greater precision.

No matter how an all-or-none measurement set is formed, its authors should make explicit the rationale used to develop the set, including the hypothesized relationships between the specified elements of care and desired outcomes and, respectively, between process and outcome measures.

Addressing Multiple Outcomes With All-or-None Measurement. With respect to any condition, multiple outcomes will commonly be of interest. These outcomes often will be affected by processes of care at different locations, such as inpatient and outpatient sites. For example, in the care of acute myocardial infarction, in-hospital mortality and 1-year mortality both matter as do both in-hospital care and subsequent outpatient care. The selection of items in an all-or-none measurement set is actually a policy choice, offering

the measurer an opportunity to focus attention on a particular outcome across a continuum of care or to focus attention on a particular location or organization with respect to 1 or more outcomes.

For example, with respect to mortality from acute myocardial infarction, a hospital all-or-none measurement might contain, among other items, the use of β -blockers and aspirin in appropriate patients at both admission and discharge. Their inclusion in the set is justified by evidence relating them to morbidity and mortality.^{10,11} This approach would emphasize and illuminate the reliability and responsibility of a specific organization (the hospital) for its contributions to both in-hospital and longer-term mortality.

Another option would be to use an all-or-none measurement that includes processes of acute myocardial infarction care for both the index hospitalization and subsequent outpatient management that are related to 1-year survival. This measure would emphasize and illuminate processes of cooperation and coordination of care across settings.

All-or-None Scoring vs Weighted Score. It is true that all the components of care are not equal with respect to their influence on patient outcomes or experience. However, a useful all-or-none measurement set will contain only important measures, close enough in their impact that the equal weighting inherent in all-or-none measurement is a reasonable starting place. As caregivers and researchers learn more about the relative contribution that the elements of care make to outcomes, weighting schemes can be developed and tested as alternatives to all-or-none measuring. All-or-none measurement is an improvement over item-by-item measurement and the simplicity of unweighted measurement may encourage more rapid adoption, even while research continues on weighting schemes.

All-or-None Measurement vs Item-by-Item Measurement. The unit of analysis for the all-or-none measurement is the individual patient. The patient either receives all the care items in the measurement set or does not. Improving that score will require changes to the system of care. A conscientious would-be improver trying to make an all-or-none score better likely will drill down into the individual care components to provide insight into what those changes might be. Examining the individual components within this larger framework has the added advantage of discovering whether a component may not add significantly to improved outcomes when it is included in the combined bundle. Eliminating that component (or those components) would therefore reduce waste of time, effort, and cost, which is also one of the general goals of improvement. The patient-centered view of all-or-none measurement does not preclude examining individual components; indeed, it would encourage it.

An important research goal, linked to the all-or-none approach, is to distinguish between those conditions in which achieving a desired clinical outcome empirically requires reliable completion of a full set of tasks (that is, when partial completion does not gain partial benefit) and those for which

there is a more continuous relationship between completeness and effect. In the former circumstance, the case for all-or-none measurement is especially strong. For example, recent work on reducing surgical site infections suggests that reliable execution of a total bundle of specific clinical tasks leads to a far lower infection rate than when even a single element of that bundle is omitted.¹² The National Surgical Infection Prevention Collaborative, sponsored by the Centers for Medicare & Medicaid Services, used the Institute for Healthcare Improvement's Breakthrough Series Collaborative approach to reduce surgical site infection rates in this manner. In this project, the 44 hospitals reporting data reduced their infection rates by an average of 27%, from 2.3% to 1.7%, from the first 3 months to the last.¹²

Conclusion

All-or-none assessment of process quality represents an important advance in the level of ambition, patient-centeredness, and system-mindedness of performance measurement and reporting as assets in the pursuit of better health care. It raises the bar and illuminates excellence in a societal enterprise that should not be satisfied with partial credit or incomplete execution and should have widespread adoption.

Financial Disclosures: None reported.

Acknowledgment: We thank Jane Roessner, PhD, Valerie Weber, BS, Erika Pabo, BA, Frank Davidoff, MD, and Don Goldmann, MD, for their assistance in preparation and revision of this article and Carol Haraden, PhD, Pat Rutherford, RN, MS, and Roger Resar, MD, for their many contributions to the development of the all-or-none measurement.

REFERENCES

1. Joint Commission on Accreditation of Healthcare Organizations. Facts about ORYX for hospitals: core measures and hospital core measures. Available at: <http://www.jcaho.org/accredited+organizations/hospitals/oryx/oryx+facts.htm>. Accessed October 3, 2005.
2. National Committee on Quality Assurance. The Health Employer Data and Information Sets (HEDIS). Available at: <http://www.ncqa.org/Programs/HEDIS/>. Accessed October 3, 2005.
3. Centers for Medicare & Medicaid Services. Medicare Health Outcomes Survey: final report on the health status of the disadvantaged. Available at: http://www.cms.hhs.gov/surveys/hos/download/HOS_Disadvantaged_ReportFinal.pdf. Accessed October 3, 2005.
4. Agency for Healthcare Research and Quality. *2004 National Healthcare Quality Report*. Washington, DC: US Dept of Health and Human Services; 2004. AHRQ Publication 05-0013.
5. Premier Inc. Summary of the composite quality scoring methodology. Available at: <http://www.premierinc.com/all/quality/hqi/resources/top-performer-summary.pdf>. Accessed January 15, 2006.
6. Centers for Medicare & Medicaid Services. 8th Scope of Work (version 080105-1). Available at: <http://www.cms.hhs.gov/QualityImprovementOrgs/Downloads/8thSOW.pdf>. Accessed January 20, 2006.
7. Amundsen G, Wehrle D. HealthPartners: 2004 clinical indicators report. Available at: <http://www.healthpartners.com/files/24880.pdf>. Accessed October 3, 2005.
8. Williams SC, Schmaltz SP, Morton DJ, Koss RG, Loeb JM. Quality of care in U.S. hospitals as reflected by standardized measures, 2002-2004. *N Engl J Med*. 2005;353:255-264.
9. Joint Commission on Accreditation of Healthcare Organizations. Attributes of core performance measures and associated evaluation criteria. Available at: <http://www.jcaho.org/pms/core+measures/attributes+of+core+performance+measures.htm>. Accessed January 15, 2006.
10. Hennekens CH, Albert CM, Godfried SL, Gaziano JM, Buring JE. Adjunctive drug therapy of acute myocardial infarction. *N Engl J Med*. 1996;335:1660-1667.
11. Antithrombotic Trialists' Collaboration. Collaborative meta-analysis of randomized trials of antiplatelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients. *BMJ*. 2002;324:71-86.
12. Dellinger EP, Hausmann SM, Bratzler DW, et al. Hospitals collaborate to decrease surgical site infections. *Am J Surg*. 2005;190:9-15.