

SOUNDING BOARD

Accountability Measures — Using Measurement to Promote Quality Improvement

Mark R. Chassin, M.D., M.P.P., M.P.H., Jerod M. Loeb, Ph.D., Stephen P. Schmalz, Ph.D.,
and Robert M. Wachter, M.D.

Measuring the quality of health care and using those measurements to promote improvements in the delivery of care, to influence payment for services, and to increase transparency are now commonplace. These activities, which now involve virtually all U.S. hospitals, are migrating to ambulatory and other care settings and are increasingly evident in health care systems worldwide. Many constituencies are pressing for continued expansion of programs that rely on quality measurement and reporting.

In this article, we review the origins of contemporary standardized quality measurement, with a focus on hospitals, where such programs have reached their most highly developed state. We discuss some lessons learned from recent experience and propose a conceptual framework to guide future developments in this fast-moving field. Although many of the points we make are relevant to all kinds of quality measurement, including outcome measures, we focus our comments on process measures, both because these account for most of the measures in current use and because outcome measures have additional scientific challenges surrounding the need for case-mix adjustment. We write not as representatives of the Joint Commission articulating a specific new position of that group, but rather as individuals who have worked in the fields of quality measurement and improvement in a variety of roles and settings over many years.

A BRIEF HISTORY OF HOSPITAL
QUALITY MEASUREMENT AND
REPORTING IN THE UNITED STATES

Although the ubiquity of quality measurement and reporting makes it difficult to remember a health care landscape without them, these trends are re-

markably recent. In 1998, the Joint Commission launched its ORYX initiative, the first national program for the measurement of hospital quality, which initially required the reporting only of non-standardized data on performance measures.¹ In 2002, accredited hospitals were required to collect and report data on performance for at least two of four core measure sets (acute myocardial infarction, heart failure, pneumonia, and pregnancy)²; these data were made publicly available by the Joint Commission in 2004.

When the program started, no consensus existed regarding the kinds of measures on which data should be gathered by hospitals, no data on quality of care were collected systematically by hospitals, and little information on nationally standardized measures of hospital quality was available to the public. Few hospitals used national data on quality measures to improve clinical care processes; in fact, hospitals strongly resisted collecting data on quality measures and reporting them publicly.

The changes over the past decade have been breathtaking. The National Quality Forum has endorsed more than 600 quality measures.³ In 2004, the Centers for Medicare and Medicaid Services (CMS) began financially penalizing hospitals that did not report to the CMS the same performance data they collected for the Joint Commission, and in 2005, the CMS began its own public reporting.^{4,5} Today, hospitals provide data to the Joint Commission from a selection of 57 inpatient measures; currently, 31 of these are publicly reported, and there are plans to add the remaining, newly implemented measures over time.^{6,7} The CMS also includes additional data on patient satisfaction and outcomes (death and readmissions) for common medical conditions such as pneumonia and heart failure.

 THE EFFECT OF QUALITY
MEASUREMENT

As we consider the effect of this new quality-measurement and reporting effort, there is much to celebrate. Many measures are quite robust, with tight, evidence-based links between process performance and patient outcomes. With the use of these measures, we have seen gratifying improvements in the performance of hospitals. For example, in 2009, a total of 98.3% of eligible patients with acute myocardial infarction received a beta-blocker at hospital discharge, as compared with 87.3% of such patients in 2002⁸ (the Joint Commission's hospital performance-measure data warehouse; 2009 data will be available to the public in September 2010). Equally important, the consistency of hospital performance on key quality measures — such as prescribing beta-blockers and angiotensin-converting-enzyme inhibitors (or angiotensin-receptor blockers) to patients with an acute myocardial infarction and, in selected patients undergoing surgery, administering and discontinuing prophylactic antibiotics at the appropriate times to reduce surgical site infections — has increased dramatically in recent years (Fig. 1). For example, in 2009 (data available to the public in September 2010), 96.8% of hospitals showed performance levels greater than 90% in administering beta-blockers at discharge to patients who had had an acute myocardial infarction, as compared with 49.1% in 2002.

Because these quality-measurement and reporting programs were not implemented with the use of an experimental design, and virtually all U.S. hospitals participate in them, it is not possible to know how many of these improvements would have occurred in the absence of standardized measurement, Joint Commission accreditation requirements, public reporting, or the threat of Medicare payment penalties. On the other hand, no other national data on quality of which we are aware show such high levels of performance, nor are there other national examples of the greatly narrowed variation around high levels of performance that these data currently exhibit.

This quality-measurement and improvement effort is not without cost. Although some information can be collected relatively inexpensively from administrative data sets, many data elements — particularly those that capture the granular clinical detail that make the data credible — require painstaking and expensive review of medical records,

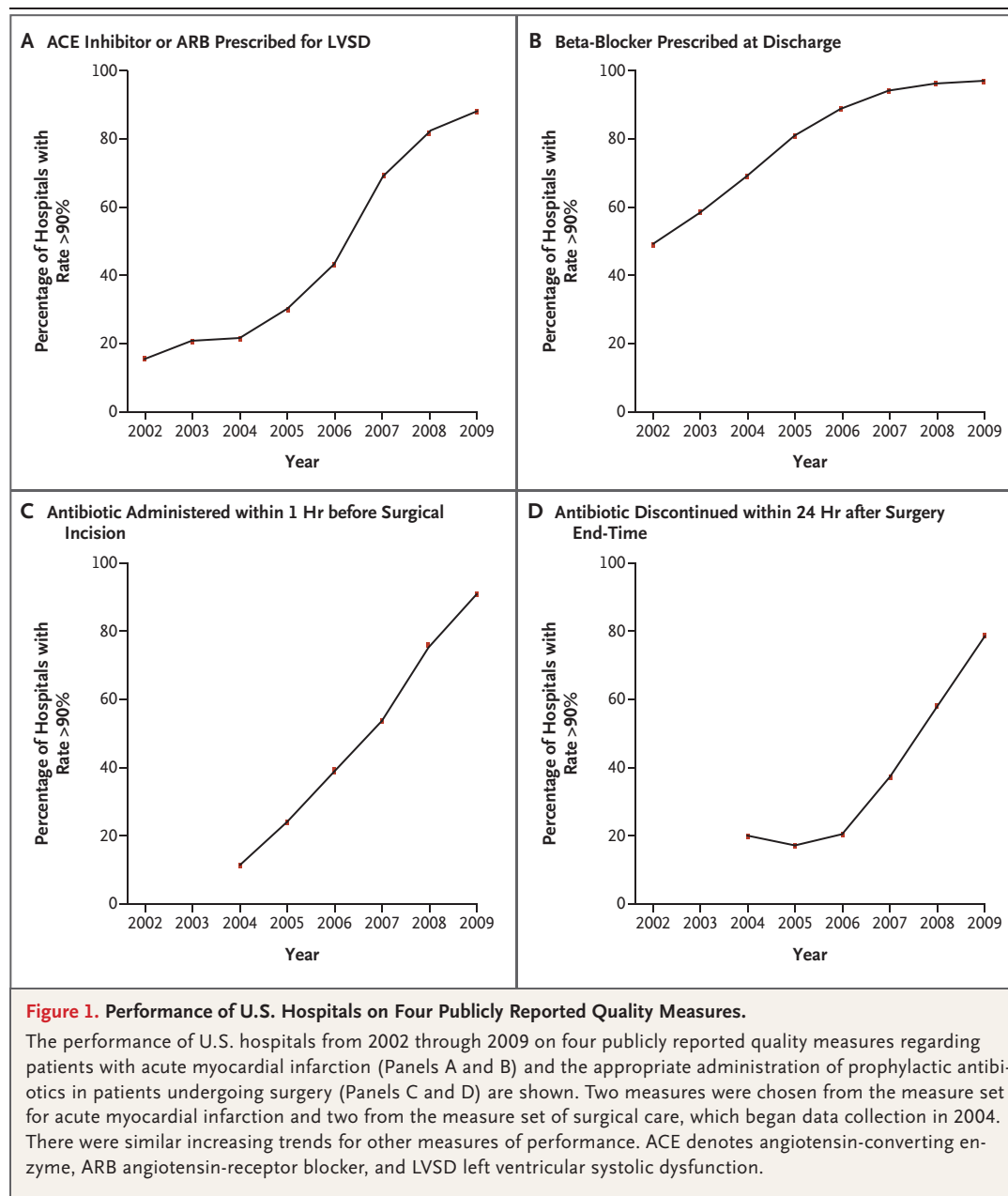
most of which are paper records. The requirements are such that a small industry of performance-measurement-system vendors, extensively vetted and operating under stringent quality standards, supports the ORYX initiative.⁹ The Joint Commission and the CMS have worked hard to ensure that in the case of the measures that are common to both programs, definitions and requirements for data collection are identical, allowing most data elements to be collected only once. ORYX vendors then submit the same data to both the Joint Commission and the CMS, satisfying both accreditation and payment requirements.¹⁰

In other words, over the past decade we have learned that standardized data can be collected by thousands of hospitals to identify and implement substantial improvements in care. Although measure specifications must keep up with emerging and evolving science, these challenges have not proved to be insurmountable. We believe that the “proof of concept” phase of national quality measurement and public reporting has now been completed.

 ROOM FOR IMPROVEMENT

Despite the progress that has been made, even proponents of the national quality programs of the Joint Commission and the CMS identify room for improvement. To address legitimate concerns about the program, we propose that such programs now focus explicitly on maximizing health benefits to patients. Achieving this goal requires examining closely the roster of measures currently included in these programs, establishing criteria to separate measures that advance this goal from those that do not, and replacing poorly performing measures with better ones. To make these goals operational, we suggest that all quality measures used in national transparency and payment programs — both existing ones and proposed new ones — be vetted against four criteria.

First, a measure must be based on a strong foundation of research showing that the process addressed by the measure, when performed correctly, leads to improved clinical outcomes. We note here that a strong foundation means more than one study, however persuasive any single investigation might be. We do not expect that this evidence base will consist solely of data from randomized trials, though much of it will. We believe that a high bar, one that exceeds the typical standard used for the development of practice



guidelines, is appropriate for measures that are used in national programs of quality measurement and improvement, since these programs affect thousands of hospitals and millions of patients. Fortunately, the state of the science has advanced to the point that we now have many measures from which to choose that meet this criterion.

Second, the measurement strategy must accurately capture whether the evidence-based care has been delivered. For example, the Joint Commission and the CMS currently measure aspirin administration after an acute myocardial infar-

tion by reviewing a medication-administration record (or its equivalent) — a measure that genuinely captures the process of interest.¹¹ On the other hand, we measure the presence of comprehensive discharge planning and of smoking-cessation counseling by whether a clinician has checked off a box or otherwise documented that such activities occurred. We know that for patients with heart failure, comprehensive education at discharge and coordination of care after discharge lead to improvements in functional outcomes, reductions in emergency department vis-

Table 1. Four Criteria for Accountability Measures That Address Processes of Care.

1. There is a strong evidence base showing that the care process leads to improved outcomes.
2. The measure accurately captures whether the evidence-based care process has, in fact, been provided.
3. The measure addresses a process that has few intervening care processes that must occur before the improved outcome is realized.
4. Implementing the measure has little or no chance of inducing unintended adverse consequences.

its, and fewer hospitalizations,¹² but our current measure is incapable of judging the quality of the process (i.e., whether the process is delivered with sufficient effectiveness to make improved outcomes likely). Organizations that wish to improve their performance record may be tempted to create clever discharge-instruction forms with just the right check-boxes and printed information summaries to satisfy the chart reviewers' rules concerning compliance with the measure, instead of doing the hard work of improving their clinical care. We were, therefore, not surprised when researchers recently found no relationship between hospital performance on the discharge-instruction measure for heart failure and readmission rates.¹³ We need a better measure for this important process; until we find one, measuring a check-box serves only to give us a false sense of accomplishment and reward "gaming."

Third, the measure should address a process quite proximate to the desired outcome, with relatively few intervening processes. Measures of appropriately administered medications meet this test, whereas the measure calling for an assessment of left ventricular function in patients with heart failure does not. With respect to the latter measure, although all patients with heart failure should have their ventricular function measured at some point, many other correctly performed clinical processes must occur after the test has been performed for the patient to have an improved outcome. The beneficial effect of processes as far upstream from outcomes as this one will be nullified if important processes closer to the outcome are not performed effectively. In such cases, we believe that the measurement of these processes is of little value, especially in the hospital inpatient setting. This criterion should be applied somewhat differently in ambulatory care settings, where it will be appropriate for some accountability measures to address processes that are quite upstream

from outcomes, such as measures of the evidence-based use of mammography or Pap smears. Even in these cases, though, we believe that such upstream measures will be inadequate by themselves to serve as accountability measures. To provide a more complete assessment of quality, they should be coupled with measures of more downstream processes, such as the timeliness of follow-up and communication of results and the occurrence and appropriateness of definitive treatment when abnormal test results are found.

Fourth, the measure should have minimal or no unintended adverse consequences. Some evidence suggests that administering the first dose of an antibiotic to a patient with community-acquired pneumonia within the first several hours after the patient's arrival at the hospital improves outcomes.¹⁴ However, the initial Joint Commission and CMS measure of that process (first dose of antibiotic within 4 hours [later relaxed to 6 hours] after arrival at the hospital) undoubtedly led to the inappropriate administration of antibiotics to patients who did not truly have pneumonia.^{15,16} Although "diagnostic uncertainty" was added to the measure criteria as a data element, permitting hospitals to exclude some such patients, the fundamental flaw in the measure remains.¹⁶

In summary, measures currently used in national quality programs that do not meet the criteria for accountability measures include: three measures concerning smoking-cessation counseling — those for adults with acute myocardial infarction, adults with heart failure, and adults with pneumonia — and the measure concerning discharge instructions for patients with heart failure, because these measures fail to accurately capture the care process; a measure concerning the evaluation of left ventricular systolic function in patients with heart failure, because it is not sufficiently proximate to the outcome; and a measure calling for the initial administration of antibiotics in patients with pneumonia within 6 hours after the patient's arrival at the hospital, because it has the potential to cause adverse consequences (see Table A in the Supplementary Appendix, available with the full text of this article at NEJM.org).

A WAY FORWARD — A FOCUS ON
ACCOUNTABILITY MEASURES

We believe that measures that meet all four criteria (Table 1) will have the greatest likelihood of

Table 2. Improvement in Performance on Accountability Core Measures from 2002 through 2009.*

Year	No. of Core Measures	No. of Accountability Measures	Median No. of Accountability Measures per Hospital†	No. of Hospitals Reporting‡	No. of Opportunities to Provide Care in Accordance with Measures‡	Overall Performance on All Accountability Measures§	Hospitals with >90% Performance†§
						percent	
2002	16	8	5	3250	957,000	81.8	20.4
2003	16	8	5	3286	2,173,000	83.9	24.6
2004	25	16	12	3254	3,651,000	83.3	16.5
2005	25	16	12	3225	4,490,000	84.9	21.9
2006	30	20	12	3283	5,322,000	88.2	41.5
2007	34	24	12	3170	7,911,000	90.0	60.0
2008	31	22	16	3178	13,222,000	93.1	70.8
2009	31	22	16	3123	12,476,000	95.4	85.9

* Data are from the Joint Commission's hospital performance-measure data warehouse.

† For data in this column, in each year, hospitals are included only if they reported a minimum of 30 cases across all their accountability measures.

‡ The numbers in this column represent the sum of all opportunities across all hospitals and all accountability measures.

§ The temporal trends were similar when the analysis was restricted to the subgroup of 2662 hospitals that reported data on acute myocardial infarction, heart failure, and pneumonia for all 8 years.

improving patient outcomes. Therefore, although other measures may be useful for internal quality-improvement purposes, we propose that only those measures that meet all four criteria be used for purposes of accountability (e.g., for accreditation, public reporting, or pay-for-performance). Of the 28 Joint Commission 2010 core measures that are aligned with Medicare, we believe that 22 meet all four criteria and could be deemed “accountability measures” (see Table B in the Supplementary Appendix).

Achieving the goal of improving health outcomes requires, of course, that hospitals make improvements in the clinical processes of care assessed by these accountability measures. Experience to date shows that such improvement is taking place at an accelerating pace. Table 2 shows the progress that hospitals have made in improving their performance on these measures — from a performance rate of 81.8% in 2002 to a rate of 95.4% in 2009. Moreover, by 2009, among all 3123 reporting hospitals, the 22 accountability measures that were in use at that time assessed about 12.5 million opportunities to provide specific elements of evidence-based care. The percentage of hospitals whose performance across all their accountability measures exceeded 90% increased substantially — from 20.4% in 2002 to 85.9% in 2009.

CHALLENGES IN IMPLEMENTING A PROGRAM OF ACCOUNTABILITY MEASURES

Implementing these criteria presents several challenges to all the key stakeholders, but we believe that these challenges are manageable. We recognize that many current measures will not meet the stringent accountability criteria. We need to be certain that measures that do not qualify still remain available for other important purposes, when they are appropriate. For example, individual health care organizations could consider using them for their own quality initiatives. After local experimentation and modification, some may ultimately be added to the set of accountability measures.

There are other challenges as well. A narrow focus on quality measures in hospitals may miss the importance of postdischarge care for a patient — for example, a patient with heart failure. The proposed development of bundled payments and accountable care organizations may facilitate the development of inpatient and outpatient measures that are more integrated, which will be particularly useful when high-quality care requires the coordination of care across the continuum. As indicated earlier, the four criteria for accountability measures may require some adaptation for

the assessment of ambulatory care. We believe, however, that these criteria can serve as a useful framework for identifying accountability measures in nonhospital settings.

Finally, the process of improving our system of high-stakes quality measurement requires perpetual vigilance. Although some unintended adverse consequences can be anticipated and avoided during the initial evaluation of a measure, others may not become evident until many hospitals use the measures. A vital part of this program, largely absent today, will be a formal process of assessing experience with the measures and using that information to improve the development of measures and decisions regarding deployment.¹⁶

THE GOAL — MEASUREMENT
FOR IMPROVEMENT

We call on all stakeholders that promulgate, support, or advocate for programs that use incentives of various sorts designed to promote quality in hospitals and health systems and among physicians to consider adopting this framework for accountability measures. For its part, the Joint Commission is incorporating this framework into its programs. We believe that the time is right for such a consensus to emerge. Far from the past attitude of resistance to all measurement, hospitals and physicians have embraced the measurement, and even the reporting, of robust and authentic quality metrics as an important mechanism to drive the improvement of clinical processes. In doing so, they have achieved substantial gains that have undoubtedly saved thousands of lives.

Fortunately, as the science has advanced, we now have a surfeit of measures that meet all four accountability criteria with which to populate accreditation, public reporting, and pay-for-performance programs. Eliminating measures that do not pass these accountability tests and replacing them with ones that do will reduce unproductive work on the part of hospitals, enhance the credibility of the program with physicians and other key stakeholders, and increase the positive effect that all these programs will have on health outcomes for patients.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

From the Joint Commission, Oakbrook Terrace, IL (M.R.C., J.M.L., S.P.S.); and the Department of Medicine, University of California, San Francisco, San Francisco (R.M.W.).

This article (10.1056/NEJMs1002320) was published on June 23, 2010, at NEJM.org.

1. Lee KY, Loeb JM, Nadzam DM, Hanold LS. An overview of the Joint Commission's ORYX initiative and proposed statistical methods. *Health Serv Outcomes Res Methodol* 2000;1:63-73.
2. 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;353:255-64.
3. NQF-endorsed standards. Washington, DC: National Quality Forum. (Accessed May 27, 2010, at http://www.qualityforum.org/Measures_List.aspx.)
4. Centers for Medicare and Medicaid Services. Medicare modernization update. (Accessed May 27, 2010, at <http://www.cms.hhs.gov/MMAUpdate/>.)
5. *Idem*. Hospital quality initiatives: Hospital Quality Alliance. (Accessed May 27, 2010, at http://www.cms.hhs.gov/HospitalQualityInits/33_HospitalQualityAlliance.asp.)
6. The Joint Commission. 2009 ORYX performance measure reporting requirements for hospitals and guidelines for measure selections. (Accessed May 27, 2010, at http://www.jointcommission.org/NR/rdonlyres/64C5EDF0-253A-42CE-9DE5-C8AFE6755267/0/09_Oryx_hap.pdf.)
7. *Idem*. Performance measurement initiatives. (Accessed May 27, 2010, at <http://www.jointcommission.org/PerformanceMeasurement/PerformanceMeasurement/default.htm>.)
8. *Idem*. Improving America's hospitals: the Joint Commission's annual report on quality and safety. 2009. (Accessed May 27, 2010, at http://www.jointcommission.org/NR/rdonlyres/22D58F1F-14FF-4B72-A870-378DAF26189E/0/2009_Annual_Report.pdf.)
9. *Idem*. Listed systems supporting core measures. (Accessed May 27, 2010, at http://www.jointcommission.org/NR/rdonlyres/3E3D16BE-3D28-47B9-86C3-DBC4D736407D/0/CoreDirectory3_24.pdf.)
10. Hospital quality data: CMS needs more rigorous methods to ensure reliability of publicly released data. Washington, DC: Government Accountability Office, January 2006. (Accessed May 27, 2010, at <http://www.gao.gov/new.items/d0654.pdf>.)
11. The Joint Commission and Centers for Medicare and Medicaid Services. Specifications manual for national hospital quality measures. 2009. (Accessed May 27, 2010, at <http://www.jointcommission.org/performancemeasurement/performancemeasurement/historical+nhqm+manuals.htm>.)
12. Phillips CO, Wright SM, Kern DE, Singa RM, Shepperd S, Rubin HR. Comprehensive discharge planning with postdischarge support for older patients with congestive heart failure: a meta-analysis. *JAMA* 2004;291:1358-67.
13. Jha AK, Orav EJ, Epstein AM. Public reporting of discharge planning and rates of readmissions. *N Engl J Med* 2009;361:2637-45.
14. Houck PM, Bratzler DW, Nsa W, Ma A, Bartlett JG. Timing of antibiotic administration and outcomes for Medicare patients hospitalized with community-acquired pneumonia. *Arch Intern Med* 2004;164:637-44.
15. Kanwar M, Brar N, Khatib R, Fakhri MG. Misdiagnosis of community-acquired pneumonia and inappropriate utilization of antibiotics: side effects of the 4-h antibiotic administration rule. *Chest* 2007;131:1865-9.
16. Wachter RM, Flanders SA, Fee C, Pronovost PJ. Public reporting of antibiotic timing in patients with pneumonia: lessons from a flawed performance measure. *Ann Intern Med* 2008;149:29-32.

Copyright © 2010 Massachusetts Medical Society.