

Is Zero the Ideal Death Rate?

Thomas H. Lee, M.D., David F. Torchiana, M.D., and James E. Lock, M.D.

Massachusetts recently joined New York, New Jersey, and Pennsylvania in reporting death rates associated with cardiac surgery for individual surgeons — another wave in the tide of public reporting that is sweeping the country. Such reporting raises questions about distinguishing the goals that define one's work from the targets used to measure success. For hospitals and physicians, minimizing death and complications is an undisputed goal. But is zero the ideal target for measures of performance with respect to death and complications?

In this era of public accountability, the answers to questions such as this turn out to be surprisingly complex. Should providers be ranked according to their patients' death rates? Could quality measures that drive providers toward apparent perfection sometimes paradoxically compromise quality by encouraging excessive focus on the measures themselves?

As strategies for improving health care by measuring quality are implemented through pay-for-performance contracts and public "report cards," subtle but important differences among performance measures are emerging. The simplest performance measures describe compliance with a standard of care, such as administration of aspirin after acute myocardial infarction in the absence of contraindications. If the denominator for these measures is defined correctly, perfect performance should be achievable

through education and the adoption of systems (e.g., electronic medical records with decision support) that make it easier to "do the right thing" — or difficult or impossible not to do it. Indeed, physicians in many hospitals administer aspirin after acute myocardial infarction in 100% of cases.

For a subgroup of measures of this type, perfect performance may be unattainable. For example, it is impossible to reduce blood pressure below 140/90 mm Hg in 100% of patients with hypertension because therapy has side effects in some of them. With such measures, "benchmarks" from organizations with superior performance can serve as targets. Sometimes, seemingly unattainable targets turn out to be attainable, as demonstrated by the achievement of zero or near-zero rates of catheter-related infections in some intensive care units.¹

But for a second type of measure, no ideal standard of care exists, so perfect performance cannot be defined. Instead of measuring "defects," data are used to describe variation. Physicians whose performance is higher or lower than the mean might have reason to question their care. For example, back when clinical judgment was the sole tool for deciding whether to operate on patients with suspected acute appendicitis, there was no ideal proportion of removed appendixes that turned out to be normal. Low rates suggested that surgeons

might be too slow to operate; high rates suggested too low a threshold. Other examples include the percentage of coronary angiograms that reveal normal coronary arteries and the percentage of patients with acute bronchitis who are treated with antibiotics. Providers often use such measures in internal quality-improvement programs, acknowledging that having higher or lower rates than colleagues may be cause for concern.

The question is: Do rates of death and complications fall into the first category or the second? At first glance, these measures have a clear goal: zero. However, public performance measures that push providers toward apparent perfection in these areas can have unintended perverse consequences.

Consider, for instance, balloon valvuloplasty for congenital aortic stenosis. If the only measured outcome is safety, one set of incentives is created. Any cardiologist knows how to minimize the complication rate: use a smaller balloon. But doing so also reduces the hemodynamic benefit and the time before another procedure is needed. Bigger balloons produce better relief of obstruction but carry a slightly higher risk of complications. Thus, two measures are needed to assess quality — one measure of effectiveness and one of safety. The interventionalists who deliver the best care with maximal benefit may well have complication rates in the middle

of the bell-shaped curve, not at the low end.

A similar dynamic applies in other domains, with trade-offs made through the choice of therapeutic strategy or referral decisions. Providers seeking to improve their measured performance are tempted to keep high-risk patients out of the denominator for a performance measure, choosing treatment options on which their performance is not being measured — even if these options have a lower expected survival rate.

For example, patients with a ruptured papillary muscle after acute myocardial infarction have a high risk of death with surgery but almost no hope of survival with medical therapy. A decision not to operate in such a case might help a surgeon preserve a low death rate but seal the patient's fate. Cardiologists who perform angioplasty in patients with acute myocardial infarction and shock improve long-term survival by 67%² but also raise their procedural death rate.

No one knows how often physicians and hospitals try to improve measured results by declining high-risk cases. One might argue that physicians should think twice before operating on very elderly patients and that tertiary care hospitals should turn away more patients who have very little chance of surviving. In theory, risk-adjustment methods allow statistical analyses to correct for differences in severity of illness so that treating high-risk patients should not adversely affect measured performance. With large enough patient populations, death and complication rates should be

dominated by “typical” patients, minimizing the risk of distortion by a few high-risk patients.

Two problems prevent this theoretical ideal from comforting practitioners. The first is the limitations of risk-adjustment meth-

***No one knows
how often
physicians and
hospitals try to
improve measured
results by
declining high-risk
cases.***

ods, which cannot adjust for variables that are not measured or included in the models, such as socioeconomic status or clinical information not reflected in administrative data. Robust risk-adjustment models are available for very few conditions, yet performance measures focusing on death related to abdominal aortic aneurysm repair, stroke, and pneumonia are being tested or are in actual use.³

The second problem is that small sample sizes lead to unstable estimates of performance. Consider the example of mortality associated with coronary-artery bypass grafting (CABG). When busy cardiac surgeons average 100 to 200 cases per year and the average mortality is about 2% (i.e., two to four deaths per year), each death makes a visible difference in a surgeon's ranking. The promise of risk adjustment may be in-

sufficient to persuade a surgeon to operate on a patient with diabetes and severe angina who has three-vessel coronary disease and poor left ventricular function, even though surgery is likely to benefit such a patient. Any theoretical leveling of the playing field accomplished by risk adjustment is, well, theoretical. For physicians being ranked on their individual performance with small sample sizes, the surest path to high marks is to have death rates as close to zero as possible.

For most outcome measures, including rates of death and complications, there is a bell-shaped performance curve, within which there can be a fair amount of lability. The Northern New England Cardiovascular Disease Study Group and others have found that a program ranking at the top one year might drop down the next and bounce back up the year after. Some variations reflect real changes within a group; others are presumably related to changes in patient-related factors that are not captured in the model. For any given year, it's hard to know which factors explain who ends up where within the curve. Opinions on how to deal with this uncertainty depend on one's vantage point on the health care system. Physicians and other providers tend to advocate a “first, do no harm” approach that emphasizes restraint in public reporting. At the other extreme, many employers and health plans favor releasing all available data to the public to create informed health care consumers and to motivate providers to improve quality.

Experiments with ways of tack-

ling this challenge are ongoing here and abroad. The New York State reporting system is widely credited with improving surgical outcomes. In contrast, Canadian investigators recently reported that in Ontario, mortality associated with CABG was sharply reduced after providers' results were confidentially disclosed at an institutional level but that public reporting had no added effect on performance.⁴

There may be no compromise that would satisfy everyone, but a middle-ground approach may be necessary to improve care so that we move the entire bell-shaped curve, not just track movement

within it. Such an approach might combine group-level measurement for public reporting and pay-for-performance contracts with confidential use of ranked outcomes at an individual-provider level for regional quality-improvement "collaboratives."

We are fortunate to live in an age in which the pursuit of perfection is becoming the culture of medicine. Nevertheless, measures that drive providers toward apparently perfect performance should be handled with care.

Dr. Lee is network president at Partners Healthcare System and an associate editor of the *Journal*. Dr. Torchiana is chief executive officer of Massachusetts General Phy-

sicians Organization and associate professor of surgery at Harvard Medical School, and Dr. Lock is cardiologist in chief at Children's Hospital and a professor of pediatrics at Harvard Medical School — all in Boston.

1. Pronovost P, Needham D, Berenholtz S, et al. An intervention to decrease catheter-related bloodstream infections in the ICU. *N Engl J Med* 2006;355:2725-32.
2. Hochman JS, Sleeper LA, Webb JG, et al. Early revascularization and long-term survival in cardiogenic shock complicating acute myocardial infarction. *JAMA* 2006; 295:2511-5.
3. Agency for Healthcare Research and Quality. National Quality Measures Clearinghouse. (Accessed June 20, 2007, at <http://www.qualitymeasures.ahrq.gov>.)
4. Guru V, Fries SE, Naylor CD, et al. Public versus private institutional performance reporting: what is mandatory for quality improvement? *Am Heart J* 2006;152:573-8.

Copyright © 2007 Massachusetts Medical Society.