

cal improvement of protocols or consent forms.³ On the contrary, this practice seems to pose a significant risk of diminishing studies' ethical integrity. Fortunately, some ways of changing this system are being explored. Recently, the Office for Human Research Protections put out for public comment a proposal to receive direct authority to take action against IRBs — as distinct from the institutions conducting the research — for noncompliance with regulations.⁴ The intent is to encourage greater reliance on outside (and central) IRBs by assuring the individual institutions participating in multisite studies that they would not be blamed if an outside IRB were responsible for violations.

Another approach to reducing the number of IRB reviews would be to have sponsors require the use of a central IRB as a condition for participating in a study. Nothing in the existing U.S. regulations would prevent them from doing

so. The Department of Veterans Affairs currently operates exactly such a system for a select group of studies. In an attempt to constrain the duplication of review efforts for international multisite studies, the European Union is taking a different approach: it now restricts each participating country to a "single opinion" representing the ethics review for that country, "notwithstanding the number of Ethics Committees" involved.⁵

Any one or a combination of these approaches may turn out to be satisfactory. But recognizing that the problem with multiple-IRB review relates not merely to wasted time and effort but also to less-than-optimal protection of people who volunteer to participate in research should add urgency to our efforts to solve this problem.

The views expressed in this article are those of the author and are not necessarily those of the U.S. Department of Health and Human Services or its operating division, the Office of the Assistant Secretary for Health.

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

Dr. Menikoff is the director of the Office for Human Research Protections, Rockville, MD.

This article (10.1056/NEJMp1005101) was published on October 13, 2010, at NEJM.org.

1. National conference on alternative IRB models: optimizing human subject protection, November 19–21, 2006. (<http://www.aamc.org/research/irbreview/irbconf06rpt.pdf>)
 2. NCI position paper on PedCIRB conflicts of interest, July 27, 2005. (<http://www.hhs.gov/ohrp/requests/NCIpos.html>)
 3. Infectious Diseases Society of America. Grinding to a halt: the effects of the increasing regulatory burden on research and quality improvement efforts. *Clin Infect Dis* 2009; 49:328-35.
 4. Office for Human Research Protections. Request for information and comments on IRB accountability. *Fed Regist* 2009;74:9578-83. (<http://edocket.access.gpo.gov/2009/pdf/E9-4628.pdf>)
 5. Directive 2001/20/EC of the European Parliament and of the Council of 4 April 2001. On the approximation of the laws, regulations and administrative provisions of the member states relating to the implementation of good clinical practice in the conduct of clinical trials on medicinal products for human use. (<http://www.eortc.be/Services/Doc/clinical-EU-directive-04-April-01.pdf>)
- Copyright © 2010 Massachusetts Medical Society.

Public Release of Clinical Outcomes Data — Online CABG Report Cards

Timothy G. Ferris, M.D., M.P.H., and David F. Torchiana, M.D.

On September 7, 2010, Consumers Union (publisher of *Consumer Reports*) reported the results of coronary-artery bypass grafting (CABG) procedures at 221 U.S. cardiac surgery programs.¹ The voluntary reporting of risk-adjusted outcomes in approximately 20% of U.S. cardiac surgery programs is a watershed event in health care accountability.

The reported ratings derive from a registry developed by the Society of Thoracic Surgeons (STS) in 1989. More than 90% of the approximately 1100 U.S. cardiac surgery programs participate in

the registry. Registry data are collected from patients' charts and include key outcomes such as complications and death, the severity of preoperative illness, co-existing conditions, surgical technique, and medications. These data are maintained by the Duke Clinical Research Institute and are analyzed with the use of well-tested statistical methods. The data-collection and auditing methods, specifications of the measures, and statistical approaches have evolved over the course of two decades and reflect a substantial commitment by

cardiac surgeons and their leadership.^{2,3}

For years, participants in the STS registry have been examining these data and using them to make improvements. What does the public now get to see? Each surgical program that has chosen to make its data public is assigned a rating of one, two, or three stars. Stars are assigned on the basis of results on 11 performance measures (see table) that have been endorsed by the National Quality Forum. The rating depends on whether the risk-adjusted outcomes in a program fall be-

Measures of Quality Used by the Society of Thoracic Surgeons in the Ratings of Coronary-Artery Bypass Grafting (CABG) Programs.

Measure	Description
Postoperative renal failure	Percentage of patients (without preexisting renal failure) undergoing isolated CABG in whom postoperative renal failure developed or dialysis was required
Surgical reexploration	Percentage of patients undergoing isolated CABG who required a return to the operating room because of bleeding, tamponade, graft occlusion, or other cardiac reason
Antiplatelet medication at discharge	Percentage of patients undergoing isolated CABG who were receiving aspirin, safety-coated aspirin, or clopidogrel at discharge
Beta-blockade at discharge	Percentage of patients undergoing isolated CABG who were receiving beta-blockers at discharge
Antilipid treatment at discharge	Percentage of patients undergoing isolated CABG who were receiving a statin or other pharmacologic lipid-lowering regimen at discharge
Risk-adjusted operative mortality after CABG	Percentage of patients undergoing isolated CABG who died during the hospitalization in which the CABG was performed or within 30 days after the procedure
Preoperative beta-blockade	Percentage of patients undergoing isolated CABG who received beta-blockers within 24 hours before surgery
Prolonged intubation (ventilation)	Percentage of patients undergoing isolated CABG (without preexisting intubation or tracheostomy) who required intubation for more than 24 hours
Rate of deep sternal-wound infection	Percentage of patients undergoing isolated CABG in whom a deep sternal-wound infection developed within 30 days after the procedure
Stroke or cerebrovascular accident	Percentage of patients (without preexisting neurologic deficit) undergoing isolated CABG in whom a postoperative neurologic deficit developed that persisted for more than 24 hours
CABG using an internal thoracic artery	Percentage of CABG performed using an internal thoracic artery

low, are equal to, or exceed the average performance range. The performance thresholds are designed to ensure a 99% probability that outlier programs — those rated significantly below or above the mean and therefore given one and three stars, respectively — are truly below or above average. With the use of this method, 23 to 27% of the programs have been identified as outliers over the past 3 years. In addition to the star rating for overall performance, consumers see the star rating and actual performance scores (on a scale from 0 to 100) in four subcategories: 30-day survival (“patients have a 98% chance of surviving at least 30 days after the procedure and of being discharged from the hospital”), complications (“patients have an 89% chance of avoiding all five of the major complications”), use of appropriate medications (“patients have a 90% chance of receiving

all four of the recommended medications”), and surgical technique (“patients have a 98% chance of receiving at least one optimal surgical graft”).

The move on the part of the STS to make results available to the public will certainly trigger a cascade of responses. Advocates of transparency will point to the shortcomings of the ratings — the voluntary and therefore selective participation of programs (50 of the programs that have chosen to report their data have received three stars, whereas only 5 have received one star), the lack of long-term outcomes (e.g., 10-year survival, graft patency, and functional improvement), and the lack of physician-specific ratings. Expect such advocates to push for more. Nonparticipating cardiac surgery programs will come under pressure to allow the outcomes in their programs to be reported. Physicians in other

surgical specialties that are amenable to this type of approach, such as orthopedics or vascular surgery, may be expected to follow suit. And this event will fuel the debate regarding the risks and benefits of public reporting, including the question of whether it assists patients in discriminating among sites of care. While these issues play out, several aspects of this release of ratings deserve attention.

First, years of pressure from policymakers, health care purchasers, and patient-advocacy groups to provide greater accountability played a major role in bringing this publication to fruition. Public reporting of outcomes has widespread support, and cardiac surgeons have been among the principal targets of these efforts. The first statewide report card on cardiac surgical performance was mandated in New York in 1989. Early experiences with pub-

lic reporting of the outcomes of cardiac surgery spurred efforts by the STS and others to improve cardiac surgery.⁴ Although some consumer advocates pushing for transparency may view this release as a glass four-fifths empty — given the selectivity and number of programs reporting — the external pressure has been critical in stimulating improvement efforts within the medical profession.

Second, the publication of definitive analyses derived from clinical data can be a double-edged sword for providers. When performance reports are based on administrative data, physicians often justifiably argue that the data are flawed and the conclusions suspect. In contrast, with these new ratings, not only have the participants endorsed the methods, but they have volunteered to display performance results that carry the imprimatur of the physicians' specialty society. Experience with performance reporting in Massachusetts has shown that when the data and analyses are as good as possible, a public report of suboptimal performance requires a substantive public response: state Department of Public Health officials suspended a Massachusetts cardiac surgery program to conduct an external review, amidst substantial media attention, when the program was identified as a high-mortality outlier.

Third, the process of moving clinical data from the STS registry into the public domain has been long, complex, and expensive. As a member-supported organization, the STS navigated treacherous waters to bring its members to the point of permitting the publication of their data. Some key decisions facilitated this process: the STS reported

group-level rather than physician-level data, rigorously validated its data-collection and risk-adjustment models, and selected a performance-classification system that maximized specificity. Such choices helped to mitigate physicians' biggest fear: the risk of misclassification. Moreover, cardiac surgery programs have been looking at these data for years, so there shouldn't be any surprises. The success that the STS has had in leading a nontrivial fraction of its members to agree to participate suggests that public reporting can be done in a way that doesn't alienate the profession.

There is no question about the need for accountability on the part of health care providers or the central role of measurement in the improvement of health care. Nonetheless, questions remain about the role of public reporting in improving health care. Performance measurements audited by regulators are one alternative, especially in situations in which the information is too complex for patients to use in discriminating among care sites. Insofar as public reporting drives improvement of all outcomes, it benefits everyone; insofar as risk aversion leads to changes in the population receiving an indicated service, the net effect can be nil or even negative.⁵ Given the heterogeneity in the delivery of medical services, it should come as no surprise that we have developed multiple methods for assessing performance and encouraging accountability. Regardless of which approach proves most beneficial to patients, public reporting will increasingly be a fact of life for physicians.

By publishing ratings using the best available data, the STS has responded to the public in a

way that attempts to both inform patients and mitigate physicians' fears. We hope that the experience of the STS can be applied to other initiatives that are aimed at bringing performance data derived from clinical sources to the public, thereby reducing the time and expense of this process. For example, this experience may contain lessons for the Centers for Medicare and Medicaid Services as it prepares to handle the wave of clinical data it will receive through the Physician Quality Reporting Initiative and the "meaningful use" program for electronic health records. At least some of these data will almost certainly be publicly reported. The STS's success suggests that reporting can be done in a way that physicians will support. Whether the STS approach is an anomaly or a precedent that other specialty groups will emulate remains to be seen.

Disclosure forms provided by the authors are available with the full text of this article at [NEJM.org](http://www.nejm.org).

From the Massachusetts General Physicians Organization, Massachusetts General Hospital, Boston.

This article (10.1056/NEJMp1009423) was published on September 7, 2010, at [NEJM.org](http://www.nejm.org).

1. Consumer Reports home page. (<http://www.consumerreportshealth.org>)
 2. Shahian DM, Edwards FH, Ferraris VA, et al. Quality measurement in adult cardiac surgery: part 1 — conceptual framework and measure selection. *Ann Thorac Surg* 2007; 83:Suppl:S3-S12.
 3. O'Brien SM, Shahian DM, DeLong ER, et al. Quality measurement in adult cardiac surgery: part 2 — statistical considerations in composite measure scoring and provider rating. *Ann Thorac Surg* 2007;83:Suppl:S13-S26.
 4. O'Connor GT, Plume SK, Olmstead EM, et al. A regional intervention to improve the hospital mortality associated with coronary artery bypass graft surgery. *JAMA* 1996;275: 841-6.
 5. Dranove D, Kessler D, McClellan M, Satterthwaite M. Is more information better? The effects of "report cards" on health care providers. *J Polit Econ* 2003;111:555-88.
- Copyright © 2010 Massachusetts Medical Society.