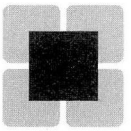

JAMES THOMAS, MD, EDITOR

Practice guidelines and physician scorecards: grading the graders

ROBERT A. VOGEL, MD, AND ERIC J. TOPOL, MD

AS MANAGED CARE takes root, physicians are finding their practices held up to scrutiny by their peers, by payers, and by the general public. Practice guidelines are being promulgated in many fields. Report cards evaluate the care physicians give patients, the skill of surgeons, and the overall quality of hospitals. But how good are the standards? In this month's Cardiology Dialogue, Robert A. Vogel, MD, of the University of Maryland and Eric J. Topol, MD, of the Cleveland Clinic Foundation, share their views and experiences in the brave new world of physician accountability.

See Editorial, page 75.

DR. VOGEL: Many physicians have gotten to the point in their education where they feel that examinations and evaluations are no longer necessary. But we in medicine are really subject to very little review, compared with airline pilots, for example. We also have very little standardization in our practice: we all read the literature about a particular problem and come to somewhat different conclusions about the correct course of action. In the past few years I have been working with several projects that are showing that there are deficiencies in how cardiology is practiced, and that formal practice guidelines and physician scorecards, properly designed and validated, can lead to improvements in the process of medicine and in patient outcomes.

■ This series is based on the Cleveland Clinic Heart Center's "Controversies in Cardiology" conferences, at which a visiting clinician-professor and a Cleveland Clinic Heart Center clinician give contrasting perspectives on the application of a current technology or the management of a cardiologic disease.

From the University of Maryland, Baltimore (R.A.V.) and the Cleveland Clinic Foundation (E.J.T.).

Address reprint requests to E.J.T., Department of Cardiology, F25, The Cleveland Clinic Foundation, 9500 Euclid Ave, Cleveland, OH 44195.

ASSESSING APPROPRIATENESS OF CARE

HCFA study of myocardial infarction care

DR. VOGEL: The first project I will discuss is the Cooperative Cardiovascular Project of the Health Care Financing Administration (HCFA), a four-state pilot study that assessed the care of 15 000 patients with acute myocardial infarction.[1] In conducting the study, we adapted 13 different criteria of process—not of outcome—from the American College of Cardiology-American Heart Association (ACC-AHA) guidelines.[2] For instance, was a patient given aspirin or counseled regarding smoking cessation? A person trained in data extraction went through each patient's chart and checked whether these things were actually done.

Almost all patients with acute myocardial infarction ought to have aspirin, and it ought to be started early. In our study, 83% of patients who had

no contraindications to aspirin did in fact receive it, but fewer than one third received it the day the myocardial infarction was diagnosed. Beta blockers were given to only 45% of patients who did not have heart failure or other contraindications to this therapy. Of the patients with heart failure, only 59% received angiotensin-converting enzyme (ACE) inhibitors. Advice regarding smoking cessation was documented in 28% of patients who smoked. In 35% of all cases, nobody documented the presence of chest pain or when it started.

TABLE 1
CONCORDANCE AMONG THREE SCORING SYSTEMS FOR APPROPRIATENESS OF CORONARY REVASCULARIZATION*

Scoring system†	Agree or minor disagreement (%)	Major disagreement (%)	Uncertain agreement (%)
RAND vs ACC–AHA	47	21‡	32
RAND vs RAS	61	7‡	32
RAS vs ACC–AHA	60	40‡	0

*Determined in 153 patients at the University of Maryland Medical Center
†RAND, the RAND corporation guidelines[3]; ACC–AHA, the American College of Cardiology and American Heart Association guidelines[2]; RAS, University of Maryland Revascularization Appropriateness Scoring system[4]
‡$P < .05$ among all groups

Admittedly, there are many weaknesses in the data. For example, a physician may counsel a patient to stop smoking, or to take aspirin after going home, but not document it in the chart. We recognize that. Nevertheless, this study underscores the need for objective assessment of physicians.

Report cards for residents

In another ongoing study, conducted at the University of Maryland, we give our residents report cards to remind them to document the presence of risk factors (hypertension, dyslipidemia, smoking, diabetes, and physical inactivity) in patients admitted to a coronary intensive care or telemetry unit. The report card also includes whether they prescribed vasoprotective drugs such as aspirin, beta blockers for post-Q-wave infarctions, ACE inhibitors for heart failure, and hormone replacement therapy for postmenopausal women. (The residents also document the presence of contraindications to these drugs.)

When we started using report cards, our residents documented 67% of risk factors, and treated 49% of them. We can argue about the importance of any of the particular risk factors, but when residents are given report cards, treatment improves. This year the treatment rate increased to 68%.

How good are the guidelines?

Although scorecards can illuminate how treatment guidelines are followed, it is still important to assess the guidelines themselves. Do different guideline systems tell you to do the same thing? And do patients treated according to the guidelines fare better than patients not treated according to the guidelines? In another project at the University of Maryland, we are addressing these questions by comparing three different sets of guidelines for appropriateness of angioplasty and coronary artery bypass grafting (CABG): the RAND guidelines,[3] those of the ACC-AHA,[2] and the University of Maryland Revascularization Appropriateness Scoring (RAS) system, which we developed.[4]

When we applied all three systems prospectively in patients in the catheterization lab, we found that the systems gave different recommendations about the appropriateness of a procedure (*Table 1*).[5] For angioplasty, the ACC-AHA criteria are very strict about what constitutes an appropriate procedure, whereas the RAND criteria are much more lenient. On the other hand, the ACC-AHA guidelines are very lenient about the appropriateness of bypass surgery. The RAND criteria yielded uncertain ratings in about 30% of clinical situations, more for angioplasty than for CABG. That high yield of uncertain ratings is one of the limitations of the RAND system.

We have followed up 153 patients for 15 months to see if their outcomes differed if they received treatment that was concordant or discordant with RAND, ACC-AHA, or RAS criteria (*Table 2*). In this cohort, 60% of the patients had undergone angioplasty, 18% had bypass surgery, and 22% were medically treated.[6] We found that the mortality rate was less in patients treated in accordance with ACC-AHA or RAS criteria, but not RAND criteria. We were amazed that these differences were apparent, given the small size of the cohort.

TABLE 2
ADVERSE OUTCOMES (%) WITH CONCORDANT VS DISCORDANT TREATMENT*

Outcome	Scoring system†		
	RAND	ACC–AHA	RAS
Mortality	9 vs 10	5 vs 19‡	5 vs 18‡
Myocardial infarction	4 vs 10	5 vs 3	4 vs 7
Percutaneous transluminal coronary angioplasty	10 vs 10	7 vs 19§	10 vs 9
Coronary artery bypass grafting	8 vs 38‡	11 vs 16	7 vs 25‡
Angina	36 vs 52	35 vs 51	34 vs 50§
Congestive heart failure	7 vs 14	7 vs 11	6 vs 14

*Determined in 153 patients at the University of Maryland Medical Center
†RAND, the RAND corporation guidelines[3]; ACC–AHA, the American College of Cardiology and American Heart Association guidelines[2]; RAS, University of Maryland Revascularization Appropriateness Scoring system[4]
‡$P < .01$
§$P < .05$

DR. TOPOL: Are you thinking about using the University of Maryland's RAS system in clinical practice?

DR. VOGEL: We do use it in clinical practice. Every time we do a catheterization report, we generate recommendations by all three systems. We do not have to follow the recommendations, but we are part of a capitated, managed-care plan, and we look at resource utilization carefully.

THE DOWNSIDE OF SCORECARDS

DR. VOGEL: Nonetheless, I think that scorecards have been more damaging than beneficial. They can help to improve practice, but raw and subjectively adjusted evaluations can lead to real problems.

In New York State in the early 1990s, the mortality rate of patients undergoing CABG fell by 21%.[7] This is encouraging, but the New York State Department of Health went on to calculate that the risk-adjusted mortality rate fell by 41%, as sicker patients were undergoing this procedure.[7] In reality, I think the operators "gamed" the system by rating their patients as sicker than they really were.

DR. TOPOL: I worked on the New York advisory committee, and I do not fully believe the decrease in risk-adjusted mortality either, especially considering that the prevalence of renal failure—a strong predictor of mortality—apparently increased by 70% during those years. But this was not an attempt to deceive anyone. Rather, the people who gathered the data are being more thorough now.

PROBLEMS WITH RATING HOSPITALS

DR. TOPOL: Another use of this type of exercise is to rate hospitals and physicians. We are in the embryonic stages of being able to compare institutions by outcome. Rather than jump into this, we need to work out some problems with the models, and figure out who is going to pay for it.

Rating operators and institutions was first done in bypass surgery, because the features that predict mortality had already been worked out from years of experience in thousands of patients, making risk adjustment possible. New York and Pennsylvania could therefore derive the data for particular hospitals and surgeons, and shut down some programs. Unfortunately, meaningful numbers are not as easy to derive in other situations.

The New York State Department of Health would like to do a study of angioplasty similar to their CABG study, but this is proving difficult to do, even though only 31 hospitals there perform angioplasty. The mortality rate is so low in angioplasty, less than 1%, that it is impossible to conclude anything. They would also like to study emergency bypass surgery and the adverse outcomes of acute MI, but cannot because of the administrative difficulty of getting the data, especially the creatine kinase (CK) concentrations and the electrocardiograms. We have advised New York to systematically collect CK data, and to have all electrocardiograms reviewed by a core laboratory, but this would cost money, which New York may not be in a position to fund.

It would be ideal to sit down with the patient or the family before a procedure and tell them exactly how many procedures you have done and exactly what the outcomes have been. That is the ideal doctor-patient relationship, with full disclosure and true data. Steve Ellis, MD, here at the Cleveland Clinic has refined an interventional scorecard with

a risk-adjusted model, which we hope will be used in other places. But I do not know who will pay for extracting these data, except insurance companies and malpractice attorneys, who would use these data in the wrong way. At one point I was extremely optimistic about scorecards. Now I am starting to wonder.

Missing data in, garbage out

DR. TOPOL: Risk adjustment is very difficult when data are missing. For example, in California, a scorecard for heart attacks is used statewide. The number of admissions and the mortality rate in acute MI are high enough to allow some comparison among hospitals. However, CK data and electrocardiograms are also needed, and these are missing. The fact that they do not even do a CK in patients with acute MI is very worrisome. But there is no way to adjust for that.

Here in Cleveland, a business coalition organized a program called Cleveland Health Quality Choice to find out where the best, most economical care is. Every quarter they rate the 37 hospitals in the Cleveland area for their risk-adjusted outcomes in five diagnoses: acute MI, heart failure, pneumonia, multiple surgeries, and stroke. (The business coalition does not pay for this. The Cleveland Clinic has spent $2.4 million over the last 2 years to extract these data.)

At one point, the Cleveland Clinic received poor ratings in heart failure, even though we have very sick patients who receive left ventricular assist devices and heart transplants, and most of the other hospitals did not even determine the patients' ejection fraction. With the garbage that goes into the model for risk adjustment, how can outcomes be compared in a meaningful fashion?

Some hospitals and physicians have very low numbers of patients, which can lead to erroneous conclusions. Then there are new doctors. Should their scorecards start right away from the first day, or should they have a grace period?

Shirking the sickest patients

DR. TOPOL: Most disturbing, the report card system penalizes the patients at highest risk, who need treatment the most. I am afraid that some physicians and institutions will shirk the sickest patients to make their scorecards look good. In fact, in 1992, the year that the Cleveland Health Quality Choice program began gathering data, other hospitals in the program increased their number of transfers to the Cleveland Clinic by 40% ($P < 10^{-8}$), while transfers from other hospitals not in the program stayed the same. The transferred patients had a mortality rate nearly three times as high as patients admitted here to begin with.[8]

IS PATIENT SATISFACTION IMPORTANT IN ASSESSING QUALITY?

DR. TOPOL: There are really three ways to assess quality of care: appropriateness, outcome, and patient satisfaction.

DR. VOGEL: We have not dealt with patient satisfaction in hard terms up to now. The health maintenance organizations are very interested in this, because if patients are satisfied they will continue to sign up year after year. However, we have no data as to how consumer satisfaction correlates with any harder index of the quality of medicine.

DR. TOPOL: So far the best variable that determines patient satisfaction is the length of time spent with the doctor or other health care provider. After that there are features such as the parking, whether coffee is provided, and whether there are good magazines in the waiting room. Of course the time spent by the physician could be inversely proportional to the quality of the care if a patient is being cared for by a very good physician who is very busy.

PHYSICIANS MUST TAKE THE LEAD

DR. VOGEL: Outcomes are clearly the bottom line, but even that measure could be subject to some manipulation. I could do incomplete revascularizations, never do a bilateral internal mammary graft, take lots of shortcuts, and come out with a good short-term mortality rate. That would not make me a good surgeon. I could underdilate lesions and probably come out with a very low myocardial infarction rate or a low complication rate. That would not make me a good interventionalist. Measuring the long-term outcomes would help, but it would also be more complicated and more expensive.

Nevertheless, the time has come to devise appropriate standards for medicine, and physicians must be the ones to do it. We do not regulate and police ourselves, but we complain that people who are not especially knowledgeable do not do it right. We

cannot have it both ways. In front-runner institutions like the Cleveland Clinic, we should not criticize the report-card process, but rather should try to define and validate the standards we intend to follow. In our profession, which is very costly and which is going to go through great upheaval in the next several years, physicians must take the lead. I can guarantee that if it is done by the government, the local chamber of commerce, or the local newspaper, it is going to be done very poorly.

REFERENCES

1. **Ellerbeck EF, Jencks SF, Radford MJ, et al.** Quality of care for Medicare patients with acute myocardial infarction. A four-state pilot study from the Cooperative Cardiovascular Project. JAMA 1992; **273:**1509–1514.
2. **Gunnar RM, Bourdillon PDV, Dixon DW, et al.** Guidelines for the early management of patients with acute myocardial infarction: a report of the American College of Cardiology/American Heart Association Task Force on Assessment of Diagnostic and Therapeutic Cardiovascular Procedures (Subcommittee to Develop Guidelines for the Early Management of Patients with Acute Myocardial Infarction). JACC 1990; **16:**249–292.
3. **Hilborne LH, Leape LL, Kahan JP, et al.** Percutaneous transluminal coronary angioplasty: a literature review and rating of appropriateness and necessity. Santa Monica, Calif: RAND; 1991. JRA-01.
4. **Vogel RA, Lauer MA, Ziskind AA.** A new revascularization appropriateness score for coronary angioplasty and bypass surgery. Clin Res 1994; **42:**290A. Abstract.
5. **Lauer MA, Ziskind AA, Lemmon CC, Vogel RA.** Prospective assessment of revascularization appropriateness scoring systems: a comparison of RAND expert panel ratings, ACC/AHA guidelines, and the University of Maryland revascularization appropriateness score. JACC 1995; **25:**344A. Abstract.
6. **Ziskind AA, Lauer MA, Lemmon CC, Vogel RA.** Prospective application of RAND expert panel ratings, ACC/AHA guidelines, and the University of Maryland revascularization appropriateness score (RAS): Only RAS predicts clinical outcome. JACC 1995; **25:**344A. Abstract.
7. **Hannan EL, Kilburn H Jr, Racz M, Shields E, Chassin MR.** Improving the outcomes of coronary artery bypass surgery in New York State. JAMA 1994; **271:**761–766.
8. **Clough JD, Kay R, Gombeski WR, Nickelson DE, Loop FD.** Mortality of patients transferred to a tertiary care hospital. Cleve Clin J Med 1993; **60:**449–454.

Correction

A table in the article "Lipid-regulating and antiatherosclerotic therapy: current options and future approaches" (*Cleveland Clinic Journal of Medicine* 1996; 63:31–41) contained an error. In *Table* 6 on page 37, the values for the effects of the various drugs on HDL-C and LDL-C were reversed through an editing error. The corrected table appears below.

TABLE 6
APPROVED DRUGS FOR DYSLIPIDEMIA*

Bile-acid sequestrants		
Lipid effects:†	LDL-C:	↓ 15%–30%
	HDL-C:	↑ 3%–5%
	TG:	↑ or no effect
Drugs and daily dose:	Cholestyramine	4–24 g
	Colestipol	5–30 g
HMG-CoA reductase inhibitors		
Lipid effects:	LDL-C:	↓ 20%–40%
	HDL-C:	↑ 5%–15%
	TG:	↓ 10%–20%
Drugs and daily dose:	Fluvastatin	20–40 mg
	Lovastatin	10–80 mg
	Pravastatin	10–40 mg
	Simvastatin	5–40 mg
Nicotinic acid (NA)		
Lipid effects:	LDL-C:	↓ 10%–25%
	HDL-C:	↑ 15%–35%
	TG:	↓ 20%–50%
Drugs and daily dose:	Crystalline NA	1.5–6 g
Fibric-acid derivatives‡		
Lipid effects:	LDL-C:	↓ 10%–15% (may↑)
	HDL-C:	↑ 10%–15%
	TG:	↓ 20%–50%
Drugs and daily dose:	Gemfibrozil	1200 mg
	Clofibrate	2000 mg
	Fenofibrate	300 mg

*Adapted from information in the second Adult Treatment Panel report, reference 24, and Yeshurun and Gotto, reference 25
†LDL-C, low-density lipoprotein cholesterol; HDL-C, high-density lipoprotein cholesterol; TG, triglyceride
‡Clofibrate is not considered a first-line agent because of associated toxicity; fenofibrate is approved but not currently available in the United States