

# Alternative Explanations for Poor Report Card Performance

**CONTEXT.** Many managed care organizations grade physician groups with “report cards” developed from administrative data sets and chart reviews.

**OBJECTIVE.** To investigate the accuracy of five report cards on a single group practice.

**DESIGN.** Determination of report card accuracy by using the practice capitation list and a review of the patients’ medical records.

**SETTING.** Academic practice in Philadelphia, Pennsylvania (19 physicians), evaluated with five report cards by two capitated health plans between 1994 and 1997.

**RESULTS.** Four major problems were uncovered. First, four of the five report cards included patients who were enrolled in our practice for only a portion of the reporting year (for the four report cards, the proportion of partial-year enrollees was 8%, 15%, 23%, and 100%). Second, there was a considerable number of false-positive diagnoses in the administrative algorithms. Eight of the 61 patients labeled with hypertension did not have this condition (error rate, 14%). Other error rates were 44% for coronary artery disease, 50% for congestive heart failure, 33% for atrial fibrillation, and 0% for diabetes. Third, the administrative data often failed to capture laboratory data. Laboratory performance measures for patients with diabetes (hemoglobin A<sub>1c</sub> and cholesterol measurement and screening for microalbuminuria) were 3 to 10 times higher when assessed by chart review. Finally, the uniformly small sample sizes used in the report cards make the estimates of performance imprecise. No report card reported 95% CIs.

**CONCLUSION.** Five report cards on a group practice contained methodologic problems that led to systematic underestimation of the practice’s performance. Larger surveys are needed to determine the accuracy of report cards in current use.

Most managed health care plans regularly send reports to participating primary care physicians. These reports gauge physicians’ hospital utilization and specialty referral patterns. In addition, many plans evaluate physicians’ clinical performance using standard HEDIS (Health Plan Employer Data and Information Set) measures and other indicators. These “report cards” present clinical opportunities. Physicians can learn how completely (or incompletely) they vaccinate their elderly patients or how often their patients with asthma are seen in emergency departments. By identifying areas for improvement, report cards can foster better care.

Report cards may also affect a physician’s reputation. In California, two health plans released physician-specific ratings from clinical audits and patient satisfaction surveys to members.<sup>1</sup> In addition, some plans publish their physician ratings on the Internet.<sup>2</sup> The Federation of State Medical Boards recommends that state medical boards consider using health plan report cards to help identify substandard physicians.<sup>3</sup> The reliability and accuracy of these data, however, are not well known. Poor-quality report cards may unnecessarily impugn good care.

*The abstract of this paper is available at [ecp.acponline.org](http://ecp.acponline.org).*

I describe one practice's experience with report cards. My colleagues and I were disturbed to learn about our own disappointing clinical performance. One report card stated that our physicians omitted blood pressure measurements for 36% of enrollees. Another stated that we administered anticoagulation therapy to only 12% of patients with atrial fibrillation. These reports prompted chart reviews. My colleagues and I wished to understand how we had failed to accomplish the intervention or, alternatively, why the report card failed to capture the care.

## Methods

### Setting

The practice is located in downtown Philadelphia, Pennsylvania. All 19 physicians in the practice are trained in internal medicine and are affiliated with Jefferson Medical College. At the time of the study, the practice had contracts with three capitated health plans, three preferred provider organizations, and one point-of-service plan. Patients from these plans constituted over half the practice in both volume and income and represented most of its growth.

Two of the capitated plans produced a total of five report cards between 1994 and 1997. As shown in **Table 1**, each of the five report cards used a 1-year study period and focused on either preventive care or disease management. The health plans used nurse auditors to gather data from outpatient records or gleaned it from administrative sources, including claims, pharmacy, and laboratory data. Results affected capitation rates; poor performers experienced a decrease in their monthly payments.

**TABLE 1**  
**Proportion of Patients Who Were Not Enrolled with the Practice for the Full 12-Month Report Card Period**

REPORT CARD TYPE	SURVEY PERIOD*	PROPORTION OF PATIENTS NOT ENROLLED FOR ENTIRE PERIOD (n/n)
Preventive	1/94–12/94	100% (42/42) <sup>†</sup>
Preventive	8/95–7/96	15% (4/26)
Cardiac	7/96–6/97	8% (7/89)
Preventive	7/96–6/97	0% (0/22)
Diabetes	10/96–9/97	23% (12/52)

\*Because these surveys came from different companies, dates may overlap.

<sup>†</sup>The practice joined the managed care organization in March; therefore, no patient was enrolled for all of 1994.

### Determination of Accuracy

The two plans provided patient names for each report card. I reviewed both the patient's enrollment status (to address the question, "Were we responsible for this patient at the time of the report card?") and the outpatient chart (to evaluate the performance measure itself). The discrepancies between the report card and our records were noted and catalogued by using the following categories.

#### Partial-Year Enrollment

The practice receives a monthly capitation list from the plans that notes the date on which patients selected their primary provider. All patients who were included in the report card and were enrolled in the practice for only part of the 12-month survey period were classified as partial-year enrollees.

#### Diagnostic Misclassification

The two disease management report cards relied solely on administrative data to establish the diagnosis. The diagnosis for each patient included in the report card was reevaluated in the medical record. The entire outpatient record was reviewed to evaluate five diagnoses.

A diagnosis of hypertension was considered erroneous if the chart lacked three elevated blood pressures, the problem list and notes made no mention of past hypertension, and the patient was not receiving pharmacotherapy for hypertension. A diagnosis of coronary artery disease was considered erroneous if the patient had normal coronary arteries at cardiac catheterization or a normal noninvasive cardiac workup or if a cardiologist decided that the patient's symptoms were non-cardiac and discontinued treatment. A diagnosis of congestive heart failure was considered erroneous if the ejection fraction exceeded 50% and there was no evidence that the physician considered congestive heart failure as an explanation for the patient's symptoms. A diagnosis of atrial fibrillation was considered erroneous when no past or present evidence of atrial fibrillation was recorded in the chart. A diagnosis of diabetes was considered erroneous if all random blood sugar levels and hemoglobin A<sub>1c</sub> levels were normal.

#### Missing Laboratory Data

The diabetes report card used three laboratory tests as performance measures (hemoglobin A<sub>1c</sub> and cholesterol measurement and screening for microalbuminuria). A test was recorded as being present when a laboratory result obtained during the 12 months of the study period was in the chart. These data were compared with the report card results.

## Precision

95% CIs were calculated on the proportions in one report card measuring preventive services. The following equation was used<sup>4</sup>:

$$95\% \text{ CI} = p \pm 1.96\sqrt{pq/n}$$

where  $n$  is the total number of patients in the sample,  $p$  is the proportion of the sample meeting the performance measure, and  $q = 1 - p$ .

## Results

### Partial-Year Enrollment

Four of the five report cards included partial-year enrollees in the patient sample. Partial-year enrollees are a particular concern in the performance measurement of annually recommend services (e.g., flu shots, fecal occult blood screening, and ophthalmologic evaluation in diabetic patients). It is misleading to hold a practice responsible for an annual flu shot, for example, in a patient who has been enrolled for only 2 months. As shown in **Table 1**, the proportion of sampled patients who were not enrolled for the full 12 months of the survey period varied from 0% to 100%.

### Diagnostic Misclassification

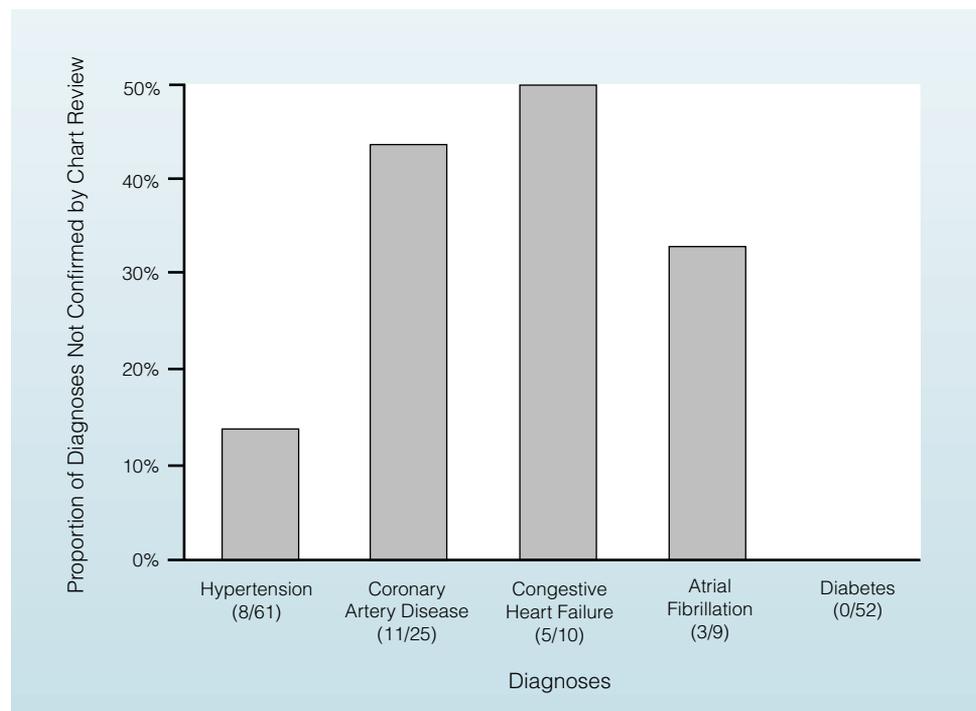
The chart could not confirm many of the administrative diagnoses for patients in the cardiac care report card. As shown in **Figure 1**, the accuracy of the administrative

diagnoses varied depending on the specific diagnosis. The rate of erroneous diagnoses ranged from 50% (5 of 10 patients) in congestive heart failure to 0% (0 of 52 patients) in diabetes (although 3 of the 52 patients developed diabetes during the survey year). **Table 2** shows the results of the chart audit for patients in whom the administrative algorithm misdiagnosed coronary artery disease, hypertension, and atrial fibrillation.

**Figure 2** explores the relevant chart findings for the reported anticoagulation rate of 12% in patients with atrial fibrillation. Three of the nine sampled patients had no chart evidence of atrial fibrillation (33% diagnostic error rate) (**Figure 1**). Physicians prescribed warfarin for two of the remaining six patients (anticoagulation rate, 33%). The chart audit clarifies why the other four patients did not receive anticoagulation: One had strong contraindications, one refused, one received salicylates, and one had never been seen by the practice (although that patient's name was on the capitation list).

### Missing Data

Some report cards rely on laboratory data obtained from administrative sources. As shown in **Table 3**, the proportion of diabetic patients who met laboratory performance measures was 3- to 10-fold higher on chart audit than in the administrative database. When partial-year enrollees were eliminated, the practice performance on diabetic indices improved further.



**FIGURE 1.** Proportion of report card diagnoses that were not confirmed by chart review.

**TABLE 2****Chart Audit Results in 22 Patients with Misdiagnosed Conditions****CORONARY ARTERY DISEASE (n = 11)**

87-year-old man with dementia and no cardiac history or symptoms.

72-year-old woman with hypertension, retinal artery occlusion, kidney stone, no clear coronary artery disease, and questionable angina. Therapy with nitrites was discontinued by the cardiologist.

66-year-old man with hypertension.

40-year-old woman with diabetes and hypertension.

61-year-old woman with noncardiac chest pain and negative results on exercise thallium test; receiving treatment for hypercholesterolemia as primary prevention.

54-year-old woman with no cardiac history.

36-year-old man with left brachial plexopathy.

50-year-old woman with sarcoidosis. Myocardial infarction was ruled out in November 1996 and March 1997. In March 1997, exercise thallium test results were negative (87% predicted heart rate).

53-year-old woman with chronic obstructive pulmonary disease, pulmonary hypertension, rheumatic fever, mild hypertension, transitional cell cancer, and normal cardiac catheterization in 1994.

67-year-old woman; smoker.

52-year-old woman with asthma who had had several surgical procedures; negative findings on cardiac catheterization in March 1996.

**HYPERTENSION (n = 8)**

48-year-old woman with breast cancer, mitral valve prolapse, elevated cholesterol level, and no hypertension on several measurements.

53-year-old woman with two normal blood pressure measurements in 1996, one measurement of 150/100 mm Hg, then four subsequent normal measurements.

40-year-old woman with sarcoidosis; receiving propranolol for migraine headaches.

39-year-old healthy woman; blood pressure readings of 135/80 mm Hg and 102/68 mm Hg.

64-year-old woman with depression, history of breast cancer, and bladder problems.

44-year-old man with bone marrow transplant for chronic lymphocytic leukemia, sleep apnea, and pulmonary hypertension.

43-year-old woman with obstructive sleep apnea, hypothyroidism, and negative stress test results for angina.

65-year-old man with osteoarthritis, bilateral prostatic hypertrophy, and elevated lipid levels.

**ATRIAL FIBRILLATION (n = 3)**

63-year-old man with chronic obstructive pulmonary disease; had prior coronary bypass surgery with perioperative paroxysmal atrial tachycardia but no atrial fibrillation.

35-year-old man with hypertension and gastroesophageal reflux disease.

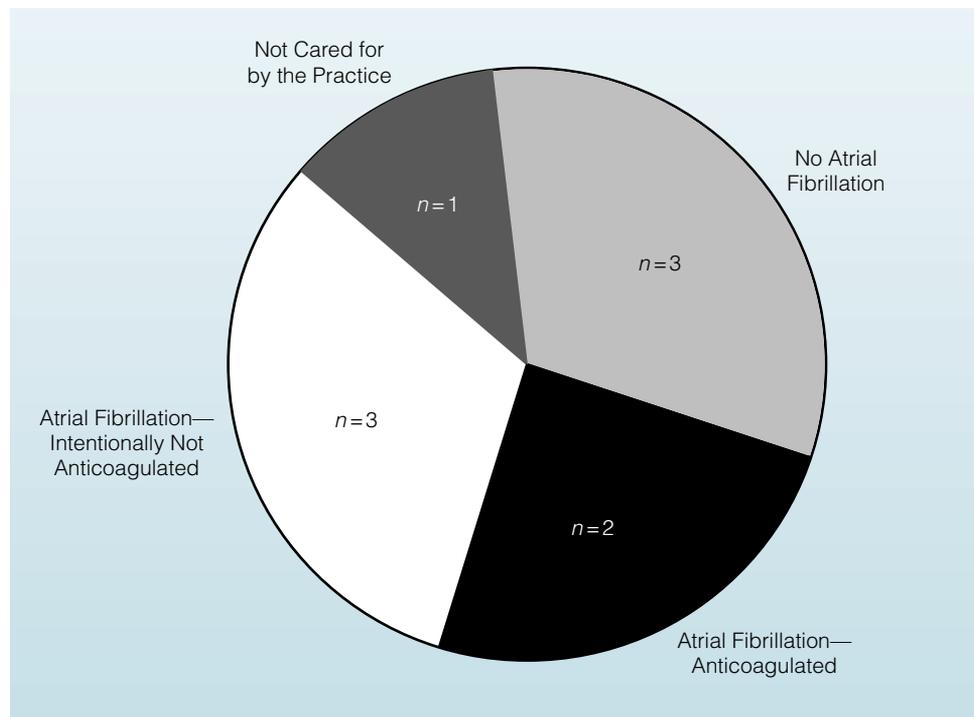
48-year-old man with coronary artery disease and no atrial fibrillation.

**Precision**

The report cards were based on sample sizes ranging from 22 to 89 patients. These samples were often further subdivided by age (preventive samples) or disease (cardiac sample). Thus, the CIs surrounding the proportions were often wide, limiting the reliability of any single quality measurement. **Table 4** provides one example of this problem, using routine preventive performance measures for enrollees 40 to 64 years of age. None of the report cards reported 95% CIs for any measurement.

**Discussion**

The National Committee for Quality Assurance judges health plan performance on the basis of data from members who have been continuously enrolled in the plan for 12 months.<sup>5</sup> However, in this study, four of five report cards judged a practice's performance without such exclusions. Partial-year enrollees confound report card results by measuring annual indices in patients whom a practice has treated for less than 1 year.



**FIGURE 2.** Chart audit findings in nine patients reported to have had atrial fibrillation in the cardiac report card. The anticoagulation rate was reported as 12% (which cannot be calculated from the data).

Our own experience is illustrative. In 1994, a report card placed our practice in the lowest 8% of physicians for quality. The plan reported that the practice neglected all clinical indicators in over one third of sampled patients; as a result, our capitation rate plummeted. The chart audit, however, revealed most of these patients had never visited or called the office, probably because they had been enrolled in the practice for only a few months. None had been enrolled for a full year. After the chart audit, the plan adjusted the quality rating; they agreed that the partial-year enrollees skewed the report card results. Because plan members change physicians, health plans should include only enrollees who have been capitated to the practice for a full year when measuring a practice's performance.

This study also shows how other methodologic problems can erroneously indicate poor performance.

The cardiac report card stated that the practice prescribed  $\beta$ -blockers in only 33% of patients with coronary artery disease. However, the chart audit found that the diagnosis of coronary artery disease was erroneous in 44% of patients, perhaps because ICD-9 (International Classification of Diseases, ninth revision) codes lack a designation for the "rule-out" diagnosis.<sup>6</sup> Eliminating the 11 patients with a misdiagnosed disease leaves 14 patients with coronary artery disease and a  $\beta$ -blocker prescription rate of 50%. The algorithm's diagnostic errors obscure the practice's real behavior and created a sample that was too small for meaningful comparisons.<sup>7</sup>

The accuracy of administrative algorithms for determining diagnoses may depend on the specific diagnosis. In hospitalized patients, claims data missed 52% of patients with congestive heart failure and 32% of

**TABLE 3**  
**The Effect of Measurement Method on Diabetes Care Performance**

DIABETES PERFORMANCE MEASURE	REPORT CARD RESULTS (n = 52)	CHART AUDIT	
		ALL PATIENTS (n = 52)	PATIENTS ENROLLED FOR ENTIRE PERIOD (n = 40)*
Hemoglobin A <sub>1c</sub> measurement (biannually)	5.8%	51%	67%
Cholesterol measurement (annually)	11.6%	50%	65%
Screening for microalbuminuria (annually)	7.8%	25%	32%

\*Twelve partial-year enrollees were excluded from the analysis.

TABLE 4

**95% CIs for Three Variables in a Report Card\***

VARIABLE	FINDINGS (95% CI)
Blood pressure recorded	64% (48%–80%)
Screened for alcohol and substance abuse	42% (26%–58%)
Cholesterol measured	61% (45%–77%)

\*Measured in enrollees 40 to 64 years of age (n = 36).

patients with hypertension,<sup>8</sup> although clinical and claims data correlated strongly for the diagnosis of diabetes.<sup>9</sup> In addition, Medicare claims data and chart audits concur on the proportion of diabetic patients in whom hemoglobin A<sub>1c</sub> levels have been measured.<sup>10</sup> However, in this study, the chart audit diverged dramatically from administrative data in identifying laboratory information on diabetic patients. This is explained by the fact that most of our patients are referred to a laboratory that is not connected to the health plan's database.

This case report, although clearly limited in scope, has a simple message for physicians: Carefully scrutinize your report cards. These reports increasingly influence physician compensation, and public dissemination of report cards may unfairly tarnish the reputations of physicians. This study also provides evidence that plans need to field-test report cards to determine their accuracy. Administrative algorithms and computerized databases need to be validated with large-scale chart audits. Sample sizes should be increased to measure physician behavior more precisely.<sup>7, 11, 12</sup> At a minimum, all report cards should report CIs.

The larger untested question is whether report cards will improve care. Although my practice achieved better scores than the report cards indicated (and often better than the plan mean), the chart reviews also confirmed genuine deficiencies. We are making efforts to improve care of patients with diabetes and are developing plans to check all patients with congestive heart failure for use of angiotensin-converting enzyme inhibitors,  $\beta$ -blockers, and spironolactone. Report cards, despite their inaccuracies, can provide guidance for clinical improvement.

Even if report cards were perfectly accurate, however, it remains unclear what portion of ideal performance is under physician control.<sup>7</sup> Business leaders assume that report cards, even if they are flawed, will enhance care.<sup>1</sup> Others warn that this places too much confidence in report cards.<sup>1, 8, 11–13</sup> Although report cards undoubtedly have the potential to enhance the quality of medical care, whether they in fact do so depends in large part on their accuracy.

## Take-Home Points

- Report cards for one practice suggested disturbing failures in clinical performance.
- Chart audits were performed to confirm the results, and the enrollment status of sampled patients was checked against capitation lists.
- Four of the five report cards included patients who were enrolled in the practice for only a portion of the reporting year.
- Other problems included diagnostic errors, missing laboratory data, and inadequate sample size.
- Although report cards may help improve medical practice, their accuracy is open to question.

### References

1. Dalzell MD. Health care report cards: are you paying attention? *Manag Care*. 1999;8:27-8, 30-2, 34.
2. <http://www.healthpartners.com/ccs.html>.
3. Prager LO. Boards want a broader quality monitoring role. *American Medical News*. 1998; 24/31:1, 46.
4. Kuzma JW. *Basic Statistics for the Health Sciences*. Palo Alto, CA: Mayfield Publishing; 1984:137-8.
5. Thompson JW, Bost J, Ahmed F, Ingalls CE, Sennett C. The NCQA's quality compass: evaluating managed care in the United States. *Health Aff (Millwood)*. 1998;17:152-8.
6. Dans P. Caveat doctor: how to analyze claims base report cards. *Journal on Quality Improvement*. 1998;24:21-30.
7. Hofer TP, Hayward RA, Greenfield S, Wagner EH, Kaplan SH, Manning WG. The unreliability of individual physicians "report cards" for assessing the costs and quality of care of a chronic disease. *JAMA*. 1999;281:2098-105.
8. Dans PE. Looking for answers in all the wrong places [Editorial]. *Ann Intern Med*. 1993;119:855-7.
9. Jollis JG, Ancukiewicz M, DeLong ER, Pryor DB, Muhlbaier LH, Mask DB. Discordance of databases designed for claims payment versus clinical information systems. Implications for outcomes research. *Ann Intern Med*. 1993;119:844-50.
10. Weiner JP, Parente ST, Garnick DW, Fowles J, Lawthers AG, Palmer RH. Variations in office-based quality. A claims-based profile of care provided to Medicare patients with diabetes. *JAMA*. 1995;273:1503-8.
11. Bindman A. Can physician profiles be trusted? [Editorial] *JAMA*. 1999;281:2142-3.
12. Localio AR, Hamory BH. A report card for report cards [Editorial]. *Ann Intern Med*. 1995;123:802-3.
13. Kassirer J. The use and abuse of practice profiles [Editorial]. *N Engl J Med*. 1994;330:634-6.

### Acknowledgments

The author thanks Kenneth Epstein, MD, Christine Laine, MD, and Barbara Turner, MD, for comments on earlier versions of this manuscript.

### Correspondence

Rachel Sorokin, MD, 12 South Ninth Street, Suite 502, Jefferson Medical College, Philadelphia, PA 19107; telephone: 215-955-0733; fax: 215-923-9239; e-mail: rachel.sorokin@mail.tju.edu.