

# Comparing Physician-Specific Two-Year Patient Outcomes After Coronary Angiography

## Methodologic Issues and Results

Stephen G. Ellis, MD, FACC,\* Dave Miller, MS,§ Thomas F. Keys, MD,†‡ Kimberly Brown, RN,\* Renee Ellert, RN,\* Georgiana Howell,\* A. Michael Lincoff, MD, FACC,\* Eric J. Topol, MD, FACC\*  
*Cleveland, Ohio, and San Francisco, California*

- OBJECTIVES** We sought to evaluate methodologies to compare physician-related long-term patient outcomes appropriately.
- BACKGROUND** Evaluation of physicians on the basis of short-term patient outcome is becoming widely practiced. These analyses fail to consider the importance of long-term outcome, and methods appropriate to such an analysis are poorly defined.
- METHODS** All patients undergoing coronary angiography between 1992 and 1994 who received all of their cardiac care at our institution were followed for  $27 \pm 13$  months (mean  $\pm$  SD). Patients ( $n = 754$ ) were cared for by one or more of 17 staff physicians. Risk-adjusted models were developed for four candidate clinical end points and cost. Physicians were then evaluated for each outcome measure.
- RESULTS** Of the clinical end points, death could be modeled most accurately ( $c$ -statistic = 0.83). The  $c$ -statistics for other end points ranged from 0.63 to 0.70. Physicians with outcomes statistically different ( $p < 0.05$ ) from other physicians were identified more commonly than would be expected from the play of chance ( $p = 0.005$ ). However, improvement in the  $c$ -statistics by the addition of physician identifiers was very modest. Physician's evaluations by the four measures of clinical outcome were variably correlated ( $r = .00$  to  $.85$ ). Graphic display of clinical and cost results for each physician did identify certain physicians who might be judged to provide more cost-effective care than others.
- CONCLUSIONS** Although comparisons of groups of physicians on the basis of long-term patient outcomes may have merit, individual physician-to-physician comparisons will be more difficult, owing to 1) multiple physicians contributing care to individual patients; 2) the poor predictive capacity of models other than that for survival; and 3) the modest apparent impact of differences in physician providers on long-term patient outcome. With these caveats in mind, modeling to compare patient outcomes of individual physicians with homogeneous patient populations or to identify gross outliers (good or bad) may be practicable in some patient-care systems, but may be inappropriate in others. (J Am Coll Cardiol 1999;33:1278-85) © 1999 by the American College of Cardiology

Although controversial in both concept and execution, evaluation in public forums of high-profile medical specialists such as cardiac surgeons and invasive cardiologists is currently being practiced by comparison of their physician-specific patient outcomes (1). Although debated, a public health benefit of such analyses has been claimed by the State of New York because short-term mortality associated with

CABG (coronary artery bypass graft) surgery appears to have decreased since public dissemination of results began in 1991 (2). Present analyses focus on short-term results, and might penalize, for example, physicians who accept high-risk patients for coronary revascularization. Methodologies appropriate to compare long-term outcomes remain poorly developed.

We analyzed several methods of assessing physician-specific patient outcomes (adverse clinical events and cost) in a well-characterized cohort of 754 patients receiving all their cardiac care in our medical system over a period of  $27 \pm 13$  months.

## METHODS

**Patient population.** All patients undergoing coronary angiography from January 1, 1992, through December 31,

From the Departments of \*Cardiology, †Biostatistics and Epidemiology, and ‡Office of Quality Management, the Cleveland Clinic Foundation, Cleveland, Ohio, and §Ischemia Research and Education Foundation, San Francisco, California. Funding for this project was provided indirectly by Nycomed, Princeton, New Jersey, and Mallinkrodt, Chesterfield, Missouri, which supported the Cleveland Clinic Foundation Interventional Registry during this time period, and by internal institutional funds.

Manuscript received June 22, 1998; revised manuscript received November 25, 1998, accepted January 5, 1999.

#### Abbreviations and Acronyms

CABG	= coronary artery bypass graft (surgery)
CAD	= coronary artery disease
CHF	= congestive heart failure
CK	= creatine kinase
COPD	= chronic obstructive pulmonary disease
MI	= myocardial infarction
PCI	= percutaneous coronary intervention
PVOD	= peripheral vascular obstructive disease
ROC	= receiver operating characteristics
TIA	= transient ischemic (cerebrovascular) attack

1994, who responded to query that they expected to or were receiving all cardiac care at our institution, and who consented to participate in a study tracking cardiac cost and clinical outcome over time (3), were included in this study (most patients undergoing catheterization at our institution receive their primary care elsewhere). Baseline, treatment, in-hospital clinical outcome and all cost data were recorded prospectively in several hospital databases that were merged for the purposes of this and the above-mentioned cost study (3). Long-term clinical follow-up was obtained via chart review and phone contact, as described previously (3). If the patient was alive and exited our care system, patient data were censored at the time of exit. Patients having had or awaiting cardiac transplantation were excluded.

**Responsible physician(s).** In our clinic system, patients were typically assigned to one primary cardiologist, who then, as necessary, referred them to another cardiologist for specific procedures. Physicians included in this analysis would be categorized as generalist cardiologists who also performed cardiac catheterization, some of whom were also specialists in percutaneous coronary intervention (PCI), intensive care unit cardiology, or treatment of patients with congestive heart failure (CHF) (not mutually exclusive). Over time, however, the patient's primary cardiologist might change. When the primary cardiologist at the time of the catheterization and the current cardiologist were the same physician, they were assigned responsibility for the patient's outcome. When the physicians were different, responsibility was assigned two ways: 1) the initial cardiologist alone (primary cardiologist), 2) the initial and subsequent cardiologist shared responsibility equally (co-primary cardiologist). Both methods of assignment were evaluated as possible methods for optimal "scorecarding."

**Method of determining cost.** In-hospital and outpatient cardiac and cardiac-related (i.e., those related to complications of cardiac procedures) costs (laboratory tests, medications, physician visits) were obtained using a commercially available cost-accounting system (Transition System, Boston, Massachusetts) and previously described methodology (3,4).

**Variables and definitions.** Modeling of clinical outcome and cost was performed using the following baseline variables:

*Angiographic:* jeopardy score (an estimate of the amount of left ventricular myocardium at risk; scored 0-6) (5); left ventricular systolic function (0 = normal, 1 = mild dysfunction [EF = 45 to 55%], 2 = moderate dysfunction [EF = 30 to 44%] and 3 = severe dysfunction [EF < 30%] by ventriculography or echocardiography); number of diseased vessels ( $\geq 50\%$  diameter stenosis of a bypassable vessel from one or more of three major vascular territories).

*Clinical:* age, aortic valve disease (valve area  $< 1.2 \text{ cm}^2$ ,  $\geq 2+$  regurgitation, or equivalent), Canadian Cardiovascular Association angina class, cardiogenic shock, catheterization status (elective, urgent [requiring catheterization within 24 h of referral], or emergency [requiring immediate catheterization]), COPD (chronic obstructive pulmonary disease) requiring treatment, creatinine  $\geq 2 \text{ mg/dl}$ , current smoking, gender, history of ventricular tachycardia or fibrillation, height, hospital transfer, hypertension (systolic  $\geq 180 \text{ mm Hg}$ , diastolic  $\geq 100 \text{ mm Hg}$  or treated), insulin-requiring diabetes, mitral valve disease (valve area  $< 2.0 \text{ cm}^2$ ,  $\geq 2+$  regurgitation, or equivalent), New York Heart Association CHF class, non-insulin-requiring diabetes, number of major comorbidities (COPD, creatinine  $\geq 2 \text{ mg\%}$ , stroke or TIA [transient ischemic attack], PVOD [peripheral vascular obstructive disease]), pacemaker, positive stress test, prior CABG, prior MI (myocardial infarction), prior PCI, prior stroke or transient ischemic attack (TIA), race, recent MI (within 2 weeks), renal insufficiency requiring dialysis, stable angina, symptomatic PVOD (peripheral vascular obstructive disease), unstable angina, and weight.

**Statistical analyses.** All data, except for cost data, which are presented as median and interquartile range owing to their non-Gaussian distribution, are presented as mean  $\pm$  1 SD. Survival analyses were performed using Kaplan-Meier and Cox analyses. Models were developed using 35 potential clinical and anatomic covariates to predict the five major candidate end points of this study: survival; MI-free survival; event (death, MI, CABG, PCI)-free survival; severe angina and event-free survival; and cost. Creatinine kinase (CK) levels were obtained routinely at 8 and 24 h after PCI, and in the event of suspected ischemia. Myocardial infarction was defined as development of new pathologic Q waves on electrocardiogram (ECG) or CK  $> 2 \times$  normal (with elevated MB fraction) under most circumstances,  $\geq 3 \times$  after PCI, and  $\geq 5 \times$  after CABG. The accuracy for each model in predicting outcome was assessed using regression analysis and receiver operating characteristic modified for survival

analysis as described by Harrell et al. (6). Models were internally validated 10 times using a randomly selected 80% sample for fitting and the remaining 20% of the sample for testing in order to determine the extent to which the predictive accuracy of the models was overly optimistic (7). Models were internally calibrated by comparing actual to predicted outcomes for patient-risk quintiles based upon predicted outcome.

To these models were then added physician-specific identifiers to ascertain the effect of each physician(s) caring for the patient on the various outcomes. All physicians with 20 or more patients in this cohort were evaluated. Two separate analyses were performed—one using the cardiologist caring for the patient at the time of the catheterization, and the other linking both that physician and the primary cardiologist during follow-up to the patient and their outcome. Physicians were then ranked by their beta coefficient of the variable reflecting provider care. Physicians with risk-adjusted patient outcomes significantly better or worse ( $p \leq 0.05$ ) than other physicians were considered "outliers." Physician coefficients for each of the five outcomes and two methods of assigning outcome responsibility were then compared using standard regression techniques.

In addition, graphical depiction of the relation between providers' clinical and cost results was presented, using their individual correlation coefficient and its standard deviation for each parameter to develop plots in a two-dimensional grid.

To assess the subjectivity and potential for manipulation of data (8) of several potential end points and covariates, 11 cardiologists, epidemiologists, or statisticians familiar with management of patients with coronary artery disease were asked to grade each parameter on a 1-5 scale (1 = highly objective; 5 = highly subjective). These data were used to place the objectivity of parameters involved in "scorecarding" into perspective, but not for the modeling process per se.

## RESULTS

**Long-term follow-up.** The average time of clinical follow-up in survivors was  $27 \pm 13$  months. Ten patients (1.3%) were lost to follow-up and 69 patients (8.8%) left the clinic system for another medical care group at  $23 \pm 10$  months. Patients leaving the clinic system less often had baseline creatinine  $\geq 2.0$  mg% (1.6% vs. 6.9%,  $p = 0.007$ ) and CHF (4.3% vs. 10.6%,  $p = 0.02$ ), but were otherwise similar to patients remaining in the system. Cost and clinical outcome data through 12 months, when most patients who eventually left were still in the system, were nearly identical for those who eventually left and those who did not.

**Baseline characteristics, initial treatment and outcome.**  
**MEDICAL THERAPY.** After catheterization, 337 patients were referred for medical therapy (see Table 1). Of these, 243 had significant CAD. Their survival at 30 days, 1 year,

and 2 years was  $98.0 \pm 0.9\%$ ,  $94.0 \pm 1.6\%$ , and  $88.1 \pm 2.5\%$ , respectively. Their event-free survival at these same time milestones was  $97.1 \pm 1.1\%$ ,  $85.3 \pm 2.3\%$ , and  $71.2 \pm 3.5\%$ , respectively. After two years, 82% had no angina, 14% had class I-II angina and 4% had class III-IV angina. In addition, 5 patients had primary ventricular arrhythmias, 4 patients had primary valvular heart disease, 4 patients had coronary spasm and 81 patients had suspected but no significant CAD. The overall median two-year per patient cardiac cost of care for patients initially referred for medical therapy alone was \$9,386 (\$14,411 for those with demonstrated CAD).

**PERCUTANEOUS INTERVENTION.** Two hundred thirty-two patients were referred for a PCI as their initial treatment after the index catheterization (see Table 1). The primary device used was balloon angioplasty in 68.6%, stent in 12.3%, directional atherectomy in 10.6%, rotational atherectomy in 6.8% and excimer laser in 1.7%. Technical success without death, infarction or need for CABG was obtained in 94.5% of patients. Overall survival and event-free survival at 30 days, 1 year, and 2 years was  $98.7 \pm 0.7\%$ ,  $97.8 \pm 1.0\%$ ,  $95.1 \pm 1.6\%$ , and  $91.8 \pm 1.8\%$ ,  $65.2 \pm 3.2\%$ ,  $60.4 \pm 3.4\%$ , respectively. After two years, 86% had no angina, 12% had class I-II angina and 1% had class III-IV angina. Their cumulative cost of cardiac care at 2 years was \$23,514.

**BYPASS SURGERY.** One hundred eighty-five patients had CABG as their initial therapy after catheterization (see Table 1). Seventy-one percent received one or more arterial conduits and 10% had concomitant valve surgery. Their survival and event-free survival at 30 days, 1 year, and 2 years was  $97.3 \pm 1.2\%$ ,  $93.3 \pm 1.9\%$ ,  $91.9 \pm 2.1\%$  and  $91.3 \pm 2.1\%$ ,  $83.3 \pm 2.8\%$ ,  $81.2 \pm 3.0\%$  respectively. After 2 years, 88% had no angina, 9% had class I-II angina and 3% had class III-IV angina. Their cumulative cost of cardiac care at 2 years was \$35,544.

**Physician providers.** Of the 17 physician providers, all performed cardiac catheterization, and 9, 3, and 3 also subspecialized in PCI, CHF, and intensive care cardiology, respectively. The mean number of patients cared for in a primary fashion for this group was  $45 \pm 25$  (median = 47).

**Reliability of variable coding.** The results of the physician panel assessment of the subjectivity or susceptibility of selected candidate covariates and end points to manipulation are shown in Figure 1. Some variables such as death and CABG were judged to be objective and reliably determined. Conversely, other variables such as CHF class, catheterization status (elective/urgent/emergency) and unstable angina were considered to be rather unreliable owing to either subjectivity or susceptibility to manipulation.

**Modeling candidate end points.** The variables correlated with selected candidate end points, their contribution to potential models, predictive accuracy and calibration of the

**Table 1.** Baseline Patient Characteristics (by Treatment Group)

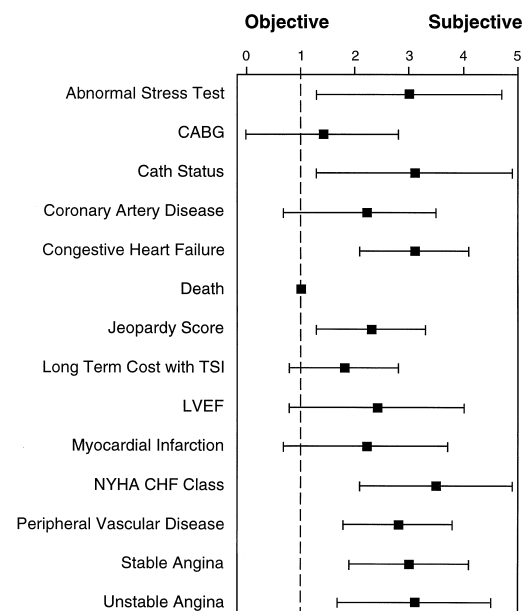
	Study Cohort				All Patients Undergoing Catheterization 1992-94
	Medicine*	PCI	CABG	Total	
Number	337	232	185	754	14,622
Age (yrs)	65 ± 11	64 ± 10	66 ± 10	65 ± 10	62 ± 11
Angina class					
I-II (%)	51	48	42	48	NA*
III-IV (%)	26	40	38	33	NA
Catheterization status					
Emergency (%)	2.0	10.7	7.0	5.8	NA
Urgent (%)	29.9	35.9	48.1	35.8	NA
Current smoker (%)	16	17	21	17	22
Diabetes (%)	28	26	33	29	29
Hypertension (%)	67	60	71	66	62
Male (%)	67	77	72	71	70
Major comorbidities					
COPD (%)	5	7	7	6	6
Creatinine ≥2 (%)	8	4	2	6	5
CVA/TIA (%)	12	8	14	11	10
PVOD (%)	18	17	19	18	19
Prior CABG (%)	31	25	20	27	25
Prior MI (%)	45	40	42	43	43
Prior VT/VF (%)	7	4	3	5	NA
LVEF < 40% (%)	13	12	11	12	15
Multivessel disease (%)	29	37	89	45	52
Jeopardy score (0-6)	1.3 ± 1.5	2.3 ± 1.3	3.8 ± 1.4	2.2 ± 1.7	NA
Unstable angina (%)	44	49	49	47	NA
Hospital transfer (%)	9	17	17	13	NA

\*During the time period studied, prospective data accrual for all patients undergoing catheterization was not performed, but rather selected data variables were captured by a trained group of individuals after chart review.

CABG = coronary artery bypass graft (surgery); COPD = chronic obstructive pulmonary disease; CVA = cerebral vascular accident (stroke); PVOD = peripheral vascular obstructive disease; TIA = transient (cerebral) ischemic attack; VT/VF = ventricular tachycardia or fibrillation.

models are shown in Table 2 and 3 and Figure 2. Of the clinical models, that for overall survival was considerably most predictive (c-statistic or area under ROC [receiver operating characteristic] curve, 0.83). C-statistics for the models for MI-free survival and angina/event-free survival were 0.63 and 0.70, respectively (data not shown). Considering that only variables available immediately after catheterization were used for these predictive models, the model for log<sub>e</sub> cost was also quite predictive (r = 0.57).

**Identifying physician outliers.** Acknowledging that our power to detect significant (p < 0.05) outliers was low because of relatively few patients for each provider, the frequency of identifications of outliers (n = 34 possible, counting primary and co-primary status separately) for various models was as follows: survival, 5.9%; MI-free survival, 20.6%; event-free survival, 5.9%; angina and event-free survival, 2.9%; cost, 17.6%. Thus, the likelihood that the identification of outliers for all outcomes combined was purely by chance was very low [expected = 34 physicians per each of 5 models/one divided by mean p value for each outlier (0.029) = 170/34 = 5.0; observed = 18, χ<sup>2</sup> = 7.88, p = 0.005].



**Figure 1.** Subjectivity of selected candidate covariates and end points. Point estimates and 95% confidence limits for selected variables (1 = most objective; 5 = most subjective) are shown. CABG in this figure refers to CABG as a complication of PCI.



**Table 2.** Independent Baseline Characteristics Correlated With Survival

Covariate	Estimate	SD	T-stat	p
Dialysis	3.30	0.59	5.54	< .001
Age/10 yrs	0.94	0.20	4.74	< .001
Diabetes requiring insulin	1.44	0.32	4.55	< .001
Left ventricular function	0.13	0.03	3.55	< .001
Caucasian race	-1.11	0.38	-2.94	.003
NYHA CHF class	0.39/unit	0.14	2.81	.005
Prior CABG	0.70	0.28	2.50	.013
Urgent catheterization	-0.66	0.30	-2.18	.029
Congestive heart failure	0.76	0.35	2.14	.032

Final log-likelihood = -280.29, p < 0.0001. ROC c-statistic = 0.829. CABG = coronary artery bypass graft (surgery); CHF = congestive heart failure; NYHA = New York Heart Association.

**Correlation of physician evaluation between different models.** Pearson linear regression coefficients (r) showing the relationship of the physicians' evaluations between each of the models are provided in Table 4 (see methodology).

Assessment of physician contribution, as primary or co-primary, by linear regression between the two revealed very good agreement (r > .90) for all models except event-free survival (r = .86) and angina and event-free survival (r = .63). Evaluation by models characterizing largely overlapping end points (e.g., event-free survival and angina and event-free survival) correlated well (r = .85), whereas the correlation between most other clinical models was weak (r ≤ .40). Physician evaluation on the basis of cost was reasonably well correlated with that for survival (r = .63 to .66), and modestly well with event-free and angina and event-free survival (r = .33 to .54), but poorly with the results of the models of MI-free survival (r = .00 to .17).

**Influence of physician modeling on models' correlation with outcome.** In general, the increment in correlation gained by adding physician identifiers to the clinical variables was very modest. Co-primary physician identifiers were somewhat more helpful than were primary physician identifiers. For example, for the end point of long-term survival, inclusion of primary physician identifiers resulted in a c-statistic of 0.831 (baseline without physician identifiers, c = 0.829), whereas co-primary physician identifier inclusion yielded a c-statistic of 0.833. The c-statistic for physician identifiers alone was 0.648. For the end point of event-free survival addition of the primary physician identifier increased the c-statistic from 0.672 to 0.685. For 2-year log cost, the inclusion of primary physician identifiers

resulted in a variance (r<sup>2</sup> = .332) (baseline r<sup>2</sup> = .326), whereas co-primary physician identifier inclusion yielded a r<sup>2</sup> = .342. The variance accounted for by the physician identifiers alone was 0.063.

**Model bias favoring different subspecialty groups.** One potential bias was identified when codes for subspecialty groups were entered into each of the five models—for MI-free survival, intensive care physicians had worse outcomes (p = .005). With no other model could a bias for or against any physician group be identified.

**"Global" evaluation of individual physician performance.** Given the superior performance of our survival and cost models, one may succinctly characterize each individual physician's patient outcomes by bipolar graphing the physician's coefficient and its 95% confidence limits for these two models (for examples, see Fig. 3).

**DISCUSSION**

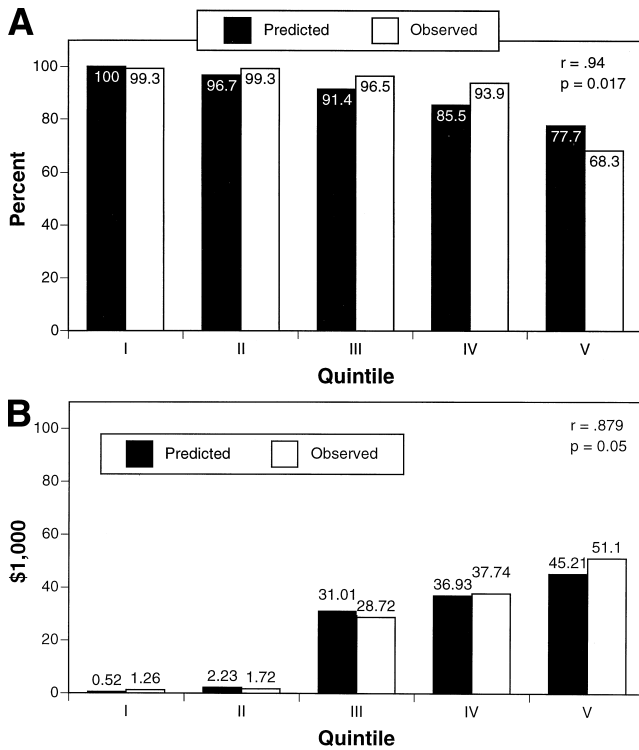
Although quantitative assessment of physician-specific outcomes is being increasingly practiced, several limitations have been identified: 1) the somewhat limited accuracy and concordance of potential models (8,9); 2) the common imprecision of point estimates of patient risk owing to the generally low volume of physician-specific outcomes from which to judge results (8); 3) the susceptibility of potential covariates and end points to manipulation (8); 4) the potential risk inherent in publicizing such results wherein high-risk patients might be declined therapy (10); and 5) the failure of such models to account for the long-term outcome of interventions.

**Table 3.** Independent Baseline Correlates of Event-Free Survival

Covariate	Estimate	SD	95% CI	T-stat	p
CAD	2.28	.58	1.14/3.43	+3.90	< .001
NYHA CHF Class	0.23	.07	.09/.38	+3.11	.002
Stable angina	0.53	.18	.17/.89	+2.87	.004
Creatinine ≥2	0.55	.29	-.02/1.12	+1.88	.064

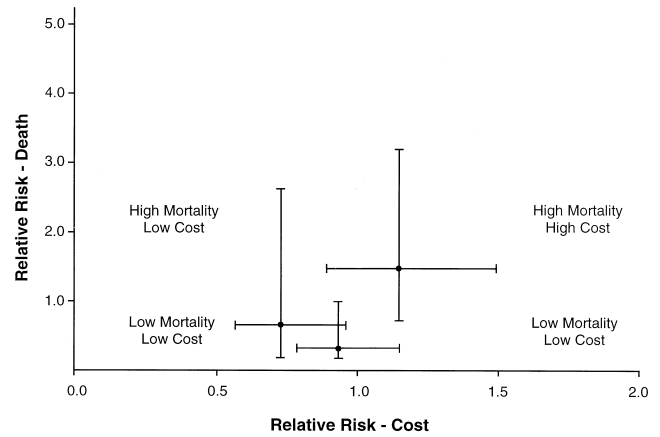
Log likelihood = -1031.499. ROC c-statistic = 0.672. CAD = coronary artery disease; CHF = congestive heart failure; NYHA = New York Heart Association.

“Scorecarding” Physicians on Patient Outcomes



**Figure 2.** Calibration of the models for (A) long-term survival and (B) long-term log<sub>e</sub> cost. Patients were divided into quintiles based upon predicted outcome; actual and predicted values are shown for each quintile.

Our results extend these observations, specifically focusing on long-term outcomes. In attempting to evaluate such outcomes, it is apparent that the first three limitations noted above pose similar problems as in short-term analyses. In addition, we identify several problems specific to the analysis of long-term patient outcome: First, over longer periods of patient follow-up, the responsible physician or physicians may change and it may be difficult to define how one or more physicians should be held “accountable” for the patient outcome. In addition to changing physician providers, when the patient changes provider systems it becomes increasingly difficult to capture all patient-related costs, such that at-



**Figure 3.** Two-dimensional graphic representation of three selected physician-specific outcomes for relative risk of long-term death and the relative risk of long-term log cost. Point estimates and 95% confidence limits are shown.

tempts at deriving estimates of cost-effectiveness of care may be problematic. Second, at least as studied in this somewhat homogeneous treatment setting, the physician provider(s) may have only a very modest impact on long-term patient outcome, relative to the patients’ “baseline” clinical state. Third, models emphasizing the importance of certain outcomes may preferentially identify certain types of physicians as having good or poor outcomes. Nonetheless, use of the best models of this analysis, those for mortality and cost, may provide a reasonable estimate of the cost-effectiveness of care rendered by different providers.

It is axiomatic in biostatistics that unless the physiologic causes of different end points are congruent, that the predictive accuracy of models assessing composite end points will be less accurate than those assessing single end points (assuming comparably powered analyses). Thus, it is not surprising in this analysis that the model for the clinical end point death was the most accurate (c-statistic = 0.83). However, as the capacity to identify physician outliers is directly related to the number of adverse events, composite models are somewhat more likely to allow identification of providers with either good or poor outcome. Nonetheless,

**Table 4.** Correlation Matrix (R) for Physician Outcomes\*

Coefficients for Outcomes	1*	2	3	4	5	6	7	8	9	10
1) Survival-primary		.99	.23	.12	.12	.16	.03	.10	.66	.64
2) Survival-co-primary	—		.18	.03	.04	.15	.09	.08	.67	.63
3) Infarct-free survival-primary	—	—		.91	.22	.11	.23	.02	.00	.05
4) Infarct-free survival-co-primary	—	—	—		.28	.07	.40	.02	.16	.07
5) Event-free survival-primary	—	—	—	—		.86	.85	.79	.43	.33
6) Event-free survival-co-primary	—	—	—	—	—		.57	.85	.39	.35
7) Angina and event-free survival-primary	—	—	—	—	—	—		.63	.48	.39
8) Angina and event-free survival-co-primary	—	—	—	—	—	—	—		.54	.54
9) Cost-primary	—	—	—	—	—	—	—	—		.96
10) Cost-co-primary	—	—	—	—	—	—	—	—	—	

\*1 = Survival-primary (numbers across the top refer to numbering system used along the first column).

one must be concerned about the limited capacity of other models to adjust for patient characteristics that might bias measures of outcome. Thus, as with our analyses of short-term outcomes, the capacity to identify physician outliers becomes highly dependent on the number of procedures or patients that can be analyzed (8).

As with models assessing short-term provider outcome, models assessing long-term outcome may be susceptible to "gaming" (8). Such "code creep" has been identified in longitudinal analyses of the New York State database assessing the results of bypass surgery (1). For our model of long-term survival (Table 2), the covariates assessing CHF status and the urgency of cardiac catheterization, in particular, were noted to be moderately subjective (see Fig. 1). The other elements that were evaluated would be considered rather objective. The covariates of the other models analyzed were no more, and sometimes less, objective (8). Thus, to have credible data for optimal physician-specific outcome measurement, some form of external auditing must be developed.

Even with a fee-for-service system of medical care, it is not unusual for a patient with a chronic disease managed over several years to have several physicians who have direct and major input into management decisions that influence outcome. To hold only one of these physicians responsible for the patient's outcome might be inappropriate. There are no accepted standards of "weighting" the input of various physicians. On the one hand, one might argue that the physician performing a procedure should be held accountable for the results. On the other hand, the physician referring that patient for the procedure would then bear no responsibility for either good or bad outcome. As the United States moves toward the more common use of a managed-care medical system, the identification of one or even a few providers to be held accountable for the patient outcome becomes more problematic. Therefore, it may be much more reasonable to attempt to compare the results of different management systems against one another rather than the results of the individual physicians in those managed-care systems. However, as such systems are made up of individual physicians, it would still be very desirable to identify physicians who have both good and bad outcomes within the systems.

Comparison of provider-related outcomes would be most simple if a single robust model could be used for large groups of patients such as those with CAD. The analyses described herein identify the limitations of such an approach. Patients who had undergone or were awaiting cardiac transplantation were eliminated a priori so as not to bias the model's results against the physicians caring for these patients. Even after this, certain potential models, such as that for MI-free survival, appeared to be biased in favor or against cardiologists who subspecialized in different facets of cardiology. The study was likely underpowered to detect differences between subspecialty groups regarding long-term patient survival. One might argue that disease-

status-specific models (e.g., CAD with CHF) might best be applied, but the logistics of applying this in a practical sense could be quite cumbersome. An alternative, and perhaps more reasonable approach, might be to eliminate from the analysis certain groups of patients who are generally acknowledged to be at particularly high risk—such as transplant patients or patients in cardiogenic shock.

For chronic diseases typified by recurrent symptoms it is generally acknowledged that measures of cost-effectiveness using mortality as the only end point have serious limitations (11). Yet standardized measures of quality of care have not been widely measured in clinical practice, and authorities debate the optimal measure for disease-specific entities (12). Hence, our capacity to judge clinical cost-effectiveness for chronic diseases on a widespread basis necessary for evaluation of large numbers of providers is problematic. Further complicating such an evaluation is the above-noted limited accuracy of many models, patient migration between provider systems, and also lack of standardized statistical methodology necessary to compare, for instance, the accuracy and utility of models with ranked categorical outcome and those with single-end point time-dependent variables. From this analysis it would appear that evaluation on the basis of mortality and cost in this setting would be the most reliable and acceptable measure, acknowledging the fact that our capacity to identify outliers might be more limited than with other modeled end points.

Finally, it is of very considerable interest to note that physicians whose patients tended to have the best clinical outcomes also had the lowest costs. Although it is now well recognized that complications drive up short-term costs in several settings (13), the concern has also been raised that attempts to decrease costs will attenuate the quality of patient care. In the present study, we were unable to demonstrate any evidence of this latter concern.

**Study limitations.** In reviewing the results of this analysis, it is important to take notice of several limitations. First, the data are derived from a somewhat unique cohort of patients from a single tertiary referral center. The extent to which they can be generalized into other settings must be tested. Identification of differences in provider-related outcomes might be more readily achieved in more heterogeneous settings(s) (14-17). The identification of key clinical variables influencing outcome and other methodological issues should not be influenced nearly as much by this issue. Second, the extent to which a physician may actually influence long-term patient outcome is not known. Third, a rather limited number of patients and providers were studied—it is quite likely that in a larger study, other findings not apparent here might become manifest.

Finally, given the large number of statistical tests performed evaluating each physician provider, a p value on the order of .01 to .05 on a single test should not be viewed as conclusive evidence that the physician is an outlier. Nonetheless, the findings reported herein may serve as initial

guidelines by which physicians and physician-care groups be evaluated and, perhaps more importantly, stimulate further research in this still-nascent field.

### Acknowledgments

The authors would like to acknowledge their appreciation to the following physicians who helped in our assessment of the subjectivity of potential covariates and end points: Robert Califf, MD, Kim Eagle, MD, Victor Guetta, MD, Mark Hlatky, MD, Neeraj Jolly, MD, A. Michael Lincoff, MD, David Moliterno, MD, Craig Narins, MD, Mitchell Silver, DO, Samuel Ward, MD, William Weintraub, MD; to Daniel Mark, MD, MPH, for his critical review of the manuscript; and to Ms. Patti Durnwald for her assistance in preparing the manuscript.

---

**Reprint requests and correspondence:** Dr. Stephen G. Ellis, The Cleveland Clinic Foundation, 9500 Euclid Avenue, F-25, Cleveland, Ohio 44195. E-mail: elliss@cesmtp.ccf.org.

---

### REFERENCES

1. Green J, Wintfeld N. Report cards on cardiac surgeons: assessing New York State's approach. *N Engl J Med* 1995; 332:1229-32.
2. Hannan EL, Kilburn HJ, Racz M, Shields E, Chassin MR. Improving the outcomes of coronary artery bypass surgery in New York State. *JAMA* 1994;271:761-6.
3. Ellis SG, Brown KJ, Ellert R, et al. Cost of cardiac care in the three years after coronary catheterization in a contained care system—critical determinants and implications. *J Am Coll Cardiol* 1998;31:1306-13.
4. Boston, Mass: Transition Systems, 1989.
5. Califf RM, Phillips HR III, Hindman MC, et al. Prognostic value of a coronary artery jeopardy score. *J Am Coll Cardiol* 1985;5:1055-63.
6. Harrell FE Jr, Lee KL, Mark DB. Tutorial in biostatistics. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15:361-87.
7. Effron B. Estimating the error rate of a prediction role: improvement on cross-validation. *J Am Statist Assoc* 1983; 78:316-31.
8. Ellis SG, Omoigui N, Bittl JA, et al. Analysis of operator-specific outcomes in interventional cardiology from a multi-center database of 4860 quality-controlled procedures. *Circulation* 1996;93:431-9.
9. Califf RM, Jollis JG, Peterson ED. Operator-specific outcomes: a call to professional responsibility. *Circulation* 1996; 93:403-6.
10. Omoigui NA, Miller DP, Brown KJ, et al. Outmigration for coronary bypass surgery in an era of public dissemination of clinical outcomes. *Circulation* 1995;93:27-33.
11. Spertus JA, Winder JA, Dewhurst TA, Deyo RA, Fihn SD. Monitoring the quality of life in patients with coronary artery disease. *Am J Cardiol* 1994;74:1240-4.
12. Testa MA, Simonson DC. Assessment of quality-of-life outcomes. *N Engl J Med* 1996;334:835-40.
13. Ellis SG, Miller DP, Brown KJ, et al. The in-hospital cost of percutaneous coronary revascularization: critical determinants and implications. *Circulation* 1995;92:741-7.
14. Jollis JG, Peterson ED, Nelson CL, et al. Relationship between physician and hospital coronary angioplasty volume and outcome in elderly patients. *Circulation* 1997;95:2485-91.
15. Hannan EL, Arani DT, Johnson LW, Kemp HG, Lukacik G. Percutaneous transluminal coronary angioplasty in New York State: risk factors and outcomes. *JAMA* 1993;268: 3092-7.
16. Jollis JG, DeLong ER, Peterson ED, et al. Outcome of acute myocardial infarction according to the specialty of the admitting physician. *N Engl J Med* 1996;335:1880-7.
17. Schreiber TL, Elkhatib A, Grines CL, O'Neill WW. Cardiologist versus internist management of patients with unstable angina: treatment patterns and outcomes. *J Am Coll Cardiol* 1995;26:577-82.