

The Cardiac Anesthesia Risk Evaluation Score

A Clinically Useful Predictor of Mortality and Morbidity after Cardiac Surgery

Jean-Yves Dupuis, M.D., F.R.C.P.(C.),* Feng Wang, M.D., M.Sc.,† Howard Nathan, M.D., F.R.C.P.(C.),‡ Miu Lam, Ph.D.,§ Scott Grimes, R.T.,|| Michael Bourke, M.D., F.R.C.P.(C.)#

Background: The Cardiac Anesthesia Risk Evaluation (CARE) score is a simple risk classification for cardiac surgical patients. It is based on clinical judgment and three clinical variables: comorbid conditions categorized as controlled or uncontrolled, surgical complexity, and urgency of the procedure. This study compared the CARE score with the Parsonnet, Tuman, and Tu multifactorial risk indexes for prediction of mortality and morbidity after cardiac surgery.

Methods: In this prospective study, 3,548 cardiac surgical patients from one institution were risk stratified by two investigators using the CARE score and the three tested multifactorial risk indexes. All patients were also given a CARE score by their attending cardiac anesthesiologist. The first 2,000 patients served as a reference group to determine discrimination of each classification with receiver operating characteristic curves. The following 1,548 patients were used to evaluate calibration using the Pearson chi-square goodness-of-fit test.

Results: The areas under the receiver operating characteristic curves for mortality and morbidity were 0.801 and 0.721, respectively, with the CARE score rating by the investigators; 0.786 and 0.710, respectively, with the CARE score rating by the attending anesthesiologists (n = 8); 0.808 and 0.726, respectively, with the Parsonnet index; 0.782 and 0.697, respectively, with the Tuman index; 0.770 and 0.724 with the Tu index, respectively. All risk models had acceptable calibration in predicting mortality and morbidity, except for the Parsonnet classification, which failed calibration for morbidity (P = 0.026).

Conclusions: The CARE score performs as well as multifactorial risk indexes for outcome prediction in cardiac surgery. Cardiac anesthesiologists can integrate this score in their practice and predict patient outcome with acceptable accuracy.

This article is accompanied by an Editorial View. Please see: Fleisher LA: Risk indices: What is their value to the clinician and patient? ANESTHESIOLOGY 2001; 94:191-3.

* Associate Professor, † Research Associate, ‡ Professor, || Respiratory Therapist and Research Assistant, # Assistant Professor, Department of Anesthesia, University of Ottawa Heart Institute. § Associate Professor, Department of Community Health and Epidemiology, Queen's University.

Received from the Department of Anesthesia, University of Ottawa Heart Institute, Ottawa, Ontario, Canada. Submitted for publication April 28, 2000. Accepted for publication September 25, 2000. Support was provided solely from institutional and/or departmental sources. Presented in part at the annual meeting of the Society of Cardiovascular Anesthesiologists, Seattle, Washington, April 25-29, 1998.

Address reprint requests to Dr. Dupuis: Department of Anesthesia, University of Ottawa Heart Institute, 40 Ruskin Street, Room H346, Ottawa, Ontario, Canada K1Y 4W7. Address electronic mail to: jydupuis@ottawaheart.ca. Individual article reprints may be purchased through the Journal Web site, www.anesthesiology.org.

THE growing interest in risk-adjusted analysis of outcome in cardiac surgery has led to the development and validation of several predictive models for postoperative mortality, morbidity, and prolonged hospital stay.¹⁻¹¹ Most models are multifactorial risk indexes developed using multiple regression analysis. Despite their potential usefulness for quality assurance and perioperative care planning, multifactorial risk indexes remain poorly integrated into clinical practice. This is probably because of their complexity in use, their inaccuracy in predicting outcome for individual patients, and their dependence on clinical variables that are not always available.^{10-12,16} In contrast to multifactorial risk indexes, functional classifications like the New York Heart Association classification or the American Society of Anesthesiologists (ASA) physical status classification are routinely used by anesthesiologists. However, those classifications are not designed to predict outcome after cardiac surgery. Consequently, their predictive ability in this setting is limited and inconsistent.¹⁷

Previous studies in cardiac surgery have demonstrated that a large amount of prognostic information can be obtained from a few clinical variables^{18,19} or clinical judgment alone.^{17,20,21} On that basis, we developed the Cardiac Anesthesia Risk Evaluation (CARE) score, which is a simple risk classification with an ordinal scale (table 1). The CARE score combines clinical judgment and the recognition of three risk factors previously identified by multifactorial risk indexes: comorbid conditions categorized as controlled or uncontrolled, the surgical complexity, and the urgency of the procedure.

In this study, we hypothesized that the CARE score would be a valid predictor of outcome after cardiac surgery and that clinicians would easily integrate this risk model into their practice. Accordingly, the study had three specific objectives: first, to determine the predictive performance of the CARE score in predicting mortality and major postoperative complications; second, to compare the predictive performance of the CARE score with that of three existing multifactorial risk indexes¹⁻³ for cardiac surgical patients; and finally, to determine the interrater variability and predictive performance of the CARE score when used by experienced cardiac anesthesiologists.

Table 1. Cardiac Anesthesia Risk Evaluation Score

1. Patient with stable cardiac disease and no other medical problem. A noncomplex surgery is undertaken.
2. Patient with stable cardiac disease and one or more controlled medical problems.* A noncomplex surgery is undertaken.
3. Patient with any uncontrolled medical problem† or patient in whom a complex surgery is undertaken.‡
4. Patient with any uncontrolled medical problem *and* in whom a complex surgery is undertaken.
5. Patient with chronic or advanced cardiac disease for whom cardiac surgery is undertaken as a last hope to save or improve life.
- E. Emergency: surgery as soon as diagnosis is made and operating room is available.

* Examples: controlled hypertension, diabetes mellitus, peripheral vascular disease, chronic obstructive pulmonary disease, controlled systemic diseases, others as judged by clinicians. † Examples: unstable angina treated with intravenous heparin or nitroglycerin, preoperative intraaortic balloon pump, heart failure with pulmonary or peripheral edema, uncontrolled hypertension, renal insufficiency (creatinine level > 140 $\mu\text{mol/l}$), debilitating systemic diseases, others as judged by clinicians. ‡ Examples: reoperation, combined valve and coronary artery surgery, multiple valve surgery, left ventricular aneurysmectomy, repair of ventricular septal defect after myocardial infarction, coronary artery bypass of diffuse or heavily calcified vessels, other as judged by clinicians.

Methods

Population

This was a prospective observational study approved by the Human Research Ethics Committee of the University of Ottawa Heart Institute, Ottawa, Ontario, Canada. Written consents were not obtained from the individual patients because the study was based on data collected for routine care. A total of 3,548 consecutive patients who underwent a cardiac surgical procedure at the University of Ottawa Heart Institute were included. The first 2,000 patients, who had surgery between November 12, 1996 and March 18, 1998, served as a reference group to develop logistic regression models for each risk model tested in the study. The next 1,548 patients, operated on between March 19, 1998 and April 2, 1999, were used for validation of the risk models. Patients undergoing heart transplantation or implantation of ventricular assist devices as a primary surgery were excluded because of the infrequency of those procedures. Patients who underwent more than one cardiac surgical procedure during the same hospitalization were counted as single cases. However, subsequent cardiac or noncardiac procedures were computed as postoperative complications, unless planned before the primary cardiac intervention.

Development of the Cardiac Anesthesia Risk Evaluation Score

To facilitate its use by anesthesiologists, the CARE score was designed to resemble the ASA physical status classification, a model also familiar to surgeons. The rationale behind the definition of each risk category of the CARE score is based on general and accepted knowl-

edge in cardiac surgery (table 1). For example, the definitions of the first category (CARE 1) and last category (CARE 5) are based on the fact that clinicians can correctly identify very low- and very high-risk cardiac surgical patients.^{17,21}

In contrast to the assessment of the two risk extremes in cardiac surgery, the subjective estimation of patients with intermediate risk is often inaccurate and inconsistent.²¹ The use of a few objective clinical variables results in better risk prediction for those patients.^{18,19,22,23} Therefore, two general, but objective, groups of risk factors were selected to define the intermediate risk levels in the CARE score (CARE 2-4): the complexity of the surgery and the presence of various comorbid conditions, which may be categorized as controlled or uncontrolled. The ranking or relative importance of those covariates in the CARE score is consistent with the findings from most existing multifactorial risk indexes.¹⁻⁸ Patients with controlled medical problems (*e.g.*, diabetes mellitus, hypertension, etc.) are at greater risk than patients without any disease. However, they are at lower risk than patients with uncontrolled diseases (heart failure with pulmonary edema, renal insufficiency, etc.) thus, the rationale for the CARE 2 and 3 categories. Uncontrolled comorbid factors and complex or difficult procedures have comparable scores in most multiple logistic regression models. Thus, the same prognostic weight was given to both groups of factors in the CARE score. This explains why one or the other group can be used to define the CARE 3 category. The CARE 4 category accounts for the fact that uncontrolled medical conditions and complex procedures have an additive effect on risk.¹⁻⁸

Finally, special consideration is given to emergency in the CARE score, so that emergency cases can be easily differentiated from others. This is because emergencies or catastrophic states are the most important predictors of outcome in cardiac surgery.¹⁻⁸ Therefore, the CARE score has eight possible risk categories: scores 1-5 for elective or urgent procedures, and scores 3E, 4E, and 5E for emergency conditions requiring immediate surgery. Unstable cardiac conditions requiring surgery within 24 h, but not immediately, are considered uncontrolled medical problems. By definition, emergency never applies to CARE 1 or 2.

Data Collection

Preoperative and intraoperative data were collected prospectively by the attending anesthesiologists, who completed a database form at the time of surgery. The database form contains 130 preoperative variables pertaining to the severity of the patient's disease and comorbid factors before the operation and 80 variables documenting intraoperative procedures and events. Research assistants who also looked after the Cardiac Surgical Unit database verified the accuracy and complete-

ness of the collected data on a daily basis. Postoperative outcome data were retrieved from the medical charts after patients' discharge and from the Cardiac Surgical Unit database, which contains 92 variables related to postoperative evolution. The quality of the collected data was assessed by an independent observer, who extracted 50% of the database information from 175 randomly selected patients and compared them with those found in the charts. An agreement rate of 98% was found between the database information and the data obtained from the charts.

Outcomes

The primary outcomes in this study were in-hospital mortality, regardless of length of stay (LOS), and morbidity, defined as one or more of the following: (1) cardiovascular—low cardiac output, hypotension, or both treated with intraaortic balloon pump, with two or more intravenous inotropes or vasopressors for more than 24 h, or with both, malignant arrhythmia (asystole and ventricular tachycardia or fibrillation) requiring cardiopulmonary resuscitation, antiarrhythmia therapy, or automatic cardioresuscitator implantation; (2) respiratory—mechanical ventilation for more than 48 h, tracheostomy, reintubation; (3) neurologic—focal brain injury with permanent functional deficit, irreversible encephalopathy; (4) renal—acute renal failure requiring dialysis; (5) infectious—septic shock with positive blood cultures, deep sternal or leg wound infection requiring intravenous antibiotics, surgical debridement, or both; (6) other—any surgery or invasive procedure necessary to treat a postoperative adverse event associated with the initial cardiac surgery.

In the absence of morbidity data, prolonged postoperative LOS in hospital has been used as a surrogate for morbidity in other studies.^{3,10} This was, therefore, analyzed as a secondary outcome in this study. Prolonged postoperative LOS in the hospital was defined as a stay of 14 days or more. This corresponds to the 90th percentiles for postoperative LOS in the entire study population. This cutoff point was proposed in a previous study because it likely reflects a LOS resulting from complications rather than differences in discharge practice.³

Risk Classification of All Patients

Using the validated preoperative database information, two investigators (JYD and SG) gave a CARE score to each patient. Throughout the study, the investigators used strict definitions of uncontrolled medical problems and complex surgical procedures, as presented in the footnotes of table 1. Multifactorial risk scores were also determined for each patient according to the risk indexes developed for general cardiac surgical populations by Parsonnet *et al.*,¹ Tuman *et al.*,² and Tu *et al.*³ (table 2). Those classifications contain variables available in most of our patients, and like the CARE score, they apply

to all cardiac surgical patients, not only to those undergoing coronary artery surgery. Patients were risk stratified according to the original criteria and definitions described in each of those risk classifications.¹⁻³ To attenuate the reported inconsistency associated with two subjective risk factors (catastrophic states and rare circumstances) in the original Parsonnet classification,¹⁶ 60 conditions potentially computable under those factors were listed and given a risk value at the beginning of the study. This list of conditions and scores was used consistently by the investigators to compute Parsonnet risk score, similar to what was done by Gabrielle *et al.*

Use of the Cardiac Anesthesia Risk Evaluation Score by Clinicians

Eight experienced cardiac anesthesiologists participated in the study. They were asked to provide a CARE score for all their patients, before surgery. The investigators' ratings served as a reference for comparison with both the anesthesiologists' ratings and the multifactorial risk indexes.

Statistical Analysis

The association between the patients' characteristics and mortality or morbidity was determined by univariate analysis, using a chi-square test or a Fisher exact test when appropriate. For each risk index considered in this study, including the CARE score, different predictive models for mortality, morbidity, and prolonged postoperative LOS were developed using logistic regression analysis. The CARE score categories were coded as 1 = 1, 2 = 2, 3 = 3, 4 = 3E, 5 = 4, 6 = 4E, 7 = 5, and 8 = 5E; those numeric scores were used as independent variables in the logistic regression models, because the models do not handle nominal scores such as 3E, 4E, and 5E. In the cases of the multifactorial risk indexes, logistic regression models were similarly developed, using the original risk categories (not the total score or integer) proposed by the developers of those indexes. The predictive performances of all predictive models were assessed by determining their discrimination and calibration for mortality, morbidity, and prolonged LOS.

Discrimination, or predictive accuracy, was assessed for all predictive models by building receiver operating characteristic (ROC) curves for mortality, morbidity, and prolonged postoperative LOS.²⁴ The ROC curve is a graphic technique plotting the true-positive rate (sensitivity) versus the false-positive rate (1-specificity) for diagnostic tests, using different cutoff points (in this study, those points were the various categories of each risk classification). The top of the y-axis represents a perfect test with a 100% true-positive rate and a 0% false-positive rate. The area under the ROC curve equals the probability of correctly identifying the patient with a complication when applying the risk classification to a pair of randomly selected patients (always one with a

Table 2. Multifactorial Risk Indexes for the Prediction of Outcome after Cardiac Surgery

Parsonnet Score		Tuman Score		Tu Score	
Age (yr)		Age (yr)		Age (yr)	
70–74	7	65–74	1	65–74	2
75–79	12	≥ 75	2	≥ 75	3
≥ 80	20		—		—
Emergency after cath.	10	Emergency within 24 h	4	Emergency within 24 h	4
	—		—	Urgent*	1
LV function		LV function		LV function	
EF 30–49%	2	EF < 35%	1	EF 35–50%	1
EF < 30%	4		—	EF 20–34%	2
	—		—	EF < 20%	3
Surgical characteristics		Surgical characteristics		Surgical characteristics	
MVR or AVR	5	MVR or AVR	1	Single valve	
CABG + valve	2	Complex†	2	Complex†	
Reoperation		Reoperation		Reoperation	
1st	5		2		
2nd	10		—		
Female gender	1	Female gender	2	Female gender	
Dialysis dependency	10	Creatinine > 125 μmol/l	2		
Systolic PAP > 60 mmHg	8	Mean PAP ≥ 25% MAP	2		
Diabetes (any type)	3		—		
Morbid obesity‡	3		—		
A-V gradient > 120 mmHg	7		—		
Preoperative IABP	2		—		
Ventricular aneurysm	5		—		
Hypertension§	3		—		
Catastrophic states	10–50		—		
Rare circumstances	2–10	Cerebrovascular disease	2		
	—	Age of previous MI			
	—	6 weeks–6 months	1		
	—	< 6 weeks#	2		
Maximum score	158	Maximum score	21	Maximum score	15

* Diagnosis and surgery within the same admission. † Multivalve or coronary artery bypass graft (CABG) plus valve. ‡ $1.5 \times$ the ideal weight in original study body mass index greater than 34 in this study. § Systolic blood pressure greater than 140 mmHg in original study¹; > 160 mmHg in this study. || 3–6 months in original study.² # < 3 months in original study.²

Cath. = coronary angiogram or percutaneous angioplasty, LV = left ventricular; EF = ejection fraction; MVR = mitral valve replacement or repair; AVR = aortic valve replacement; PAP = pulmonary artery pressure; MAP = mean arterial pressure; A-V = aorto-ventricular; IABP = intraaortic balloon pump; MI = myocardial infarction.

complication and one without a complication) on successive trials. Thus, the area under the ROC curve is commonly used to measure and compare the predictive accuracy of risk classifications. An area under the ROC curve of 1.0 indicates perfect accuracy of the risk classification, whereas an area less than 0.5 (line of no discrimination) means that it is no better than chance. Areas of 0.5 to 0.7 suggest a low accuracy and values more than 0.7 confirm the usefulness of the risk classification as a risk predictor.²⁴ In this study, the area under the various ROC curves and their standard error (SE) were measured and compared by performing the two-tailed nonparametric ROC analysis of DeLong *et al.*,²⁵ using the statistical program AccuROC for Windows 95 (Accumetric Corporation, Montreal, Canada), with correction made for multiple comparisons.

All risk indexes in the study, including the CARE score, were tested as categorical. Data were tabulated in contingency tables and calibration, which represents the precision of the probabilities generated by a prediction model, was assessed using the Pearson chi-square good-

ness-of-fit test. This is the most commonly used statistic for contingency tables.²⁶ It compares the estimated predicted outcomes (mortality, morbidity, or prolonged LOS) from the logistic regression models with the observed outcomes for each risk category of the prediction model.^{26,27} A small chi-square value (or a *P* value > 0.05) indicates acceptable calibration.

The interrater variability in using the CARE score was determined by measuring the concordance rate and the κ measure of agreement between the attending anesthesiologists' and the investigators' assessments. This analysis was first performed with the data from the entire population. It was then repeated with the data from the reference and validation groups separately, to determine any possible change with use (learning effect) of the CARE score over time. The discrimination and calibration analyses were also performed for the CARE score ratings by the attending anesthesiologists. To confirm the usefulness and uniqueness of the CARE score, its discrimination in predicting outcome was compared with that of other variables commonly used by clini-

cians: ASA physical status, New York Heart Association classification for heart failure, left ventricular ejection fraction, age, serum creatinine, operative priority, and type of surgery.

For all analyses and comparisons, a P value less than 0.05 was used to determine statistical significance.

Results

Population

The patients' characteristics in the reference group ($n = 2,000$) and their association with mortality and morbidity are presented in table 3. The mortality and morbidity rates in the reference group were 3.4% and 20.7%, respectively. Comparable patients' characteristics, mortality (3.4%), and morbidity (22.2%) rates were found in the validation group ($n = 1548$). The mortality and morbidity rates were comparable between the eight cardiac surgeons who participated in the study. The mean postoperative LOS was 8.8 ± 11.0 days in the reference population and 9.0 ± 10.3 days in the validation group, with a median of 6 days in both groups. The incidences of prolonged postoperative LOS were 10.2% and 12.3% in the reference and validation groups, respectively. The mortality rate in this study compares very well with values recently obtained through large multicenter databases.^{28,29} The morbidity rate and the incidence of prolonged postoperative LOS are more difficult to compare with those from other studies because of variations in outcome definitions.

The Cardiac Anesthesia Risk Evaluation Score as a Predictive Risk Model

Table 4 shows the probabilities of mortality, morbidity, and prolonged postoperative LOS associated with each category of the CARE score, as determined from logistic regression analysis in the reference population.

Predictive Performance of the Risk Classifications

In the reference group, all risk classifications had comparable areas under the ROC curves for the prediction of mortality and morbidity (figs. 1 and 2). With all risk classifications, the discrimination for mortality was significantly better than for morbidity. Age being a risk factor not explicitly taken into account by the CARE score, the discrimination of the CARE score in predicting mortality and morbidity was further tested in various age subgroups. The areas under the ROC curve for the prediction of mortality and morbidity were 0.791 ± 0.067 and 0.740 ± 0.024 in the patients younger than 65 yr of age, 0.763 ± 0.045 and 0.721 ± 0.026 in the patients 65–74 yr of age, and 0.795 ± 0.049 and 0.715 ± 0.031 in the patients 75 yr of age or older, respectively. For prolonged postoperative LOS, the area under the ROC curve was 0.715 ± 0.018 with the CARE score, $0.774 \pm$

0.016 with the Parsonnet classification ($P = 0.05$ vs. all other classifications), 0.730 ± 0.019 with the Tuman classification, and 0.730 ± 0.018 with the Tu classification.

In the validation group, the areas under the ROC curve for the prediction of mortality and morbidity were 0.807 ± 0.031 and 0.721 ± 0.016 with the CARE score, 0.804 ± 0.026 and 0.698 ± 0.017 with the Parsonnet classification, 0.823 ± 0.030 and 0.699 ± 0.017 with the Tuman classification, and 0.801 ± 0.032 and 0.688 ± 0.017 with the Tu classification, respectively. A significant difference was found between the CARE score and the Tu classification in predicting morbidity ($P = 0.029$). For the prediction of prolonged postoperative LOS, the area under the ROC curve was 0.728 ± 0.020 with the CARE score, 0.769 ± 0.017 with the Parsonnet classification ($P < 0.05$ vs. all other classifications), 0.720 ± 0.020 with the Tuman classification, and 0.741 ± 0.018 with the Tu classification.

The calibration analysis for mortality in the validation group showed an acceptable fit between the observed and expected values for all risk classifications (table 5–8). For morbidity, an acceptable level of agreement between the observed and expected values was found for the CARE score, the Tuman, and the Tu classifications, but not for the Parsonnet classification (table 5–8). For prolonged postoperative LOS, all classifications failed the calibration analysis: $P = 0.014$ with the CARE score, $P = 0.036$ with the Parsonnet classification, $P = 0.012$ with the Tuman classification, and $P = 0.02$ with the Tu classification.

Use of the Cardiac Anesthesia Risk Evaluation Score by Clinicians

An overall concordance rate of 85.1% in CARE score ratings was found between the investigators and the eight participating cardiac anesthesiologists, with a κ value of 0.790 (SE = 0.008; $P < 0.001$). In the reference group, a concordance rate of 86.3% and a κ value of 0.806 (SE = 0.011; $P < 0.001$) were found between the two ratings. A comparable concordance rate of 83.6% and a κ value of 0.770 (SE = 0.013; $P < 0.001$) were found in the validation group, suggesting no significant change over time in the anesthesiologists' use of the CARE score.

The CARE score used by the attending anesthesiologists had areas under the ROC curve of 0.782 ± 0.028 for mortality, 0.710 ± 0.016 for morbidity, and 0.715 ± 0.018 for prolonged postoperative LOS in the reference group. In the validation group, the areas under the ROC curves were 0.789 ± 0.031 for mortality, 0.721 ± 0.016 for morbidity, and 0.710 ± 0.019 for postoperative LOS. Those values were not significantly different from those obtained from the CARE score used by the two investigators. The Pearson goodness-of-fit test for the CARE score used by clinicians showed an acceptable fit be-

Table 3. Patient Characteristics in the Reference Group (n = 2000) and Association with Perioperative Outcomes, as Determined by Univariate Analysis

Characteristics	Prevalence (%)	Mortality		Morbidity	
		%	P Value	%	P Value
Age (yr)					
< 65	49.1	1.6	—	14.7	—
65–74	34.4	3.6	0.009	23.4	<0.001
≥ 75	16.5	8.2	<0.001	32.7	<0.001
Gender					
Male	74.8	3.1	—	19.3	—
Female	25.2	4.4	0.164	24.1	0.010
LV ejection fraction					
≥ 50%	65.5	1.7	—	16.4	—
30–49%	21.2	5.0	<0.001	25.3	<0.001
< 30%	8.3	7.1	<0.001	34.5	<0.001
Unknown	5.0	13.1	<0.001	33.3	<0.001
Operative priority					
Elective	48.7	1.5	—	15.7	—
Urgent*	46.2	3.9	0.002	21.3	0.002
Emergency (within 24 h)	1.9	10.3	<0.001	51.3	<0.001
Immediate surgery	3.2	20.3	<0.001	67.2	<0.001
Type of surgery					
CABG	75.7	2.5	—	16.6	—
ASD repair	1.5	0	0.796	6.9	0.023
Single valve	8.5	2.4	0.902	23.5	0.023
Combined or complex	14.3	9.1	<0.001	42.0	<0.001
Reoperation					
No	89.2	2.9	—	18.6	—
Yes	10.8	7.41	<0.001	37.5	<0.001
Cardiogenic shock					
No	98.4	2.7	—	19.4	—
Yes	1.6	43.8	<0.001	96.9	<0.001
Congestive heart failure					
NYHA Classes 1 and 2	85.1	1.8	—	16.6	—
NYHA Class 3	9.4	7.4	<0.001	36.0	<0.001
NYHA Class 4	5.5	20.9	<0.001	72.7	<0.001
Previous myocardial infarction					
Never	54.1	2.6	—	19.9	—
> 6 Weeks ago	26.2	3.1	0.664	16.3	0.076
Within 6 weeks	18.6	5.8	0.003	28.4	<0.001
Unstable angina treated with intravenous heparin or nitroglycerin					
No	77.1	3.1	—	19.8	—
Yes	22.9	4.2	0.275	23.4	0.090
Preoperative IABP					
No	98.2	3.2	—	19.5	—
Yes	1.8	14.3	<0.001	85.7	<0.001
Left main disease (≥ 50% stenosis)					
No	82.5	3.1	—	19.9	—
Yes	17.5	4.9	0.098	24.3	0.064
Cerebrovascular disease					
No	84.4	2.7	—	19.4	—
Yes	15.6	7.4	<0.001	27.7	<0.001
Peripheral vascular disease					
No	84.6	2.6	—	19.0	—
Yes	15.4	7.82	<0.001	29.6	<0.001
Hypertension					
No	51.9	3.0	—	18.3	—
Controlled†	44.2	3.7	0.389	22.9	0.012
Uncontrolled‡	3.9	5.2	0.461	27.3	0.006
Pulmonary hypertension§					
No	93.5	2.5	—	18.2	—
Yes	6.5	16.9	<0.001	55.4	<0.001
COPD on medications					
No	95.4	3.2	—	20.4	—
Yes	4.6	6.5	0.091	26.1	0.187
Diabetes mellitus					
No	79.1	3.5	—	20.3	—
On oral hypoglycemic agent	13.9	2.5	0.413	20.9	0.842
Insulin dependent	7.0	4.3	0.606	23.7	0.342
Creatinine ≥ 125 μmol/l					
No	86.9	2.2	—	17.7	—
Yes	13.1	11.1	<0.001	40.5	<0.001
Body mass index					
21–34	88.0	3.4	—	20.7	—
≤ 20	3.5	8.6	0.021	25.7	0.314
> 34	8.5	1.8	0.378	17.8	0.360

For each characteristic, the first category was taken as the reference for the 2 × 2 contingency table analysis.

* Diagnosis and surgery during same admission. † Systolic blood pressure 160 mmHg or less and diastolic blood pressure 105 mmHg or less. ‡ Systolic blood pressure greater than 160 mmHg or diastolic blood pressure greater than 105 mmHg. § Systolic pulmonary artery pressure greater than 50 mmHg.

LV = left ventricular; CABG = coronary artery bypass graft; ASD = atrial septal defect; NYHA = New York Heart Association; IABP = intraaortic balloon pump; COPD = chronic obstructive pulmonary disease.

Table 4. Probabilities of Mortality, Morbidity and Prolonged Postoperative Length of Stay in Hospital, as Predicted by the CARE Score

CARE Score	Mortality (%)	Morbidity (%)	Prolonged LOS (%)
1	0.5 (0.3–0.9)	5.4 (4.3–6.8)	2.9 (2.2–3.9)
2	1.1 (0.7–1.7)	10.3 (8.9–12.1)	5.1 (4.2–6.3)
3	2.2 (1.6–3.1)	19.0 (17.2–20.9)	8.8 (7.6–10.2)
3E	4.5 (3.5–5.7)	32.1 (29.3–35.0)	14.7 (12.8–16.8)
4	8.8 (6.9–11.3)	48.8 (44.1–53.6)	23.5 (20.1–27.3)
4E	16.7 (12.4–22.1)	65.8 (59.5–71.6)	35.4 (29.3–42.0)
5	29.3 (20.8–39.6)	79.6 (73.2–84.7)	49.4 (40.4–58.5)
5E	46.2 (32.4–60.5)	88.7 (83.5–92.5)	63.6 (52.5–73.4)

Values obtained from the logistic regression analysis performed in the reference population ($n = 2,000$). Numbers in parentheses are 95% confidence intervals.

CARE = Cardiac Anesthesia Risk Evaluation; LOS = length of stay.

tween the observed and expected rates of mortality (chi-square = 3.056; $df = 8$; $P = 0.931$) and morbidity (chi-square = 14.174; $df = 8$; $P = 0.077$), but not for prolonged postoperative LOS ($P = 0.045$).

When compared with other clinical variables used by clinicians in the reference population, the CARE score predicted mortality and morbidity significantly better than any of those markers alone (table 9), confirming its uniqueness as a clinical tool for risk assessment and classification of cardiac surgical patients.

Discussion

The results of this study show that the CARE score is an accurate predictor of mortality and morbidity after

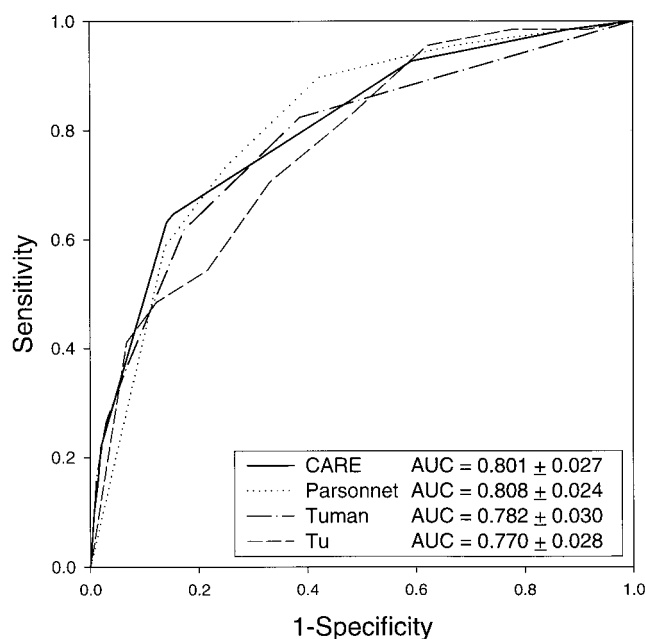


Fig. 1. Receiver operating characteristic curves obtained with each risk model for prediction of mortality in the reference group ($n = 2,000$ patients). CARE = Cardiac Anesthesia Risk Evaluation; AUC = area under the curve.

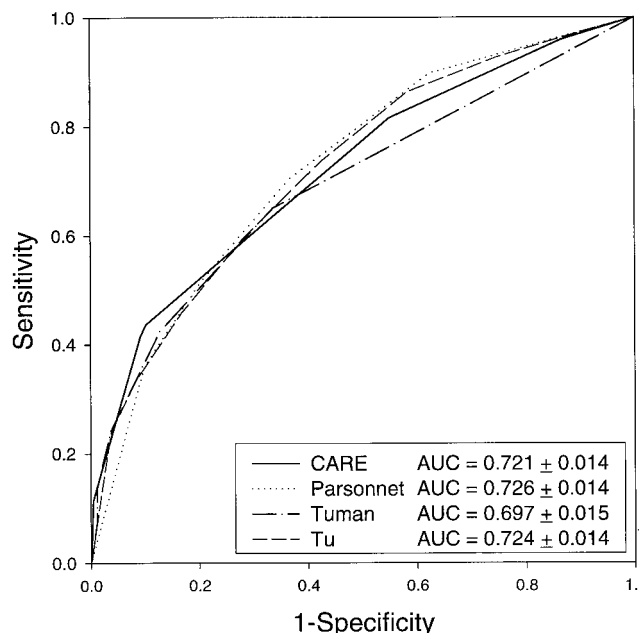


Fig. 2. Receiver operating characteristic curves obtained with each risk model for prediction of morbidity in the reference group ($n = 2,000$ patients). CARE = Cardiac Anesthesia Risk Evaluation; AUC = area under the curve.

cardiac surgery. Its discrimination and calibration for the prediction of mortality and morbidity compare very well with those of more complex multifactorial risk indexes. The study also suggests that experienced cardiac anesthesiologists can integrate the CARE score into their clinical practice in a consistent manner that will provide accurate predictions of outcome after cardiac surgery.

Many mathematical modeling techniques are available to quantify the risk associated with cardiac surgery. Of those techniques, multiple regression analysis is probably the most commonly used.^{1–8} With this statistic, independent risk factors are identified and entered in a complex equation that expresses the probability of adverse outcomes. For practical reasons, continuous data derived from this equation are converted to a multifactorial risk index, where the predictive value of each risk factor is reduced to an integer, and the sum of the integers determines the risk category to which individual patients belong. In general, the discrimination provided by those risk indexes, as determined by the area under the ROC curve, is in the range of 0.65–0.85 for mortality, morbidity, or increased postoperative LOS in hospital.^{3–5,7,9–13,15} So, most existing multifactorial risk indexes meet the acceptability criteria for risk-adjusted analysis of outcomes in cardiac surgery.

Despite their apparent simplicity, multifactorial risk indexes are difficult to memorize. For example, the Tu classification,³ which is the simplest model tested in this study, counts six risk factors and 17 point strata from which nine risk categories can be derived. This is an obvious handicap for daily application in the practice of

Table 5. Pearson Chi-square Goodness-of-fit Test for Prediction of Mortality and Morbidity with the CARE Score

CARE Score	n	Mortality				Morbidity			
		No		Yes		No		Yes	
		Observed	Expected	Observed	Expected	Observed	Expected	Observed	Expected
1	169	169	168.098	0	0.902	159	159.839	10	9.161
2	482	478	476.721	4	5.279	428	431.983	54	50.017
3	634	619	619.829	15	14.171	493	513.812	141	120.188
3E	30	27	28.648	3	1.352	14	20.372	16	9.628
4	197	179	179.507	18	17.493	105	100.781	92	96.219
4E	7	3	5.827	4	1.173	3	2.39	4	4.610
5	16	11	11.304	5	4.696	2	3.267	14	12.733
5E	13	9	6.998	4	6.002	0	1.465	13	11.535
$\chi^2 = 12.847; df = 8; P = 0.117$					$\chi^2 = 13.960; df = 8; P = 0.083$				

CARE = Cardiac Anesthesia Risk Evaluation; n = total number of patients in each risk category; df = degrees of freedom.

cardiac anesthesia and surgery. The CARE score is a proposed alternative to this cumbersome approach. It is a simple risk ranking system designed for routine use by cardiac anesthesiologists and surgeons. It allows some clinical judgment within a framework of general concepts related to risk in cardiac surgery. This approach may become appealing to clinicians, especially if shown to be as good a predictor as more complicated multifactorial risk indexes.

In this study, the CARE score was submitted to the usual evaluation criteria for predictive risk models in two consecutive cohorts of patients, a reference and a validation group. Two investigators gave a CARE score to each patient and determined their risk category according to the multifactorial indexes developed by Parsonnet *et al.*,¹ Tuman *et al.*,² and Tu *et al.*³ The CARE score predicted postoperative mortality or morbidity with as much accuracy as the multifactorial risk indexes. In fact, only the Parsonnet classification (which is the most complex model tested in this study) and the CARE score predicted mortality with areas under the ROC curve equal to or more than 0.80 in both cohorts of patients. Furthermore, the CARE score is the only model that provided areas under the ROC curve more than 0.70 for the prediction of morbidity in the two groups of patients. It also provided an acceptable level of agreement

between the observed and expected rates of mortality and morbidity in the validation set of patients. For those two outcomes, the fit was not perfect for all the CARE score categories, but none of the multifactorial indexes had a perfect fit for all its categories. In the case of the CARE score, poor fits were observed with the 3E and 4E categories, where the predicted mortality was underestimated, and with the 3E category, where the predicted morbidity was also underestimated. The small number of patients and expected outcome in those categories may explain those results.³⁰

A common finding to all the risk indexes tested in this study was their lower accuracy in predicting morbidity or prolonged postoperative LOS as compared with mortality. This difference has also been observed in previous studies.^{3,5} Those results suggest that some complications defining morbidity and some causes of prolonged postoperative LOS are not well accounted for by the risk factors used by the CARE score or the tested multifactorial risk indices. Another common finding to the CARE score and the tested multifactorial risk indexes was their poor calibration for the prediction of prolonged postoperative LOS. The reason for this lack of calibration is unclear. The mean and median postoperative LOS were comparable between the reference and the validation groups, suggesting that a significant change in practice

Table 6. Pearson Chi-square Goodness-of-fit Test for Prediction of Mortality and Morbidity with the Parsonnet Classification

Risk Category	n	Mortality				Morbidity			
		No		Yes		No		Yes	
		Observed	Expected	Observed	Expected	Observed	Expected	Observed	Expected
1 (0-4)	525	524	522.716	1	2.284	465	482.838	60	42.162
2 (5-9)	350	344	346.636	6	3.364	297	302.644	53	47.356
3 (10-14)	266	258	259.466	8	6.534	207	207.750	59	58.250
4 (15-19)	163	153	153.707	8	9.293	117	108.491	46	54.509
5 (≥ 20)	244	214	213.077	30	30.923	118	128.400	126	115.600
$\chi^2 = 2.842; df = 5; P = 0.724$					$\chi^2 = 12.770; df = 5; P = 0.026$				

Numbers in parentheses represent the range of risk scores in points for each risk category. n = total number of patients in each risk category; df = degrees of freedom.

Downloaded from https://www.atsjournals.org/ by guest on 04 March 2024

Table 7. Pearson Chi-square Goodness-of-fit Test for Prediction of Mortality and Morbidity with the Tuman Classification

Risk Category	n	Mortality				Morbidity			
		No		Yes		No		Yes	
		Observed	Expected	Observed	Expected	Observed	Expected	Observed	Expected
1 (< 3)	941	935	928.521	6	12.479	821	828.599	120	112.401
2 (4-5)	295	287	286.977	8	8.023	223	231.361	72	63.639
3 (6-7)	167	152	157.821	15	9.179	100	107.206	67	59.794
4 (8-9)	90	80	80.286	10	9.714	42	42.234	48	47.766
5 (10-11)	34	27	27.163	7	6.837	14	10.324	20	23.676
6 (\geq 12)	21	14	13.783	7	7.217	4	3.717	17	17.283
$\chi^2 = 7.340; df = 6; P = 0.291$					$\chi^2 = 5.245; df = 6; P = 0.513$				

Numbers in parentheses represent the range of risk scores in points for each risk category. n = total number of patients in each risk category; df = degrees of freedom.

pattern between the reference and validation groups is unlikely to be the cause of poor calibration for this outcome.

One major objective of this study was to determine the performance of the CARE score when used by cardiac anesthesiologists in their daily practice. As previously observed with the ASA physical status classification,^{31,32} differences in CARE score ratings were expected between the attending anesthesiologists and the investigators. However, a very high agreement between the two ratings was observed, with an overall concordance rate of 85% for the 3,548 studied patients. Consequently, clinicians predicted mortality, morbidity, and prolonged postoperative LOS with almost as much accuracy as the investigators. Because only two ratings were obtained for each patient in this study, the whole range of variation in CARE score rating remains undetermined. However, the overall results suggest that the participating anesthesiologists can use the CARE score in a manner that is consistent enough to provide appropriate risk-adjusted analysis of outcome in their institution.

The subjective risk assessment of cardiac surgical patients, using the equivalent of a visual analog score, is

another simple alternative to multivariate risk analyses. This approach was recently tested and compared with multivariate risk model in 1,198 patients from seven centers. All patients were given a subjective risk score of 1-5 by their attending surgeon. This subjective risk model provided an area under the ROC curve of 0.70 for the prediction of mortality. This was significantly less than the 0.76 value obtained with the multivariate risk model in the same population. The subjective risk assessment was accurate in identifying the very low- and very high-risk patients, but inaccurate for those with intermediate risk. In the present study, the CARE score predicted mortality as well as any of the tested multivariate risk models. In addition, the CARE score was useful in predicting outcome for patients with intermediate risk levels. Thus, the CARE score may have certain advantages over a purely subjective risk model. Only comparisons of both methods of risk assessment in the same population could determine the true difference between the two methods.

Following public reporting of cardiac surgical outcomes over the last decade,³³⁻³⁵ patient risk stratification has been suggested to avoid unfair comparisons

Table 8. Pearson Chi-square Goodness-of-fit Test for Prediction of Mortality and Morbidity with the Tu Classification

Risk Category	n	Mortality				Morbidity			
		No		Yes		No		Yes	
		Observed	Expected	Observed	Expected	Observed	Expected	Observed	Expected
1 (0)	144	143	143.367	1	0.633	130	134.646	14	9.354
2 (1)	194	194	192.682	0	1.318	170	177.172	24	16.828
3 (2)	224	220	221.652	4	2.348	195	198.256	29	25.744
4 (3)	243	241	239.076	2	3.924	202	206.364	41	36.636
5 (4)	216	211	210.644	5	5.356	176	173.812	40	42.188
6 (5)	166	163	159.710	3	6.290	121	124.640	45	41.360
7 (8)	150	141	141.374	9	8.626	104	103.186	46	46.814
8 (7)	105	96	95.933	9	9.067	60	64.804	45	40.196
9 (\geq 8)	106	86	92.462	20	13.538	46	57.359	60	48.641
$\chi^2 = 9.040; df = 9; P = 0.434$					$\chi^2 = 13.313; df = 9; P = 0.149$				

Numbers in parentheses represent the range of risk scores in points for each risk category. n = total number of patients in each risk category; df = degrees of freedom.

Downloaded from: https://www.ash-sci.com/doi/pdf/10.1097/00000539-200102000-00006.pdf by guest on 04 March 2024

Table 9. Accuracy of the CARE Score versus Other Clinical Variables for Prediction of Mortality and Morbidity in the Reference Population

	Area under ROC Curve for Mortality	P Value	Area under ROC Curve for Morbidity	P Value
CARE score	0.786 ± 0.027	—	0.710 ± 0.015	—
ASA physical status	0.723 ± 0.027	0.015	0.646 ± 0.013	< 0.001
NYHA classification for heart failure	0.715 ± 0.034	0.05	0.652 ± 0.014	< 0.001
LVEF	0.681 ± 0.035	0.014	0.588 ± 0.016	< 0.001
Age	0.679 ± 0.033	0.006	0.610 ± 0.015	< 0.001
Serum creatinine	0.662 ± 0.032	< 0.001	0.581 ± 0.012	< 0.001
Operative priority	0.685 ± 0.032	< 0.001	0.603 ± 0.015	< 0.001
Type of surgery	0.655 ± 0.033	< 0.001	0.642 ± 0.014	< 0.001

The P value is for comparisons between the Cardiac Anesthesia Risk Evaluation (CARE) score and each clinical variable. Cutoff points for the various receiver operating characteristic (ROC) curves: left ventricular ejection fraction: ≥ 55%, 45–54%, 30–34%, 20–29%, < 20%; age: 5-yr increment from < 65 to ≥ 80 yr; serum creatinine (μM): < 125, 125–140, 141–167, 168–200, > 200; operative priority: elective, urgent (diagnosis and surgery on same admission), emergency within 24 h, immediate emergency; type of surgery: repair of atrial septal defect, coronary artery bypass graft surgery, single valve surgery, combined or complex procedures, and any of those procedures plus reoperation.

ASA = American Society of Anesthesiologists; NYHA = New York Heart Association; LVEF = left ventricular ejection fraction.

between individual cardiac surgeons and institutions.^{1,36,37} In this context, a risk classification like the CARE score may be discredited because it allows some rater's subjectivity, which may facilitate risk overrating to obtain better risk-adjusted outcome results. This phenomenon, previously called *gaming*,³⁸ is also possible with multifactorial risk indexes when certain risk factors are unavailable before surgery (e.g., left ventricular ejection fraction, pulmonary artery pressure) or when their definition is influenced by practice patterns (e.g., urgency, emergency, use of intraaortic balloon pump) or the technique of measurement (e.g., pulmonary artery pressure, ejection fraction).^{16,38} Furthermore, the use of multifactorial risk indexes usually involve data collection, entry in computers, and individual risk calculation by clerks or research assistants. The potential errors in each step of this process may add distortions to risk calculation. Because no predictive model is perfect, there will always be a risk of making mistakes by analyzing and comparing risk-adjusted outcomes to determine quality of care.³⁹ Recognizing this limitation, all characteristics of the various models must be considered before selecting one that suits an institution or department needs for risk calculation. This study does demonstrate that a model does not have to be complex and free of clinicians' subjective input to provide reliable risk prediction.

One major limitation of this study is that it was performed in a single institution. External validation of the CARE score in other centers will be necessary to confirm its predictive accuracy. Another limitation is the fact that it was tested among a small group of experienced cardiac anesthesiologists only. The predictive performance of the CARE score may be different when used by residents, cardiac surgeons, or intensivists. However, looking at table 1 and the operational criteria in its footnotes, it seems that a CARE score can be assigned to most patients with minimal clinical judgment and experience. The research assistant participating in this study (SG)

had no major difficulties using the score. This suggests that the accuracy of the CARE score would not be altered significantly by the experience of its user. Further studies will be required to confirm this hypothesis.

Currently, multifactorial risk indexes are used mainly by professional and government bodies that produce annual reports on risk-adjusted outcome 1 or 2 yr after the data collection. Many clinicians feel distant from that process, possibly because of its complexity and financial and time requirements. The CARE score is a model proposed to facilitate data collection and interpretation by busy clinicians involved in cardiac surgery. It can easily be added to patient discharge summaries and used by medical record departments to produce frequent risk-adjusted mortality reports. If shown to be an accurate risk predictor in other institutions, the CARE score may be appealing to many clinicians, mainly because it keeps the whole process of risk evaluation at a clinical level.

The authors thank Mrs. Geraldine Wells, Research Division, Department of Anesthesia, University of Ottawa Heart Institute, Ottawa, Ontario, Canada, for her help in preparing this manuscript and Ioulia Doumkina M.D., for reviewing patient charts and assessing the quality of database information.

References

1. Parsonnet V, Dean D, Bernstein AD: A method of uniform stratification of risk for evaluating the results of surgery in acquired adult heart disease. *Circulation* 1989; 79(suppl 1):3–12
2. Tuman KJ, McCarthy RJ, March RJ, Hassan N, Ivankovich AD: Morbidity and duration of ICU stay after cardiac surgery. A model for preoperative risk assessment. *Chest* 1992; 102:36–44
3. Tu JV, Jaglal SB, Naylor D, the Steering Committee of the Provincial Adult Care Network of Ontario: Multicenter validation of a risk index for mortality, intensive care unit stay, and overall hospital length of stay after cardiac surgery. *Circulation* 1995; 91:677–84
4. O'Connor GT, Plume SK, Olmstead EM, Coffin LH, Morton JR, Maloney CT, Nowicki ER, Levy DG, Tryzelaar JF, Hernandez F, Adrian L, Casey KJ, Bundy D, Soule DN, Marrin CAS, Nugent WC, Charlesworth DC, Clough R, Katz S, Leavitt BJ, Wennberg JE, for the Northern New England Cardiovascular Disease Study Group: Multivariate prediction of in-hospital mortality associated with coronary artery bypass graft surgery. *Circulation* 1992; 85:2110–8
5. Higgins TL, Estafanous FG, Loop FD, Beck GJ, Blum JM, Parandhi L: Stratification of morbidity and mortality outcome by preoperative risk factors in coronary artery bypass patients. A clinical severity score. *JAMA* 1992; 267: 2344–8
6. Edwards FH, Clark RE, Schwartz M: Coronary artery bypass grafting: The

Society of Thoracic Surgeons National Database Experience. *Ann Thorac Surg* 1994; 57:12-9

7. Magovern JA, Sakert T, Magovern GJ Jr, Benckart DH, Burkholder JA, Liebler GA, Magovern GJ Sr: A model that predicts morbidity and mortality after coronary artery bypass graft surgery. *J Am Coll Cardiol* 1996; 28:1147-53

8. Nashef SAM, Roques F, Michel P, Gauducheau E, Lemeshow S, Salamon R, the EuroSCORE Study Group: European system for cardiac operative risk evaluation (EuroSCORE). *Eur J Cardio-thorac Surg* 1999; 16:9-13

9. Gabrielle F, Roques F, Michel P, Bernard A, de Vicentis C, Roques X, Brenot R, Baudet E, David M: Is the Parsonnet's score a good predictive score of mortality in coronary artery surgery? Assessment by a French multicentre study. *Eur J Cardiothorac Surg* 1997; 11:406-14

10. Weightman WM, Gibbs NM, Sheminant MR, Thackray NM, Newman AJ: Risk prediction in coronary artery surgery: A comparison of four risk scores. *Med J Aust* 1997; 166:408-11

11. Pons JMV, Espinas JA, Borrás JM, Moreno V, Martín I, Granados A: Cardiac surgical mortality. Comparison among different additive risk-scoring models in a multicenter sample. *Arch Surg* 1998; 133:1053-7

12. Orr RK, Maini BS, Sottile FD, Dumas EM, O'Mara P: A comparison of four severity-adjusted models to predict mortality after coronary artery bypass graft surgery. *Arch Surg* 1995; 130:301-6

13. Martínez-Alario J, Tuesta ID, Plasencia E, Santana M, Mora ML: Mortality prediction in cardiac surgery patients. Comparative performance of Parsonnet and general severity systems. *Circulation* 1999; 99:2378-82

14. Sanchez García R, Nygård E, Christensen JB, Lund JT, Micheelsen F, Niebuhr-Jorgensen U, Bie P: Perioperative risk assessment in connection with heart surgery. Results from 628 patients discussed with emphasis on the difficulties associated with comparison between performances of different departments. *Ugeskr Laeger* 1995; 157:6720-5

15. Leonard RC, van Heerden PV, Power BM, Cameron PD: Validation of Tu's cardiac surgical risk prediction index in a Western Australian population. *Anaesth Intensive Care* 1999; 27:182-4

16. Parsonnet V, Bernstein AD, Gera M: Clinical usefulness of risk-stratified outcome analysis in cardiac surgery in New Jersey. *Ann Thorac Surg* 1996; 61:S8-11

17. Urzua J, Dominguez P, Quiroga M, Moran S, Irarrazaval M, Maturana G, Dubernet J: Preoperative estimation of risk in cardiac surgery. *Anesth Analg* 1981; 60:625-8

18. Jones RH, Hannan EL, Hammermeister KE, DeLong ER, O'Connor GT, Luepker RV, Parsonnet V, Pryor DB, for the Working Group Panel on the Cooperative CABG Database Project: Identification of preoperative variables needed for risk adjustment of short-term mortality after coronary artery bypass graft surgery. *J Am Coll Cardiol* 1996; 28:1478-87

19. Tu JV, Sykora K, Naylor CD, for the Steering Committee of the Cardiac Care Network of Ontario: Assessing the outcomes of coronary artery bypass graft surgery: How many risk factors are enough? *J Am Coll Cardiol* 1997; 30:1317-23

20. Marshall G, Grover FL, Henderson WG, Hammermeister KE: Assessment of predictive models for binary outcomes: An empirical approach using operative death from cardiac surgery. *Statistics in Medicine* 1994; 13:1501-11

21. Pons JMV, Borrás JM, Espinas JA, Moreno V, Cardona M, Granados A: Subjective versus statistical model assessment of mortality risk in open heart surgical procedures. *Ann Thorac Surg* 1999; 67:635-40

22. Tremblay NA, Hardy JF, Perrault J, Carrier M: A simple classification of the risk in cardiac surgery: The first decade. *Can J Anaesth* 1993; 40:103-11

23. Dupuis JY, Wynands JE: Risk-adjusted mortality to assess quality of care in cardiac surgery (editorial). *Can J Anaesth* 1993;38:91-7

24. Swets JA: Measuring the accuracy of diagnostic systems. *Science* 1988; 240:1285-93

25. DeLong ER, DeLong DM, Clarke-Pearson DL: Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics* 1988; 44:837-45

26. Hosmer DW, Lemeshow S: *Assessing the fit of the model, Applied Logistic Regression*. Edited by Hosmer DW, Lemeshow S. New York, Wiley, 1989, pp 135-75.

27. Plackett RL: Karl Pearson and the chi-squared test. *International Statistical Review* 1983; 51:59-72

28. Ferguson TB Jr, Dziuban SW, Edwards FH, Eiken MC, Shroyer ALW, Pailorero PC, Anderson RP, Grover FL: The STS national database: Current changes and challenges for the new millennium. *Ann Thorac Surg* 2000; 69:680-91

29. Wynne-Jones K, Jackson M, Grotte G, Bridgewater B, on behalf of the North West Regional Cardiac Surgery Audit Steering Group: Limitations of the Parsonnet score for measuring risk stratified mortality in the north west of England. *Heart* 2000; 84:71-8

30. Rosner B: *Fundamentals of Biostatistics*, 4th edition. Belmont, Duxbury Press, 1995, pp 419-23

31. Ranta S, Hynynen M, Tammisto T: A survey of the ASA physical status classification: Significant variation in allocation among Finnish anaesthesiologists. *Acta Anaesthesiol Scand* 1997; 41:629-32

32. Haynes SR, Lawler GP: An assessment of the consistency of ASA physical status classification allocation. *Anaesthesia* 1995; 50:195-9

33. Green J, Wintfeld N: Report cards on cardiac surgeons. Assessing New York State's approach. *N Engl J Med* 1995; 332:1229-32

34. Hannan EL, Kilburn H Jr, O'Donnell JF, Lukacik G, Shields EP: Adult open heart surgery in New York State. An analysis of risk factors and hospital mortality rates. *JAMA* 1990; 264:2768-74

35. Williams SV, Nash DB, Goldfarb N: Differences in mortality from coronary artery bypass graft surgery at five teaching hospitals. *JAMA* 1991; 266:810-5

36. Edwards FH, Albus RA, Zajtcuk R, Graeber GM, Barry M: A quality assurance model of operative mortality in coronary artery surgery. *Ann Thorac Surg* 1989; 47:646-9

37. Griffith BP, Hattler BG, Hardesty RL, Kormos RL, Pham SM, Bahnson HT: The need for accurate risk-adjusted measures of outcome in surgery. Lessons learned through coronary artery bypass. *Ann Surg* 1995; 222:593-9

38. Burack JH, Impellizzeri P, Homel P, Cunningham JN Jr: Public reporting of surgical mortality: A survey of New York State Cardiothoracic surgeons. *Ann Thorac Surg* 1999; 68:1195-202

39. Iezzoni LI: The risks of risk adjustment. *JAMA* 1997; 278:1600-7