Multivariate Prediction of In-Hospital Mortality Associated With Coronary Artery Bypass Graft Surgery

Gerald T. O'Connor, PhD, DSc; Stephen K. Plume, MD; Elaine M. Olmstead, BA;
Laurence H. Coffin, MD; Jeremy R. Morton, MD; Christopher T. Maloney, MD;
Edward R. Nowicki, MD; Drew G. Levy, MPH; Joan F. Tryzelaar, MD; Felix Hernandez, MD;
Lawrence Adrian, PA-C; Kevin J. Casey, PA, RN; Dale Bundy, PA, RN; David N. Soule, BA;
Charles A.S. Marrin, MB, BS; William C. Nugent, MD; David C. Charlesworth, MD;
Robert Clough, MD; Saul Katz, MD; Bruce J. Leavitt, MD; and John E. Wennberg, MD, MPH;
for the Northern New England Cardiovascular Disease Study Group

Background. A prospective regional study was conducted to identify factors associated with in-hospital mortality among patients undergoing isolated coronary artery bypass graft surgery (CABG). A prediction rule was developed and validated based on the data collected.

Methods and Results. Data from 3,055 patients were collected from five clinical centers between July 1, 1987, and April 15, 1989. Logistic regression analysis was used to predict the risk of in-hospital mortality. A prediction rule was developed on a training set of data and validated on an independent test set. The metric used to assess the performance of the prediction rule was the area under the relative operating characteristic (ROC) curve. Variables used to construct the regression model included age, sex, body surface area, presence of comorbid disease, history of CABG, left ventricular end-diastolic pressure, ejection fraction score, and priority of surgery. The model significantly predicted the occurrence of in-hospital mortality. The area under the ROC curve obtained from the training set of data was 0.74 (perfect, 1.0). The prediction rule performed well when used on a test set of data (area, 0.76). The correlation between observed and expected numbers of deaths was 0.99.

Conclusions. The prediction rule described in this report was developed using regional data, uses only eight variables, has good performance characteristics, and is easily available to clinicians with access to a microcomputer or programmable calculator. This validated multivariate prediction rule would be useful both to calculate the risk of mortality for an individual patient and to contrast observed and expected mortality rates for an institution or a particular clinician. (Circulation 1992;85:2110–2118)

KEY WORDS • clinical prediction rule • risk estimation • coronary artery bypass graft surgery • mortality

Clinicians have always had to make decisions under conditions of uncertainty. The decision to perform coronary artery bypass graft surgery (CABG) on a patient with angina is an example. As with virtually all medical or surgical interventions, the benefits of CABG must be balanced against its risks. Both patient counseling and optimal clinical decision making require careful assessment of the risks and benefits of alternative strategies. To estimate this risk, the clinician must simultaneously consider many types of information, including characteristics of the patient, characteristics of the disease, results and performance characteristics of hemodynamic tests, and clinical acuity. This can be a daunting task. In addition, a number of biases may contribute to the difficulty of predicting the likelihood of an event.1 Kahnemann et al2 and others3 have demonstrated that subjective probability assessment often is inaccurate.

Quantitative methods that identify factors associated with in-hospital mortality and the integration of this information by clinical prediction rules4 may benefit both clinician and patient. With respect to the short-term outcomes associated with CABG, there are two relevant questions that can be answered by using the methods described in the present study. The patient
asks, "What are my chances?" The answer is intuitively
difficult to estimate yet may be readily and reliably
calculated by these methods. The surgeon asks, "How
am I doing?" That is, "How do the short-term outcomes
of my patients compare with those of other physicians or
to my own results last year?" The comparison of observed
and predicted mortality rates can be readily
achieved using these quantitative methods. Thus, the
reliable and accurate multivariate prediction of risk is
useful to the patient and to the clinician. Accurate
estimation of clinical risk and the study of variation in
this risk among patients and institutions has implications
for patients, clinicians, and the health care system.

It was the intent of the present study to develop a
multivariate clinical prediction rule using logistic regres-
sion analysis, a statistical method that allows the
calculation of a conditional probability of death. A
regional prospective study of 3,055 consecutive iso-
lated CABG procedures performed in northern New
England provided an opportunity to identify patient
and disease factors that are associated with in-hospital
mortality and to develop and validate a prediction rule
to estimate this risk. This rule makes use of data that
are readily available before surgery, and it was vali-
dated using data other than that on which it was
derived. The performance of the prediction rule was
evaluated by comparison of observed and predicted
mortality rates and analysis of the relative operating
characteristic (ROC).

Methods

Data were collected on all patients undergoing iso-
lated CABG in Maine, New Hampshire, and Vermont
between July 1987 and April 1989. Patients undergoing
CABG that was incidental to heart valve repair or
replacement, resection of a ventricular aneurysm, or
other surgical procedure were not included in this
analysis. Participating hospitals verified the number of
patients in the data set and their status at hospital
discharge.

Data were collected on the following variables: patient
age and sex, body surface area (BSA), cardiac catheter-
ization results (degree of left main coronary artery steno-
sis, total number of significantly diseased vessels, left
ventricular end-diastolic pressure [LVEDP], ejection fraction),
priority of surgery (emergency, urgent, or elective),
prior CABG (yes or no), and status at hospital discharge
(dead or alive). The variables chosen were based on a
review of the literature and on the experience of the
clinicians in the research group.

Cardiac catheterizations were performed at the
participating or referring institutions using their own stan-
dard methods during the course of regular clinical care.
Ejection fractions were scored using the method de-
scribed by Pierpont et al.6 Angiography reports were
reviewed to assess the severity of coronary artery dis-
case, expressed as the number of diseased vessels. The
text and diagrams of the reports were read, and the
number of vessels was scored using methods adapted
from the National Heart, Lung, and Blood Institute
Coronary Artery Surgery Study.7

Priority of surgery was assessed by the cardiothoracic
surgeons using definitions previously described.8 Briefly,
emergency means that medical factors relating to the
patient's cardiac disease dictate that surgery should be
performed within hours to prevent morbidity or death,
urgent means that medical factors require the patient to
stay in the hospital for an operation before discharge,
and elective means that medical factors indicate the
need for operation, but the clinical situation allows
discharge from the hospital with readmission at a later
date.

At the time this study was initiated (early 1987), no
comprehensive system for summarizing comorbidity
burden on a patient had been described. The subse-
quent report by Charlson et al9 provided such a mecha-
anism, which has been demonstrated to predict in-
hospital mortality.10 A comorbidity index was compiled
using these methods, as modified for use with hospital
discharge data.11 The variables and their weights (in
parentheses) are as follows: peripheral vascular disease
(1), chronic lung disease (1), dementia (1), chronic liver
disease/cirrhosis (1), (preexisting) peptic ulcer disease
(1), diabetes mellitus with no sequelae (1), diabetes
mellitus with sequelae (e.g., renal, ophthomalic, neuro-
logic, or peripheral circulatory manifestations) (2),
(preexisting) renal failure (2), leukemia, lymphoma, or
solid cancer (2), liver disease with sequelae (e.g., esoph-
ageal varices, portal hypertension, hepatic coma) (3),
and metastatic or multiple cancers (6). The Charlson
index is the summation of these weights (e.g., a patient
with chronic lung disease and preexisting renal failure
would have a comorbidity score of 3). For conditions
that could be either preexisting or a consequence of
surgery (e.g., renal failure or peptic ulcer disease),
individual patient records were reviewed and the disor-
der was considered to be present only if it was present
before surgery. There is no reason to restrict the use of
this comorbidity index to retrospective studies. Prospec-
tive data collected before surgery can be used easily
with this method.

Statistical Methods

Logistic regression analysis was used to assess the
relation between patient and disease characteristics,
and in-hospital mortality and was performed using the
SAS PROC LOGIST program.12,13 The logistic regression
equation that was used to calculate multivariate odds
ratios (OR) and tests of trend predicts the probability of
an event (p) — in this case, in-hospital death —condi-
tional on patient and disease characteristics specified by
the analyst (X1, Xn). A constant (β0) and a series of
weights (βi=1) are determined by the maximum likeli-
hood estimation of the equation. The term "exp" desig-
nates e where "e" is the base of the natural logarithm.
The odds of in-hospital mortality are equal to

\[ \text{exp} \left( \beta_0 + \sum_{i=1}^{n} \beta_i X_i \right) \]

and the predicted probability equals the odds/(1+odds). The logistic regression model may be
expressed as the following function:

\[ p \text{ given } X_1, X_n = \frac{\text{exp} \left( \beta_0 + \sum_{i=1}^{n} \beta_i X_i \right)}{1 + \text{exp} \left( \beta_0 + \sum_{i=1}^{n} \beta_i X_i \right)} \]
When actual patient values are used, the logistic regression equation can be used to calculate the predicted risk of in-hospital mortality for an individual with any combination of demographic or disease characteristics. The likelihood ratio test \( (x^2_{LR}) \) was used to compare nested logistic regression models. Standard statistical methods were used for the calculation of exact binomial confidence intervals.

The entire data set was divided randomly into a "training set" that was used to develop the clinical prediction rule and a "test set" to assess the performance of the rule when applied to data other than those on which it was developed. The dependent variable was in-hospital mortality.

The preferred measure of the resolution of a diagnostic system quantifies accuracy independently of the relative frequency of the events and also is independent of the diagnostic systems detection bias. The ROC, a measure developed in the field of signal detection theory, fulfills these criteria. It is a graphic representation of the relation between the true-positive rate (i.e., the sensitivity of the test) and the false-positive rate (i.e., \( 1 \)-the specificity of the test). The true-positive rate is plotted on the ordinate, and the false-positive rate is plotted on the abscissa. A suitable single-valued measure of test accuracy is the area of the entire graph that lies beneath the curve. This area may vary between 0.5 and 1.0. A useless test would have an area of 0.5, i.e., a positive test would be as likely to be a false-positive as a true-positive at every threshold. A perfect diagnostic test would have an area of 1.0. ROC curves were plotted using the RULE MAKER microcomputer program. Estimates of the area under the ROC curves were calculated using the ROC ANALYZER program. Standard error of the area under the ROC curve was calculated using the nonparametric method described by Hanley and McNeil. Further details regarding the relative operating characteristic may be found in the "Statistical Appendix." The calibration of the prediction rule is assessed by comparing the observed and expected numbers of deaths by category of predicted risk. In addition, the Lemeshow-Hosmer goodness-of-fit statistic \( C \) was calculated.

Data were collected on 3,404 consecutive patients undergoing isolated CABG in Maine, New Hampshire, and Vermont between July 1, 1987, and April 15, 1989. Information on clinical variables, comorbidity, and outcome was available on 3,055 patients. Among these 3,055 patients, there were 132 deaths (4.3%). Coronary angiography results were available on 2,884 patients (94.4%). Ejection fraction and LVEDP measurements were available on 2,790 (91.3%) and 2,377 (77.8%), respectively.

### Results

#### Risk Factors for In-Hospital Mortality

Assessment of predictors of in-hospital mortality was conducted by calculating OR and \( x^2 \) tests; results are summarized in Table 1. To present multivariate odds ratios by category, several continuous variables were categorized for presentation in Table 1. Age was stratified into 5-year age groups. BSA and LVEDP were divided into approximate quartiles. Ejection fraction categories were based on the method described by Pierpont et al. In this multivariate analysis, age was significantly (\( P_{trend} = 0.0001 \)) and positively associated with the risk of in-hospital mortality. The risk for an individual \( \geq 75 \) years old was almost fivefold greater (OR=4.7) than that of an individual less than 55 years old. Female sex was not a significant predictor of in-hospital mortality (OR=1.2, \( p < 0.460 \)). BSA was inversely associated with

### Table 1. Associations Between Patient and Clinical Variables and In-Hospital Mortality

<table>
<thead>
<tr>
<th>Variable</th>
<th>Multivariate analysis</th>
<th>Odds ratio</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient age (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;55</td>
<td></td>
<td>1.0</td>
<td>0.0001</td>
</tr>
<tr>
<td>55–59</td>
<td></td>
<td>1.5</td>
<td>(trend)</td>
</tr>
<tr>
<td>60–64</td>
<td></td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>65–69</td>
<td></td>
<td>2.6</td>
<td></td>
</tr>
<tr>
<td>70–74</td>
<td></td>
<td>3.4</td>
<td></td>
</tr>
<tr>
<td>≥75</td>
<td></td>
<td>4.7</td>
<td></td>
</tr>
<tr>
<td>Patient sex (female vs. male)</td>
<td></td>
<td>1.2</td>
<td>0.460</td>
</tr>
<tr>
<td>Body surface area (m(^2))</td>
<td></td>
<td>1.0</td>
<td>0.009</td>
</tr>
<tr>
<td>≥2.00</td>
<td></td>
<td>1.3</td>
<td>(trend)</td>
</tr>
<tr>
<td>1.80–1.99</td>
<td></td>
<td>1.8</td>
<td></td>
</tr>
<tr>
<td>1.60–1.79</td>
<td></td>
<td>2.4</td>
<td></td>
</tr>
<tr>
<td>&lt;1.60</td>
<td></td>
<td>2.3</td>
<td></td>
</tr>
<tr>
<td>Charlson comorbidity score</td>
<td></td>
<td>1.0</td>
<td>0.002</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td>1.5</td>
<td>(trend)</td>
</tr>
<tr>
<td>≥2</td>
<td></td>
<td>2.3</td>
<td></td>
</tr>
<tr>
<td>Prior coronary artery bypass grafting (yes vs. no)</td>
<td></td>
<td>3.6</td>
<td>0.0001</td>
</tr>
<tr>
<td>Ejection fraction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥60%</td>
<td></td>
<td>1.0</td>
<td>0.013</td>
</tr>
<tr>
<td>50–59%</td>
<td></td>
<td>1.4</td>
<td>(trend)</td>
</tr>
<tr>
<td>40–49%</td>
<td></td>
<td>1.6</td>
<td></td>
</tr>
<tr>
<td>&lt;40%</td>
<td></td>
<td>1.9</td>
<td></td>
</tr>
<tr>
<td>Left ventricular end-diastolic pressure (mm Hg)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤14</td>
<td></td>
<td>1.0</td>
<td>0.005</td>
</tr>
<tr>
<td>15–18</td>
<td></td>
<td>1.3</td>
<td>(trend)</td>
</tr>
<tr>
<td>19–22</td>
<td></td>
<td>1.6</td>
<td></td>
</tr>
<tr>
<td>&gt;22</td>
<td></td>
<td>2.1</td>
<td></td>
</tr>
<tr>
<td>Priority of surgery</td>
<td></td>
<td>1.0</td>
<td>0.001</td>
</tr>
<tr>
<td>Elective</td>
<td></td>
<td>2.1</td>
<td>(trend)</td>
</tr>
<tr>
<td>Urgent</td>
<td></td>
<td>4.4</td>
<td></td>
</tr>
<tr>
<td>Emergency</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left main coronary artery stenosis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;50%</td>
<td></td>
<td>1.0</td>
<td>0.929</td>
</tr>
<tr>
<td>50–89%</td>
<td></td>
<td>1.0</td>
<td>(trend)</td>
</tr>
<tr>
<td>≥90%</td>
<td></td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>No. of diseased coronary vessels</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One</td>
<td></td>
<td>1.0</td>
<td>0.166</td>
</tr>
<tr>
<td>Two</td>
<td></td>
<td>1.3</td>
<td>(trend)</td>
</tr>
<tr>
<td>Three</td>
<td></td>
<td>1.6</td>
<td></td>
</tr>
</tbody>
</table>
risk of in-hospital mortality. Compared with individuals with BSA ≥2.0 m², those with BSA <1.6 have greater than a twofold increase in risk (OR=2.4, p=0.009). Comorbidity, as indicated by a higher score on the Charlson comorbidity index, was associated with increased mortality; patients with a Charlson score ≥2 had more than twice the risk (OR=2.3, p=0.002) of in-hospital mortality than did individuals with a score of 0. In these data, 94.6% of the comorbidity was related to three diagnoses: peripheral vascular disease, chronic obstructive pulmonary disease, and diabetes mellitus. No other diagnoses, with the exception of preexisting renal failure (1.2%) and preexisting peptic ulcer disease (0.8%), were mentioned in more than 0.5% of CABG patients. The prevalence of comorbid conditions and their association with in-hospital mortality are summarized in Table 2.

Individuals with a history of a prior CABG procedure had an approximately 3.5-fold risk (OR=3.6, p=0.0001) of in-hospital mortality compared with patients with no prior CABG procedure.

Cardiac catheterization measurements of ejection fraction and LVEDP were associated with in-hospital mortality. Low ejection fraction was associated with higher rates of in-hospital mortality (p<0.013). An individual with an ejection fraction <40% had a mortality rate approximately twice as high as did an individual with an ejection fraction ≥60% (OR=1.9). LVEDP was positively associated with risk of in-hospital mortality. Individuals with LVEDP >22 had approximately a twofold increase in risk (OR=2.1, p=0.005) compared with those with LVEDP ≤14. Priority of surgery influenced the risk of in-hospital mortality. Emergency or urgent surgery was associated with greater risk than was elective surgery (p<0.001). The risk of urgent surgery was twofold greater than that of elective surgery (OR=2.1), and the risk of emergency surgery was fourfold (OR=4.4) as high as that of elective surgery. In these multivariate analyses, neither the severity of stenosis of the left main coronary artery nor the number of diseased coronary vessels was significantly associated with the risk of in-hospital mortality.

**Development and Validation of Multivariate Prediction Rule**

The data set was randomly divided into a training set and a test set. A clinical prediction rule was developed on the training set using logistic regression analysis. This logistic regression model included patient sex, age, the square root of BSA, a prior CABG procedure, Charlson comorbidity score, ejection fraction score, LVEDP, and priority at surgery. All of these variables were significantly associated with outcome in the multivariate analyses except sex. Gender was not a statistically significant predictor of mortality in the multivariate analyses. However, it was retained in the multivariate prediction rule for two reasons: the patient’s sex is always known, it is “free” information, and whereas the adjusted OR associated with sex is statistically nonsignificant (OR=1.2, 95% confidence intervals [95% CI]=[0.6 to 2.2]), it is not null. Thus, the inclusion of this variable will tend to increase the accuracy of prediction among women.

Parameterization of the variables and the calculation of predicted risk are detailed in Table 3. Coefficients from the logistic regression equation may be used to calculate predicted probability of in-hospital mortality for an individual patient. The dependent variable was in-hospital mortality. The regression model significantly (χ² χ²=57.7, p≤0.0001) predicted the occurrence of in-hospital mortality in this training data set. The addition of angiographic data that included the number of diseased vessels and the percent stenosis of the left main coronary artery did not substantially improve this multivariate prediction model (χ² Lr=2, χ²=0.62, p=0.733).

### Table 2. Prevalence of Comorbid Conditions and the Association of Comorbidity With In-Hospital Mortality

<table>
<thead>
<tr>
<th>Condition</th>
<th>Prevalence (%)</th>
<th>Odds ratio</th>
<th>95% Confidence Intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diabetes</td>
<td>18.1</td>
<td>1.5</td>
<td>1.0, 2.3</td>
</tr>
<tr>
<td>Chronic obstructive pulmonary disease</td>
<td>11.2</td>
<td>1.5</td>
<td>0.9, 2.4</td>
</tr>
<tr>
<td>Peripheral vascular disease</td>
<td>5.5</td>
<td>2.9</td>
<td>1.8, 4.8</td>
</tr>
<tr>
<td>Preexisting renal failure</td>
<td>1.2</td>
<td>5.2</td>
<td>2.5, 11.1</td>
</tr>
<tr>
<td>Preexisting peptic ulcer</td>
<td>0.8</td>
<td>3.2</td>
<td>1.0, 10.2</td>
</tr>
</tbody>
</table>

Dementia, moderate-to-severe liver disease, and various cancers occur in 0.5% or fewer of the cases.

### Table 3. Prediction of Risk of In-Hospital Mortality

<table>
<thead>
<tr>
<th>Variables</th>
<th>Odds</th>
<th>Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Years</td>
<td>0.056</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td>0.278</td>
</tr>
<tr>
<td>0=Male</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1=Female</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BSA</td>
<td>VBSA</td>
<td>-4.021</td>
</tr>
<tr>
<td>Comorbidity score</td>
<td></td>
<td>0.381</td>
</tr>
<tr>
<td>0=0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1=1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2=2 or greater</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prior CABG</td>
<td></td>
<td>1.288</td>
</tr>
<tr>
<td>0=No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1=Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ejection fraction score</td>
<td></td>
<td>0.095</td>
</tr>
<tr>
<td>6=≥60%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10=50–59%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12=40–49%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14=&lt;40%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LVEDP</td>
<td></td>
<td>0.236</td>
</tr>
<tr>
<td>1=≤14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2=15–18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3=19–22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4=&gt;22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Priority at surgery</td>
<td></td>
<td>0.726</td>
</tr>
<tr>
<td>1=Elective</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2=Urgent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3=Emergency</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Constant = -4.374

Using the logistic regression model to predict the risk of in-hospital mortality:

1. Calculate the odds using the patient’s values and the coefficients from the regression equation: Odds = exp(-4.374+(0.056×age(years))+(-4.021×BVS)+(0.381×comorbidity score)+(1.288×prior CABG)+(0.095×ejection fraction score)+(0.236×LVEDP quartile)+(0.726×priority at surgery)).

2. Use the odds to calculate the predicted probability of in-hospital mortality: Probability=odds/(1+odds).
The observed death rates in the test set were compared with the predicted death rates by category of risk (Figure 1). The correlation of observed to expected deaths was 0.99. Table 4 displays observed versus expected mortality rates by risk group along with the exact binomial confidence intervals for the observed mortality rates. The Lemeshow-Hosmer goodness-of-fit statistic, a comparison of observed and expected deaths by decile of risk, was not statistically significant (C_{10} = 7.38, p=0.689), indicating little departure from a perfect fit. ROC curves for the multivariate risk prediction rule were calculated. The performance of the prediction rule on the training set (i.e., the data from which it was derived) and its ability to predict in-hospital mortality on the test set data were compared (Figure 2). The nonparametric estimate of the area under the ROC curve for the training set was 0.74, whereas that on the test set was 0.76.

To study the performance of the prediction rule on subgroups of patients, the data were stratified by sex, age, LVEDP, ejection fraction, priority at surgery, comorbidity, reoperation, and BSA, and the respective areas under the ROC curves were calculated (Table 5). The total area under the ROC curve was 0.76. This was somewhat lower for women (0.66) than for men (0.78). Among the 6% of patients undergoing repeat CABG, the ROC area was 0.60 with a large standard error (0.067); otherwise, it was consistently close to 0.75, indicating that the prediction rule performs well among most subgroups of patients. Elucidating reasons for the poorer performance characteristics of the ROC curve among women and those undergoing reoperation will require further detailed studies with very large sample sizes.

**Table 4. Observed Versus Expected Mortality Rates**

<table>
<thead>
<tr>
<th>Risk category</th>
<th>Expected mortality rate</th>
<th>Observed mortality rate</th>
<th>95% Confidence intervals for observed mortality rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;1.5%</td>
<td>1.0%</td>
<td>0.3%</td>
<td>0.0%, 1.5%</td>
</tr>
<tr>
<td>1.5–2.4%</td>
<td>2.0%</td>
<td>1.0%</td>
<td>0.2%, 2.8%</td>
</tr>
<tr>
<td>2.5–3.4%</td>
<td>3.0%</td>
<td>3.6%</td>
<td>1.6%, 6.9%</td>
</tr>
<tr>
<td>3.5–6.4%</td>
<td>4.8%</td>
<td>4.5%</td>
<td>2.5%, 7.4%</td>
</tr>
<tr>
<td>≥6.5%</td>
<td>12.0%</td>
<td>12.5%</td>
<td>9.1%, 16.7%</td>
</tr>
</tbody>
</table>

n=1,516.

Discussion

In this regional prospective study, a multivariate clinical prediction rule was developed and validated. This mathematical model is a highly significant predic-

![Graph](https://example.com/graph.png)

**Figure 1.** Observed versus expected mortality by predicted risk category among those in the test set. Prediction rule was developed from a model based on patients in the training set (n=1,539) and then applied to patients in the test set (n=1,516). Risk categories were <1.5%, 1.5–2.4%, 2.5–3.4%, 3.5–6.4%, and ≥6.5%.

tor of the dependent variable, in-hospital mortality. This statistical method calculates the independent weight associated with each of the variables and allows a combination of this information in a quantitative manner that would be difficult or impossible without this statistical tool. The rule uses patient and clinical data that are available before surgery. The prediction rule had relatively good performance characteristics (area under ROC curve, 0.76), and performs equally well when used to predict outcomes on a test set of data and within clinically relevant subgroups of patients.

A question that arises from these data is what is responsible for the imprecision in this clinical prediction rule (the remaining 0.24 of the area under the ROC curve)? In our view, at least four factors contribute. First, when variables such as ejection fraction are measured, there is a certain error that may be inherent in the measurement. If this error is random, it will create a bias toward the null when the variable is used to predict an outcome. When there are multiple measured variables in a multivariate model, the cumulative error tends to bias the entire model toward the null and decrease the area under the ROC curve.

Second, there are actual differences in outcome associated with the individual medical centers and surgeons. This has been demonstrated to be substantial in these data but is not itself responsible for the entire difference between perfect and achieved prediction.

Third, unmeasured variables are likely to contribute to imprecision in this prediction model that describes risk using eight variables. The magnitude of this effect is difficult to determine. An indication that the effect is small in these data is the relatively small remaining area under the ROC curve.

Last, clearly, prediction is optimal with complete data. Yet, in real settings, data are not always available in a timely fashion. With respect to ejection fraction, Pierpont et al suggest substitution of an ejection fraction over 60 when ejection fraction is not available. This would obviously cause an underestimate of risk if the patient had a poor ejection fraction. In our data, those with missing ejection fraction had a mortality rate nearly identical to those with ejection fraction over 60; substitution of normal ejection fraction score for missing values seemed justified for our analysis. If the ejection fraction could be estimated from the ventriculogram, this value could be substituted for missing data. With respect to LVEDP, we have chosen to substitute the median value for missing data because, in our data set, those with missing data for LVEDP showed a mortality rate similar to those with midrange LVEDP.
values. We do not believe that this creates any substantial bias in prediction although exceptions could be found.

The multivariate estimation of risk of in-hospital or perioperative mortality has been studied by several investigators during the past two decades. In 1975, Loop and coauthors published a multivariate comparison of 1,188 survivors (operated on during 1973) with 60 patients who died from cardiac-related causes associated with CABG operations (operated on during 1967 through 1973). Using discriminant analysis, they identified marked cardiomegaly, uncompensated congestive heart failure, three-vessel disease and/or obstruction of the left main coronary artery, and elevated LVEDP as independent risk factors. Coefficients, their standard errors, and examples of calculations were given.21

Using the techniques of multivariate discriminant analysis as applied to data from the National Heart, Lung, and Blood Institute Coronary Artery Surgery Study (CASS), Kennedy et al22 described a multivariate prediction rule. Variables that were independent predictors of risk of perioperative mortality among 6,176 patients undergoing isolated CABG operations were increasing age, left main coronary stenosis ≥90%, female sex, left ventricular wall motion score, and the presence of rales. Regression coefficients were presented along with calculations of risk for individual patients. No comorbidity variables were included. There was a comparison of the observed and expected mortality rates at the 15 CASS sites.22 Junod et al23 also used the multivariate prediction rule developed by the CASS study to calculate the predicted mortality on 913 patients undergoing isolated primary CABG and compared observed with expected outcomes.

Wright et al24 studied 50 clinical and angiographic variables collected from 6,257 patients undergoing CABG between 1970 and 1984 at the Loyola University Medical Center. The analysis was based on 11 years of retrospective and 3 years of prospective data collection. For isolated CABG, those variables associated with perioperative mortality included age, severity of coronary heart disease, the presence of diffuse coronary heart disease, the number of coronary arteries bypassed, and a family history of coronary heart disease. The resulting multivariate prediction rule was then applied to the population from which it was derived. The coefficients used in the regression model are not reported.

Edwards et al25 developed a prediction model based on 20 variables using a Bayesian analysis of data from 300 patients undergoing isolated CABG at their institution. This model then was used to predict perioperative mortality on 400 additional patients. The model had face validity, evidenced by a positive association between predicted and observed mortality. The conditional probability matrix is reported but the Bayesian algorithms that are required for calculation are not revealed. In another report from the same research group, a Bayesian model was derived from the CASS registry data. Variables including age, sex, ventricular function, previous myocardial infarction, extent of coronary artery disease, unstable angina and surgical priority were used to calculate retrospectively the expected mortality for a group of 840 patients at their institution. The report does not describe the actual equations used in calculating the risk of perioperative mortality.26

Based on retrospective data collection on 3,500 consecutive open heart patients (CABG and repair or replacement of mitral and/or aortic valves), Parsonnet et al27 developed two multivariate prediction models: one additive using multiple regression analysis and another using logistic regression. The logistic regression model consisted of 17 variables: age, aortic valve disease, bypass only, bypass plus other procedure, elevated cholesterol, diabetes, catastrophic states (e.g., acute structural failure), family history of coronary artery disease, female sex, hypertension, left ventricular aneurysm, left ventricular ejection fraction, mitral valve disease, obesity, preoperative intra-aortic balloon pump, reoperation, and smoking. Five of these variables (aortic valve disease, elevated cholesterol, family history, left ventricular aneurysm, and smoking) were not statistically significant but were retained in the model. The additive model consisted of 14 variables, including female sex, morbid obesity, diabetes (unspecified type), hypertension, ejection fraction, reoperation, preoperative intra-aortic balloon pump placement, left ventricular aneurysm, emergency surgery after percutaneous transluminal coronary angioplasty or catheterization complications, dialysis dependency, "other rare circumstances" (e.g., paraplegia, pacemaker dependency, congenital heart disease in an adult, severe asthma), mitral valve pulmonary artery pressure ≥60 mm Hg, and aortic valve pressure gradient >120 mm Hg. These prediction models were tested prospectively on independent data sets and were found to be highly correlated with the observed death rates.
The use of multivariate analysis facilitates the distillation of multiple types of information into a single and reproducible metric for risk. The validated clinical prediction rule described in this report was developed specifically to predict the risk of in-hospital mortality associated with isolated CABG. Although local validation should precede actual clinical use of the prediction model, this multivariate model has several desirable characteristics. Contemporaneous regional data were obtained in this prospective study. The prediction rule was developed using statistically appropriate methods including logistic regression analysis and the area under the ROC curve as a metric for performance. The prediction rule deals in a straightforward manner with comorbidity and uses a total of only eight variables, all of which usually are available to the clinician before CABG surgery. Overall, the performance characteristics, accuracy and reproducibility, have been shown to be relatively good, and the calculations may be made easily using either a microcomputer or a programmable calculator. This prediction rule allows the rapid and accurate quantification of risk of in-hospital mortality and may be useful for patient counseling and quality assurance studies and to augment clinical judgment.

Statistical Appendix: The ROC

Assessing the accuracy of a multivariate prediction rule is not a trivial problem. The prediction equation yields a conditional probability for each individual patient. This is a number between 0 and 1.0 that expresses the estimated likelihood of the outcome of interest based on the supplied values for each of the independent variables. However, each patient either does or does not experience the outcome of interest. Thus, a direct comparison of predicted and observed outcomes for each individual patient is not useful as a performance metric.

The ROC, a technique developed in the field of signal detection theory, may be the optimal performance metric for multivariate prediction rules. The use of this technique in assessing the accuracy of diagnostic systems has recently been reviewed by Swets.28 The ROC is a graphic technique, a plot of the true-positive rate (sensitivity) versus the false-positive rate (1-specificity) evaluated at a number of cutoff points or decision thresholds (Figure 3). The area under the ROC becomes a suitable single number summary of the diagnostic accuracy of the prediction rule. A useless test, such as a coin flip, would yield an area of 0.5 (i.e., a positive test would be equally as likely to be a false-positive as a true-positive). A perfect diagnostic test would have an area under the ROC of 1.0. According to Swets, areas “between 0.5 and 0.7 or so represent a rather low accuracy—the true-positive proportion is not much greater than the false-positive proportion. Values of A (area) between 0.7 and 0.9 are useful for some purposes, and higher values represent a rather high accuracy.”

The ROC has a number of desirable characteristics as a performance metric for multivariate prediction rules. The area under the ROC is independent of the relative frequencies of the events (e.g., death rate). It also is unaffected by the diagnostic system’s decision biases or decision thresholds. Last, it allows the comparison of different diagnostic systems by putting them on a common scale.

Further, this measure of performance often is readily available. The C statistic (also referred to as the concordance index), which is often calculated by many

\[
\text{C statistic} = \frac{\text{true positive rate} - \text{false positive rate}}{2}
\]

The C statistic is the area under the ROC curve, which can be interpreted as the probability that the sample score of a randomly chosen positive case will be greater than the score of a randomly chosen negative case. Values of the C statistic range from 0.0 to 1.0, with higher values indicating better discrimination. A C statistic of 0.5 indicates a useless test, while a C statistic of 1.0 indicates a perfect test.

\[
\text{C statistic} = \frac{\text{true positive rate} - \text{false positive rate}}{2}
\]

The C statistic is the area under the ROC curve, which can be interpreted as the probability that the sample score of a randomly chosen positive case will be greater than the score of a randomly chosen negative case. Values of the C statistic range from 0.0 to 1.0, with higher values indicating better discrimination. A C statistic of 0.5 indicates a useless test, while a C statistic of 1.0 indicates a perfect test.

\[
\text{C statistic} = \frac{\text{true positive rate} - \text{false positive rate}}{2}
\]

The C statistic is the area under the ROC curve, which can be interpreted as the probability that the sample score of a randomly chosen positive case will be greater than the score of a randomly chosen negative case. Values of the C statistic range from 0.0 to 1.0, with higher values indicating better discrimination. A C statistic of 0.5 indicates a useless test, while a C statistic of 1.0 indicates a perfect test.

\[
\text{C statistic} = \frac{\text{true positive rate} - \text{false positive rate}}{2}
\]

The C statistic is the area under the ROC curve, which can be interpreted as the probability that the sample score of a randomly chosen positive case will be greater than the score of a randomly chosen negative case. Values of the C statistic range from 0.0 to 1.0, with higher values indicating better discrimination. A C statistic of 0.5 indicates a useless test, while a C statistic of 1.0 indicates a perfect test.

\[
\text{C statistic} = \frac{\text{true positive rate} - \text{false positive rate}}{2}
\]

The C statistic is the area under the ROC curve, which can be interpreted as the probability that the sample score of a randomly chosen positive case will be greater than the score of a randomly chosen negative case. Values of the C statistic range from 0.0 to 1.0, with higher values indicating better discrimination. A C statistic of 0.5 indicates a useless test, while a C statistic of 1.0 indicates a perfect test.

\[
\text{C statistic} = \frac{\text{true positive rate} - \text{false positive rate}}{2}
\]

The C statistic is the area under the ROC curve, which can be interpreted as the probability that the sample score of a randomly chosen positive case will be greater than the score of a randomly chosen negative case. Values of the C statistic range from 0.0 to 1.0, with higher values indicating better discrimination. A C statistic of 0.5 indicates a useless test, while a C statistic of 1.0 indicates a perfect test.

\[
\text{C statistic} = \frac{\text{true positive rate} - \text{false positive rate}}{2}
\]

The C statistic is the area under the ROC curve, which can be interpreted as the probability that the sample score of a randomly chosen positive case will be greater than the score of a randomly chosen negative case. Values of the C statistic range from 0.0 to 1.0, with higher values indicating better discrimination. A C statistic of 0.5 indicates a useless test, while a C statistic of 1.0 indicates a perfect test.
logistic regression programs such as SAS PROC LOGIST, has been shown by Hanley and McNeill to be identical to the area under the ROC.

Appendix

Northern New England Cardiovascular Disease Study Group

Catholic Medical Center, Manchester, N.H.

Christopher T. Maloney, MD (Institutional Representative); Paula Banahan, RN; David C. Charlesworth, MD; Robert C. Dewey, MD; J. Beatty Hunter, MD; Donna Pulsifer, RN; Benjamin M. Westbrook, MD.

Dartmouth-Hitchcock Medical Center, Lebanon, N.H.

Stephen K. Plume, MD (Institutional Representative); Gordon Defoe, CCP; Eugenia Hamilton, MSHA; Douglas James, MD; Jo Ann Kairys, MSPH; Helen M. Lemal, BS, RCP; Betsy Maislen, RN; Charles A.S. Marrin, MB, BS; William C. Nugent, MD; John F. Robb, MD; William Schults; Jon Wahrnerberger, MD; John H. Wasson, MD; John E. Wennberg, MD; Dorothy Williams.

Data Coordinating Center, Lebanon, N.H.

Gerald T. O’Connor, PhD, DSc (Director); Patricia S. Johnson, BA, RN; Michael Diehl, MD; Joseph F. Kasper, ScD; Terry Kneeland, MPH; Drew G. Levy, MPH; David Malenka, MD; Lynn L. Moore, MPH; Elaine M. Olimstead, BA.

Eastern Maine Medical Center, Bangor

Robert Clough, MD; Felix Hernandez, MD (Institutional Representatives); William Crouss, PA; Robert M. Hoffman, MD; Gordon Schaedel, PA.

Elliot Hospital, Manchester, N.H.

Michael J. Hearne, MD.

Maine Medical Center, Portland

Jeremy R. Morton, MD (Institutional Representative); Lawrence Adrian, PA-C; Eric Anderson; Michael Brennan, PA-C; Desmond Donegan, MD; Lee Ann Genthen, CORT; Clement A. Hiebert, MD; Saul Katz, MD; Mirle A. Kellett Jr., MD; Robert Kramer, MD; Costas T. Lambrew, MD; Jean Letourneau; Chris A. Lutes, MD; Edward R. Nowicki, MD; Russell Stodgill, PA; Kim Tierney, RN; Joan F. Tryzelaar, MD; Richard L. White, MD.

Maine Medical Assessment Program, Augusta

David N. Soule, BA; Daniel F. Hanley, MD, DSc.

Medical Center Hospital of Vermont, Burlington

Laurence H. Coffin, MD (Institutional Representative); Dale Bundy, RN, PA; Kevin Casey, RN, PA; Roy V. Ditchez, MD; Frank Harris; Frank P. Ittleman, MD; Richard S. Jackson, MD; Bruce J. Leavitt, MD; Dean Lee; Diane Pappalardo.

Portsmouth Hospital, Portsmouth, N.H.

James Schmitz, MD (Institutional Representative); Amy Coombs, PA; John R. O’Meara, MD.

References


_Circulation_. 1992;85:2110-2118
doi: 10.1161/01.CIR.85.6.2110

_Circulation_ is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 1992 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7322. Online ISSN: 1524-4539

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circ.ahajournals.org/content/85/6/2110

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in _Circulation_ can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to _Circulation_ is online at:
http://circ.ahajournals.org//subscriptions/