TABLE 1. Corrected Regression Data: Plasma Endothelin-1 (Ordinate) Versus Systemic and Pulmonary Hemodynamics (Abscissa)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Regression</th>
<th>r</th>
<th>r²</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heart rate</td>
<td>Linear</td>
<td>0.22</td>
<td>0.05</td>
<td>NS</td>
</tr>
<tr>
<td>Arterial pressure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Systolic</td>
<td>Linear</td>
<td>0.09</td>
<td>0.008</td>
<td>NS</td>
</tr>
<tr>
<td>Diastolic</td>
<td>Linear</td>
<td>0.12</td>
<td>0.015</td>
<td>NS</td>
</tr>
<tr>
<td>Mean</td>
<td>Linear</td>
<td>0.08</td>
<td>0.006</td>
<td>NS</td>
</tr>
<tr>
<td>Pulmonary pressure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Systolic</td>
<td>Polynomial</td>
<td>0.72</td>
<td>0.52</td>
<td>&lt;0.002</td>
</tr>
<tr>
<td>Diastolic</td>
<td>Polynomial</td>
<td>0.80</td>
<td>0.65</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Mean</td>
<td>Polynomial</td>
<td>0.79</td>
<td>0.62</td>
<td>&lt;0.0003</td>
</tr>
<tr>
<td>Right atrial pressure</td>
<td>Linear</td>
<td>0.53</td>
<td>0.28</td>
<td>&lt;0.02</td>
</tr>
<tr>
<td>Pulmonary wedge pressure</td>
<td>Linear</td>
<td>0.4</td>
<td>0.16</td>
<td>NS</td>
</tr>
<tr>
<td>Cardiac index</td>
<td>Linear</td>
<td>0.28</td>
<td>0.08</td>
<td>NS</td>
</tr>
<tr>
<td>Stroke volume index</td>
<td>Linear</td>
<td>0.38</td>
<td>0.14</td>
<td>NS</td>
</tr>
<tr>
<td>Systemic vascular resistance</td>
<td>Linear</td>
<td>0.05</td>
<td>0.003</td>
<td>NS</td>
</tr>
<tr>
<td>Pulmonary vascular resistance</td>
<td>Polynomial</td>
<td>0.68</td>
<td>0.46</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Resistance ratio</td>
<td>Polynomial</td>
<td>0.82</td>
<td>0.67</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

The stepwise multiple regression analysis now actually provides greater evidence of the selectivity of the relation of plasma endothelin to pulmonary hemodynamics. The only variables contributing to the relation are mean pulmonary artery pressure and pulmonary wedge pressure: the numerator of the resistance ratio. With the corrected endothelin values, the previous small contribution of mean arterial pressure was rejected in the current multiple regression computation. These correlations were not significantly altered by the error in four of 20 heart failure samples, since three of the four erroneous values were at the low end of the endothelin range. We are therefore able to provide a more accurate range of plasma endothelin values in normal subjects. It should be noted that this is the only peer-reviewed publication from our group that is affected by this error.

Robert J. Cody, MD
Division of Cardiology
The Ohio State University Hospitals
Columbus, Ohio

Reference

Risky Business: Prospective Applicability of Models

Recently, O'Connor et al developed a multivariable model to predict in-hospital mortality associated with coronary artery bypass graft surgery based on a small set of descriptors that "...are available to the clinician before...surgery" and concluded that the model "...may be useful for patient counseling." Although the authors do not say so, some readers might interpret these statements to mean that this model can be used to help them decide who to refer for surgery. If so, several matters relating to the prospective applicability of the model are deserving of comment.

The authors necessarily developed their model retrospectively in patients who had already been referred to surgery. For this reason, the model cannot be used prospectively to help decide who should be so referred. To see why, imagine that we know that low left ventricular ejection fraction is associated with high surgical risk, and we thereby tend not to refer such patients for surgery. As a consequence of good clinical judgment, (relatively) high-risk patients exhibiting (relatively) low ejection fractions will have been selected out of the study population. A retrospective analysis that includes ejection fraction as one of the candidate predictor variables will conclude incorrectly that it is not an important prospective predictor of risk. If, on the other hand, ejection fraction is excluded from the retrospective analysis because we know it has been distorted by selection bias, then the resultant prediction model will be irrelevant because it fails to consider a predictor of known prospective importance. Any such model is either incorrect or irrelevant.

The authors properly assessed the accuracy of their model both in terms of its resolution (receiver-operating characteristic curve area) and its calibration (goodness-of-fit). They observed a 76% resolution for their prediction model and speculated at some length on the reasons for the residual 24% loss. There is, however, a much simpler explanation for this loss. A prediction model cannot be 100% accurate in terms of both resolution and calibration; in fact, a perfectly calibrated model will have a resolution of only 83%. Because the logistic regression algorithm tends to optimize calibration rather than resolution, the authors' observation of imperfect resolution is not at all surprising.

This trade-off of resolution for calibration can be considered entirely appropriate: Calibration is more relevant to individual (clinical) decisions, whereas resolution is more relevant to group (epidemiological) decisions. Predicting the average risk for a group (knowing how many will die) is very different from predicting the specific risk for an individual member of that group (knowing who will die). Accordingly, risk predictions must be...
precise as well as accurate if they are to be applicable to the care of individual patients. Unfortunately, the predictions are often very imprecise. Thus, Work et al. documented a highly significant inverse correlation between resting left ventricular ejection fraction and subsequent risk of coronary events among 646 postinfarction patients and developed a logistic model to predict that risk. Although the model accurately predicted the average risk of groups defined by different levels of ejection fraction, the 95% confidence interval of risk for an individual patient with a depressed ejection fraction of 0.20 was very wide (11–36%) and overlapped that for a patient with a normal ejection fraction of 0.60 (0–14%).

Although the authors say their model should undergo “local validation” before its clinical use, few potential users will actually follow this sage advice. Many will simply choose not to use the model at all (few prediction models, unfortunately, have diffused into routine clinical practice), whereas others will embrace it on the basis of the current validation study alone. Unfortunately, the transportability of a logistic prediction model such as this (its ability to perform as well at sites remote from that of its origin) is very sensitive to differences in the prevalence of the predicted outcome (in this case, mortality) between the local and remote sites. Accordingly, the authors should assess the transportability of their model by reporting the range of mortality rates at different study sites (e.g., state by state or hospital by hospital).

If the mortality rate at a “remote” site is materially different from that at the “local” site, the prediction model can be modified to account for these differences. The local mortality rate in this study was 4.3% (132 of 3,055). Suppose the model will be used at a remote site where we know the mortality rate averages 10%. If we do not adjust for this difference, the local prediction model will systematically underestimate the risk at the remote site. Fortunately, we can adjust the model’s logistic intercept (what the authors denoted as β0) to account for this difference. If we convert each mortality rate μ into its equivalent odds Ω = μ/(1 − μ), then the adjusted logistic intercept α0 = ln(Ωremote/Ωlocal) is the sum of the unadjusted intercept β0 + ln(Ωremote/Ωlocal) and the logarithm of the ratio of these odds: α0 = β0 + ln(Ωremote/Ωlocal).

The comorbidity index used in this study was originally developed in a hospitalized general internal medicine population and was prospectively “validated” over 1 year in a population of women with breast cancer. How applicable is such a model to the in-hospital mortality of patients undergoing coronary artery bypass surgery? The authors might instead consider developing a more specific comorbidity index from their own data. Prediction models such as the one proposed by O’Connor et al promise to improve the quality and cost of health care 1) by placing our limited personal experience into broader perspective through comparison to a larger repository of clinically relevant information, 2) by making explicit the assumptions implied by our decisions, and 3) by alerting us whenever our decisions do not appear to be consistent with these assumptions, with the available information, or with the conventional rules of logic. The models must be prospectively applicable if this promise is to be realized. Nobel Laureate Manfrid Eigen once remarked that a scientific theory only risks being wrong, whereas a model might be right but irrelevant.

George A. Diamond, MD
Timothy A. Denton, MD
Division of Cardiology
Cedars–Sinai Medical Center
Los Angeles, California


Reply

Diamond and Denton have commented on five points in our recent report of the development and validation of a multivariate prediction rule,1 namely that 1) the rule is validated only on patients who actually have surgery; 2) there are upper limits of the predictive ability of logistic regression models; 3) there are limits in the accuracy of prediction rules; 4) prediction rules should be recalibrated to reflect observed local rates of mortality; and 5) they suggest that we might consider developing a measure of comorbidity that is specific for patients undergoing coronary artery bypass graft surgery (CABG).

It is true that a multivariate rule predicting the risk of operative mortality is developed only on those who have surgery. Because of this, there is concern regarding the generalizability of the results of prediction rules. For example, individuals with very poor prognoses may not be referred for surgery, and thus the sample on which the prediction rule is developed and validated may differ in some substantial manner from the population potentially eligible for surgery. This effect is difficult to evaluate, since those not referred are obviously not included in the study. However, the 3,055 patients evaluated represent individuals with a wide spectrum of clinical severity. They range in age from 25 to 89 years old; >25% are ≥70 years old. Their ejection fractions vary from 14% to 96%; 10% of patients in this data set have ejection fractions <40%.

It is true that the rule cannot be validated among those who do not have surgery. The actual risk of operative mortality among people who are potentially eligible for surgery but elect not to have it is, of course, unknowable. The most relevant estimate of the risk of operative mortality that might be experienced by an individual patient considering surgery is the experience of others with similar characteristics who have actually undergone surgery. Thus, although in an absolute sense the use of the multivariate prediction rule for the counseling of patients considering surgery cannot be validated, these estimates may still be the best context-specific estimates available.

We agree with Diamond and Denton that there are limits on the predictive ability of regression models. There are practical upper limits, and there may be theoretical limits, to the predictive ability of multivariate models. We have documented physician and institutional factors2 that substantially affect the outcomes of CABG, and if these are not included in a prediction rule, it will not predict optimally. In addition, measurement errors and unmeasured variables also contribute to the difference between achievable and achieved accuracy in any mathematical model.

We recognize their agreement with our point that predictive models must be modified to suit local circumstances. In our regional study of CABG, we have done just that by including a categorical variable for each medical center. In the method suggested by Diamond and Denton, the intercept term is altered to adjust the results of the prediction model on the basis of the local mortality rate. This method is mathematically simple but is valid only if there is no effect modification by site; that is, the

References
1. O’Connor GT, Plume SK, Olmstead EM, Coffin LH, Morton JR, Maloney CT, Nowicki ER, Levy DG, Tryzelaar JF, Hernandez F,
Risky business: prospective applicability of models.
G A Diamond and T A Denton

Circulation. 1993;87:1054-1056
doi: 10.1161/01.CIR.87.3.1054

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/87/3/1054.citation

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/