

An Externally Validated Model for Predicting Long-Term Survival after Exercise Treadmill Testing in Patients with Suspected Coronary Artery Disease and a Normal Electrocardiogram

Michael S. Lauer, MD; Claire E. Pothier, MPH; David J. Magid, MD, MPH; S. Scott Smith, MD; and Michael W. Kattan, PhD

Background: The exercise treadmill test is recommended for risk stratification among patients with intermediate to high pretest probability of coronary artery disease. Posttest risk stratification is based on the Duke treadmill score, which includes only functional capacity and measures of ischemia.

Objective: To develop and externally validate a post-treadmill test, multivariable mortality prediction rule for adults with suspected coronary artery disease and normal electrocardiograms.

Design: Prospective cohort study conducted from September 1990 to May 2004.

Setting: Exercise treadmill laboratories in a major medical center (derivation set) and a separate HMO (validation set).

Patients: 33 268 patients in the derivation set and 5821 in the validation set. All patients had normal electrocardiograms and were referred for evaluation of suspected coronary artery disease.

Measurements: The derivation set patients were followed for a median of 6.2 years. A nomogram-illustrated model was derived on the basis of variables easily obtained in the stress laboratory, including age; sex; history of smoking, hypertension, diabetes, or typical angina; and exercise findings of functional capacity, ST-segment changes, symptoms, heart rate recovery, and frequent ventricular ectopy in recovery.

Results: The derivation data set included 1619 deaths. Although both the Duke treadmill score and our nomogram-illustrated model were significantly associated with death ($P < 0.001$), the nomogram was better at discrimination (concordance index for right-censored data, 0.83 vs. 0.73) and calibration. We reclassified many patients with intermediate- to high-risk Duke treadmill scores as low risk on the basis of the nomogram. The model also predicted 3-year mortality rates well in the validation set: Based on an optimal cut-point for a negative predictive value of 0.97, derivation and validation rates were, respectively, 1.7% and 2.5% below the cut-point and 25% and 29% above the cut-point.

Limitations: Blood test-based measures or left ventricular ejection fraction were not included. The nomogram can be applied only to patients with a normal electrocardiogram. Clinical utility remains to be tested.

Conclusion: A simple nomogram based on easily obtained pretest and exercise test variables predicted all-cause mortality in adults with suspected coronary artery disease and normal electrocardiograms.

Ann Intern Med. 2007;147:821-828.

www.annals.org

For author affiliations, see end of text.

Exercise treadmill testing is recommended for assessing prognosis in patients with symptoms suggestive of coronary artery disease and at least an intermediate probability for disease, as assessed by age, sex, and symptoms (1). Current guidelines focus on the Duke treadmill score, which incorporates functional capacity, electrocardiographic evidence of ischemia, and test-induced angina pectoris (2–4). However, this scheme of risk stratification based primarily on age, sex, symptom status, and the Duke treadmill score does not take into account other simple historical measures that have known diagnostic and prognostic value (5, 6), as well as more recently discovered exercise test predictors of risk (7), such as heart rate recovery (8–11) and stress-related ventricular ectopy (12).

Because combining clinical and exercise variables improves prognostic accuracy (13–15), we developed a practical prognostic model for all-cause mortality based on measures that are easily obtained at the time of stress testing and do not require blood testing. We then compared this model with the current standard Duke treadmill score and externally validated the model for its ability to accurately identify low-risk persons for whom conservative management would be appropriate. We focused on all-cause mortality as an end point because, unlike cardiac

mortality, it is wholly unbiased, objective, and clinically relevant (16, 17). For ease of use, we present this model as a nomogram.

METHODS

Patients

Since September 1990, all patients undergoing any type of stress testing at our institution have had demographic, clinical, and stress data entered prospectively into a computerized database as part of routine clinical reporting (8, 9, 12). For our analysis, we focused on adult patients 30 years of age or older who were referred for symptom-limited exercise treadmill testing for evaluation of

See also:

Print

Editors' Notes 822

Web-Only

CME quiz

Conversion of graphics into slides

Audio summary

Context

How might we best stratify low-risk patients with normal electrocardiograms but suspected coronary artery disease?

Contribution

This large prospective study evaluated a post-treadmill test mortality prediction rule in adults with normal electrocardiograms and suspected coronary artery disease. A nomogram that included clinical variables (age, sex, smoking, and diabetes) and treadmill variables (exercise capacity, abnormal heart rate recovery, and ventricular ectopy during recovery) better identified patients with less than a 3% mortality rate over 3 years than did a standard Duke treadmill score.

Implication

The nomogram warrants additional validation to verify that it reliably identifies low-risk patients who need no further cardiac work-up after a treadmill test.

—The Editors

suspected coronary disease between 1 September 1990 and 31 December 2002. None had known coronary disease, heart failure, documented left ventricular systolic dysfunction, any type of cardiomyopathy, valvular or congenital heart disease, previous renal or cardiac transplantation, atrial fibrillation, digitalis use, pacemaker or defibrillator placement, or end-stage renal disease. Furthermore, none had any baseline electrocardiographic abnormalities, such as Q-wave myocardial infarction, right or left bundle-branch block, left ventricular hypertrophy with accompanying repolarization abnormalities, preexcitation syndrome, or resting ST-T wave changes precluding exercise interpretation (defined as ≥ 1 mm of horizontal or down-sloping ST-segment depression 80 ms after the J point). We focused on patients with normal electrocardiograms to allow for ST-segment interpretation and to minimize the likelihood of undiagnosed left ventricular systolic dysfunction (18). Our institutional review board granted permission to use electronically obtained clinical data for research purposes and waived the requirement for informed consent.

Clinical Data

As described elsewhere, all patients underwent a routine structured chart review and interview with prospective data entry of demographic and clinical data before exercise testing (8, 9, 12, 19). For our analysis, we focused on age, sex, smoking history, diabetes mellitus, hypertension, and history of typical angina (5, 6)—variables that can be easily obtained in a stress laboratory without blood tests or access to medical records and have been correlated elsewhere with severe angiographic coronary disease and death (5, 6). We considered patients who reported regular use of cigarettes any time within the past year to be recent or current smok-

ers. We classified diabetes on the basis of whether it was being treated regularly with insulin or treated with oral agents only. If resting systolic blood pressure was at least 140 mm Hg, resting diastolic blood pressures was at least 90 mm Hg, or medications were regularly used for treating elevated blood pressure, we considered hypertension to be present. We considered typical angina present if, on reporting a history of chest pain or discomfort, patients answered affirmatively to all 3 of the following questions: “Is the pain or discomfort substernal?” “Is the pain or discomfort brought on by physical or emotional stress?” and “Is the pain quickly relieved by rest and/or use of nitroglycerin?” We did not consider other symptoms, such as atypical angina, dyspnea, or nonanginal chest pain, that were not routinely recorded before 1 January 1996 (20). Trained personnel prospectively recorded all clinical data online before exercise testing; they were blinded to both the hypothesis of this study and the patient outcomes.

Exercise Testing

We performed symptom-limited exercise testing mostly by using the Bruce and modified Bruce protocols; protocol selection was based on aiming for a test that would last 8 to 12 minutes. We estimated exercise capacity in metabolic equivalents, where 1 metabolic equivalent is an oxygen consumption of 3.5 mL per kg of body weight per minute, and calculated the age- and sex-adjusted proportion of predicted exercise capacity achieved by using previously published regression equations (21). We measured ST-segment deviation to the nearest 0.5 mm and considered it abnormal if horizontal or sloping away from the isoelectric line at least 80 ms after the J point in at least 2 contiguous leads for at least 3 consecutive beats was observed. We recorded exercise-induced angina as being test-limiting or non-test-limiting. The Duke treadmill score (2, 3) was calculated as follows: (Bruce equivalent number of minutes until test termination) $- (5 \times$ the amount of ST-segment deviation in millimeters) $- (4 \times$ the angina index). ST-segment deviation had to be horizontal or sloping away from the isoelectric baseline. The angina index had a value of 0 if no angina occurred during exercise, 1 if non-test-limiting angina occurred, and 2 if test-limiting angina occurred. A score of 5 or greater was considered low risk; between 5 and -10 , intermediate risk; and less than -10 , high risk.

We calculated heart rate recovery as the difference between heart rate at peak exercise and that 1 minute later; we considered a heart rate recovery of 18 or fewer beats/min to be abnormal for patients undergoing stress echocardiography (10) and a heart rate recovery of 12 or fewer beats/min to be abnormal for patients undergoing other types of testing that used an upright cooldown (8, 9). If at least 7 premature ventricular beats/min, frequent ventricular couplets, any ventricular triplets, nonsustained or sustained ventricular tachycardia or torsade de pointes, or ventricular fibrillation occurred in the first 5 minutes of

recovery, we considered frequent ventricular ectopy to be present (12).

Outcomes

The primary outcome was all-cause mortality as determined with the Social Security Death Index (22). Accuracy of this measure can be described in terms of specificity (death has occurred when the Index indicates that it has) and sensitivity (the Index indicates that death has occurred when in fact it has). Previous investigators have demonstrated specificity greater than 99% (22); we have shown that, among patients seen in our exercise laboratory, the sensitivity of the Index exceeds 95% (9). The final censoring date was 1 May 2004. Data on cardiac deaths were not available.

External Validation Data

In 2001, physicians from Kaiser Permanente Colorado, Denver, Colorado, visited the stress laboratory of the Cleveland Clinic, Cleveland, Ohio. They agreed to adopt the Cleveland Clinic database and definitions for their tests. After receiving local institutional review board approval in May 2006, they sent the Cleveland Clinic's de-identified data for analyses from consecutive patients who underwent testing in Denver and met the same inclusion and exclusion criteria as those of patients at the Cleveland Clinic.

Statistical Analysis

We used the Kaplan–Meier method to describe survival according to the Duke treadmill score and tested differences by using the log-rank chi-square statistic. We used Cox proportional hazards modeling (23) to predict mortality rate based on all the variables listed in Table 1 because they are easily measured, are routinely available, and have theoretical support. Any angina during testing was considered a composite variable (angina scores of 1 and 2 were combined) because relatively few patients had test-limiting angina (angina score of 2). To allow for nonlinear associations, we used restricted cubic splines for continuous or ordinal variables (24). We confirmed the proportional hazards assumption for all variables by calculating scaled

Schoenfeld residuals and inspecting time-based hazard ratio plots. We constructed our nomogram on the basis of the results of the proportional hazards model.

Prediction model accuracy was assessed in 2 different ways. First, we assessed model discrimination by measuring the concordance index (c-index) for censored data after 10-fold cross-validation (25). The c-index is analogous to the area under a receiver-operating characteristic curve, where a value of 1 indicates perfect discrimination and 0.5 indicates only random discrimination. Second, we assessed model calibration by grouping patients with regard to nomogram-predicted 10-year survival probabilities obtained by 10-fold cross-validation and comparing the probabilities with actual Kaplan–Meier survival rates.

We compared the predictions from our model with those of the Duke treadmill exercise score. An obvious concern is that the Duke treadmill exercise score was derived on a different data set from that of our nomogram; however, we obtained predictions from our model using 10-fold cross-validation, refitting the model to omit one tenth of the data set and obtaining predictions for the patients in the omitted tenth. Thus, we generated 10 models during this process and obtained a prediction for each patient based on a model that did not contain that patient. These cross-validated predictions were then compared with the Duke treadmill scores.

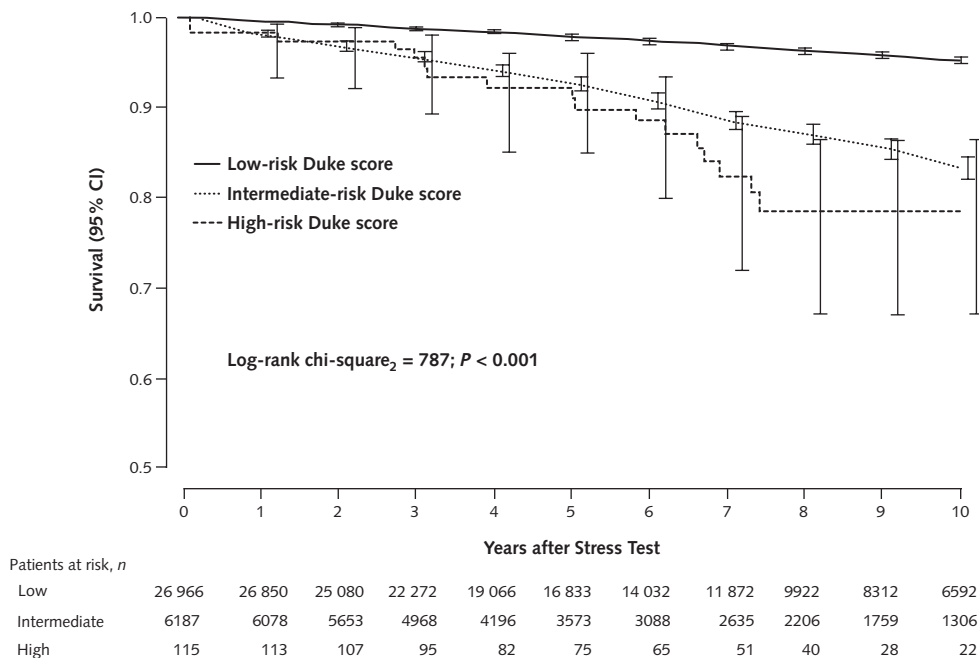
In a supplementary analysis, we tested the multivariable model on the external validation data set from Kaiser Permanente Colorado. As none of their patients were followed for more than 4.3 years, we used the model to identify a cut-point in the derivation set that would yield a 3-year negative predictive value of 0.97. This value was chosen because it is generally agreed that patients with an estimated risk of 1% per year or less should be managed conservatively. After identifying an optimal cut-point for the chosen negative predictive value of 0.97, we calculated event rates for patients below and above the cut-point in the derivation and validation sets.

All analyses were performed by using SAS, version 9.1 (SAS Institute, Cary, North Carolina), or S-plus 2000 Pro-

Table 1. Demographic, Clinical, and Exercise Test Characteristics in the Derivation and Validation Data Sets

Characteristic	Derivation Data Set (n = 33 268)	Validation Data Set (n = 5821)
Median age (25th, 75th percentiles), y	52 (45, 61)	55 (47, 64)
Male sex, n (%)	20 518 (62)	3188 (55)
Current or recent smoking, n (%)	5680 (17)	874 (15)
Non-insulin-treated diabetes mellitus, n (%)	1820 (5)	552 (9)
Insulin-treated diabetes mellitus, n (%)	665 (2)	107 (2)
Hypertension, n (%)	16 177 (49)	2416 (42)
Typical angina, n (%)	358 (1)	101 (2)
Median proportion of predicted exercise capacity achieved (25th, 75th percentiles)	1.03 (0.86, 1.21)	0.94 (0.78, 1.12)
Median ST-segment depression (25th, 75th percentiles), mm	0 (0, 0)	0 (0, 0)
Test angina, n (%)	486 (1)	465 (8)
Abnormal heart rate recovery, n (%)	5644 (17)	1535 (26)
Frequent ventricular ectopy in recovery, n (%)	1204 (4)	192 (3)

Figure 1. Survival according to the Duke treadmill exercise score.



Values below the graph are patients at risk for death each year, after accounting for previous deaths and censoring.

fessional software (Insightful, Seattle, Washington) with Harrell’s Design and Hmisc libraries (26). We derived survival probabilities at specific time points based on the nomogram by using the “nomogram” and “survfit” functions of the Design library, whereas the c-index was calculated by using the “rcorr.cens” function of the Hmisc library.

Role of the Funding Source

The National Heart, Lung, and Blood Institute provided funding for the study. The funding source had no role in the design, conduct, or analysis of the study or in the decision to submit the manuscript for publication.

RESULTS

Baseline and Exercise Characteristics

There were 33 268 patients (20 518 [62%] men) eligible for analyses (median age, 52 years; 25th and 75th percentiles, 45 and 61 years; range, 30 to 93 years). Among these patients, 6187 (19%) and 115 (<1%) had Duke treadmill exercise scores in the intermediate- or high-risk categories, respectively. Only 1246 (4%) were referred for testing as part of preoperative evaluation. Table 1 shows the baseline characteristics. Among the 21 130 patients who had testing after 1 January 1996, 4601 (22%) had atypical angina, 1504 (7%) had nonanginal chest pain, 664 (3%) had dyspnea, 455 (2%) had palpitations, and 266 (1%) had dizziness as presenting symptoms. Among all patients, 2989 (9%) were taking thiazide diuretics, 3559 (11%) were taking β-blockers, 1684 (5%) were taking nondihydropyridine calcium-channel blockers, 1419 (4%)

were taking dihydropyridine calcium-channel blockers, and 2949 (9%) were taking angiotensin-converting enzyme inhibitors. During exercise, 2875 (9%) patients developed between 1 mm and 2 mm of horizontal or downsloping ST-segment depression and an additional 1376 (4%) patients developed ST-segment depression greater than 2 mm.

Duke Treadmill Exercise Score and All-Cause Mortality

The median follow-up among survivors was 6.2 years (25th and 75th percentiles, 3.6 and 9.9 years; range, 1.3 to 13.7 years); 1619 patients died. Figure 1 shows the deaths, by the Duke treadmill score. As the Duke treadmill score decreased, the hazard for death increased (spline-based hazard ratio for score of 5 vs. 10, 5.18 [CI, 4.52 to 5.94]; score of 0 vs. 5, 1.21 [CI, 1.13 to 1.29]; and score of −5 vs. 0, 1.17 [CI, 1.09 to 1.25]) (overall *P* < 0.001). The c-index of the Duke treadmill score for prediction of death was 0.73, indicating moderately good discrimination. Figure 2 shows calibration; predicted and actual risks differed for some risk groups.

Nomogram Prediction of Death and Validation

Table 2 shows the results of multivariable Cox proportional hazards modeling. Statistically significant predictors of death include older age, male sex, current or recent smoking, diabetes mellitus, lower exercise capacity, an abnormal heart rate recovery, and frequent ventricular ectopy during recovery. Hypertension, history of typical angina, ST-segment deviation, and test angina were not predictive of death. A prespecified test for interaction between sex

and exercise capacity was not significant. **Figure 3** shows the nomogram based on our proportional hazards model. The c-index for the model was 0.83, which is substantially better than the value of 0.73 obtained for the Duke score ($P < 0.001$).

Model validation indicated that the degree of optimism was trivial, with essentially no change between the c-index based on the original data set and that based on 10-fold cross-validation. Model calibration, again based on 10-fold cross-validation, was also excellent even after accounting for optimism (**Figure 4**).

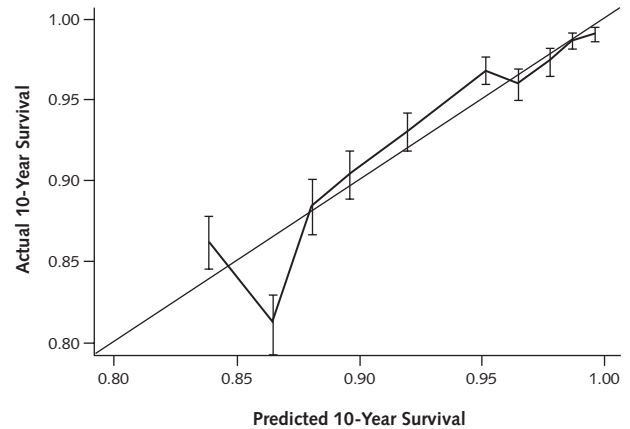
Table 3 compares the classification of patients as low risk (0% to 3% mortality rate over 3 years) or intermediate to high risk (>3% mortality rate over 3 years) by using the Duke treadmill score versus our nomogram-based model. Through use of the Duke score, 26% were classified as intermediate or high risk, whereas the nomogram model classified only 14% as intermediate or high risk. Among the patients classified as intermediate or high risk by the Duke score, the nomogram model reclassified 64% of them as low risk. Among all patients, 21% were reclassified.

External Validation

The external validation data set included 5821 patients. These patients were older and were more likely to have diabetes than patients in the derivation data set; they were also more likely to have exercise-induced angina and an abnormal heart rate recovery (**Table 1**). Over the maximum follow-up of 4.3 years, 62 deaths occurred. As shown in **Figure 5**, mortality rates below and above the optimal cut-point for a negative predictive value in the derivation set of 0.97 were similar in both data sets. Thus, the multivariable model generated from the derivation data set appropriately and successfully identified a low-risk subset in the validation data set.

Our model also discriminated well in the validation data set, with a c-index of 0.774. In post hoc analyses, consideration of clinical variables alone and a model of just

Figure 2. Calibration plot for the Duke treadmill exercise score based on 10-fold cross-validation.



Patients are divided into deciles of predicted risk. The 45° line represents an ideal treadmill exercise score with perfect prediction. Vertical bars represent 95% CIs.

age and percent-predicted exercise capacity yielded lower c-indexes of 0.749 and 0.755, respectively.

DISCUSSION

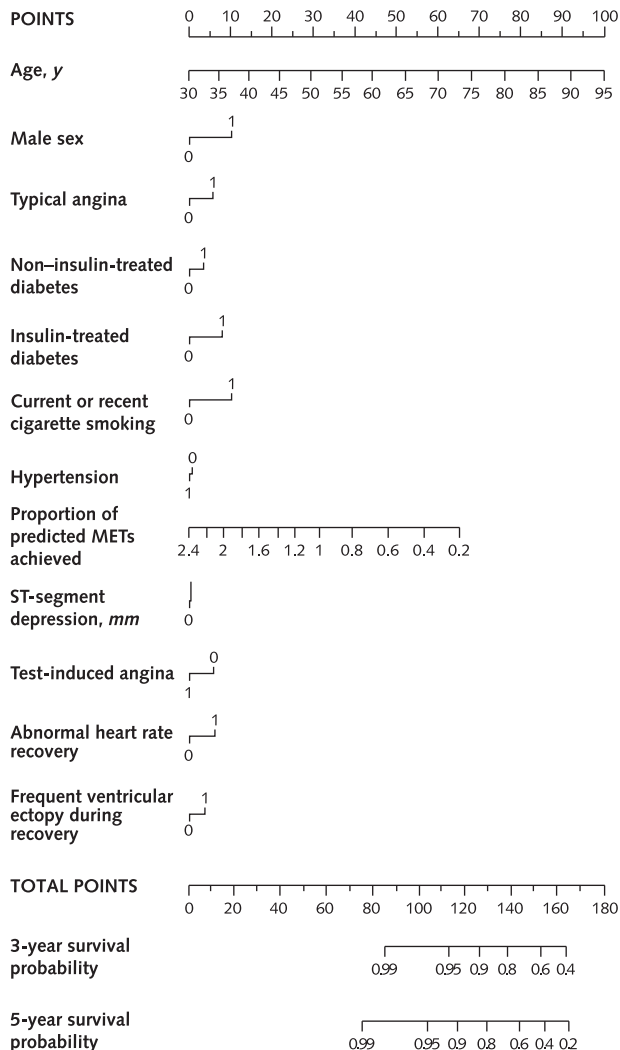
Analyzing the outcomes of a large cohort of more than 30 000 patients with suspected coronary artery disease and a normal electrocardiogram, we developed and externally validated a nomogram-illustrated model for predicting all-cause mortality. This model incorporates pretest variables that can be easily obtained in a stress laboratory without any need for blood tests or previous medical records (5, 6). Our model also takes advantage of recent advances in our understanding of the prognostic value of stress testing by including heart rate recovery (8, 9) and ventricular ectopy during recovery (12). The model performs substantially better than the Duke treadmill score both in discrimina-

Table 2. Results of Cox Proportional Hazards Analysis

Characteristic	Comparison	Hazard Ratio (95% CI)*	P Value
Age	61 y vs. 45 y	4.04 (3.53–4.62)	<0.001
Male sex	Present vs. absent	1.83 (1.63–2.04)	<0.001
Current or recent smoking	Present vs. absent	1.88 (1.67–2.12)	<0.001
Non-insulin-treated diabetes mellitus	Present vs. absent	1.31 (1.12–1.53)	<0.001
Insulin-treated diabetes mellitus	Present vs. absent	1.85 (1.50–2.28)	<0.001
Hypertension	Present vs. absent	0.98 (0.89–1.09)	0.77
Typical angina	Present vs. absent	1.20 (0.77–1.87)	0.43
Proportion of predicted exercise capacity achieved	1.21 vs. 0.86	0.53 (0.49–0.58)	<0.001
ST-segment depression	2.0 vs. 0	1.09 (0.95–1.25)	0.23
Test angina	Present vs. absent	0.77 (0.50–1.20)	0.25
Abnormal heart rate recovery	Present vs. absent	1.46 (1.31–1.62)	<0.001
Frequent ventricular ectopy in recovery	Present vs. absent	1.23 (1.02–1.49)	0.032

* Hazard ratios for continuous and ordinal variables take into account nonlinear associations.

Figure 3. Nomogram of multivariable proportional hazards prediction model for all-cause mortality.



To determine risk, draw a vertical line from each risk marker to the top line, labeled “POINTS,” to calculate points for each risk marker. The sum of all these points is then marked on the line labeled “TOTAL POINTS.” Drop vertical lines from there to yield the 3- and 5-year survival probabilities. For binary variables, 1 means yes and 0 means no. MET = metabolic equivalent.

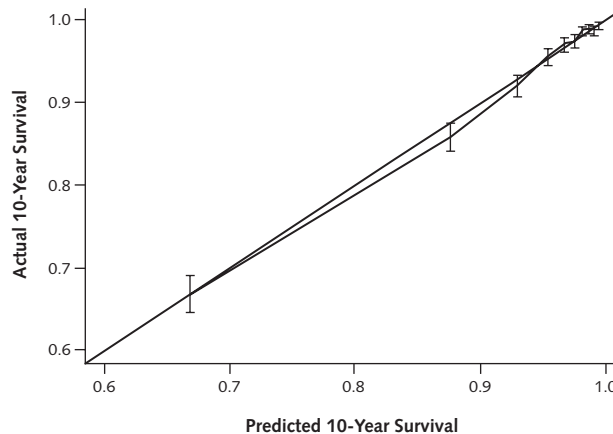
tion (c-index, 0.83 vs. 0.73; $P < 0.001$) and calibration (Figures 2 and 4), even when penalized for its derivation from an internal data set.

Other studies (27) have shown that expanded models outperform the Duke treadmill score for diagnosing coronary disease; our model demonstrates better performance for predicting all-cause mortality. The better performance of the new model makes it possible to confidently identify more low-risk patients in whom further work-up studies, such as cardiac imaging or coronary angiography, would not be needed. This is evident from Table 3, as 64% of the patients identified as intermediate or high risk by the Duke treadmill score were reclassified as low risk with our model.

An important advantage of our Cox proportional hazards model is that it readily incorporates nonlinear associations between risk factors and hazard of death. However, nonlinear spline-based coefficients are difficult to interpret or readily apply to real clinical practice. Therefore, we present the model as a nomogram and also as a simple, user-friendly software application (available free at <http://clinicriskcalculators.org>), whereby a user can enter in values for the variables listed in Table 1 and obtain instant predicted mortality risks.

Most variables on the nomogram have intuitive interpretations. Obvious exceptions were resting typical and test angina: Patients who experienced test angina had better survival than those who did not. The effect of test angina in the statistical model is not significant ($P = 0.20$). However, we did not remove the variable from the model, nor did we limit the model to those variables that were significant on univariable analysis, because either approach is inferior in accuracy to simply fitting the full model (24, 26). As a general rule, variable selection and univariable screening based on statistical significance reduce predictive accuracy of the final model because the variables that remain after this process have overstated effects; their coefficients are biased high in absolute value. A possible explanation for this counterintuitive relationship is that patients obtain points on the nomogram from each variable in our multivariable model. Therefore, the patient who experiences angina may actually have more nomogram points because of the effects of other variables. For example, the patient who experiences angina and has his test stopped prematurely will have a lower recorded exercise capacity and thus more nomogram points from the “percent-predicted metabolic equivalents” variable. In other words, patients who experience test angina but do not get test angina

Figure 4. Calibration curve for 10-year nomogram prediction based on 10-fold cross-validation.



Patients are divided into deciles of predicted risk. The 45° line represents an ideal nomogram with perfect prediction. The vertical bars represent 95% CIs.

Table 3. Classification of 3-Year Risk according to Duke Treadmill Score and Multivariable Nomogram-Based Models

Duke-Predicted Risk	Nomogram-Predicted Risk		Total Patients
	0–3%	>3%	
0–3%			
Patients, <i>n</i>	18 877	1316	20 193
Events, <i>n</i> (%)	148 (0.8)	82 (6.2)	
Kaplan–Meier score (95% CI)*	0.007 (0.006–0.008)	0.053 (0.043–0.065)	
>3%			
Patients, <i>n</i>	4560	2582	7142
Events, <i>n</i> (%)	81 (1.8)	271 (10.5)	
Kaplan–Meier score (95% CI)*	0.015 (0.012–0.019)	0.084 (0.075–0.094)	
Total patients, <i>n</i>	23 437	3898	27 335

* 3-year mortality rate.

nomogram points may be at higher risk once other variables are considered.

It is noteworthy that 2 of the 3 Duke treadmill score variables, ST-segment deviation and test-induced angina, were not predictive of risk. Similar observations have been noted by others (28). One possible reason for this is that the original sample from which the Duke treadmill score was derived included only patients who were referred for coronary angiography (2), whereas subsequent cohorts did not have this potential selection bias (28).

We acknowledge some important limitations. Our nomogram-based model was derived from 1 cohort from a single center, whereas the Duke treadmill score was derived elsewhere and has been externally validated at multiple centers (4, 9). We deliberately and systematically penalized our model with serial 10-fold cross-validations and incorporated only variables that had previously been validated in non–Cleveland Clinic cohorts. We also externally validated the model by using data from patients cared for at Kaiser Permanente Colorado.

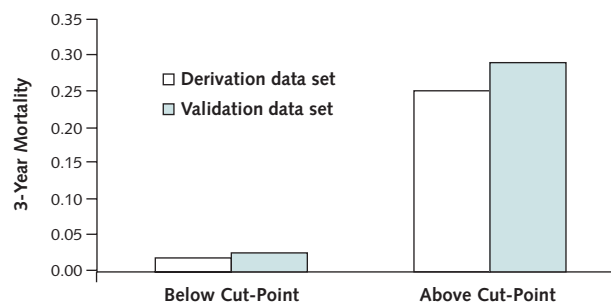
We did not include other variables that are known to be strongly predictive of risk, such as left ventricular ejection fraction, lipid panels, and renal function. We doubt that left ventricular ejection fraction would have substantially improved our model fit, given that all patients had a normal electrocardiogram (18). Assessment of renal function and lipid status requires a blood test; we deliberately restricted our model to variables that would not require a blood test or even access to medical records. Although one might think that such test results would be routinely available, only 56% of patients in the Kaiser Permanente Colorado validation cohort had a lipid profile within the previous 3 months and only 23% had their creatinine levels measured. Importantly, the nomogram-illustrated model discriminated and calibrated well despite the exclusion of these variables.

Only a few patients presented with typical angina. The performance of stress testing in asymptomatic patients is the subject of considerable debate (11, 29, 30), although

the importance of symptoms for assessing prognosis has also been challenged (20). The lack of association between death and typical angina (or such standard risk factors as hypertension) when considering exercise variables is consistent with previous reported findings from other centers (28).

We did not measure cardiac deaths, but we focused instead on all-cause mortality as a wholly unbiased and objective end point (16, 17). Other investigators have validated the Duke treadmill score or functional capacity as predictive of all-cause mortality (4, 28). Because few high-risk patients were in our cohort, we can only comment on distinguishing patients who are at low risk (for whom further cardiac work-up would not be indicated) and those at intermediate or high risk (for whom some kind of work-up would be indicated).

Finally, the clinical utility of our nomogram has not been formally tested. Ideally, a randomized trial could be performed in which patients, providers, or both are randomly assigned to use or nonuse of the nomogram. End points could include appropriate utilization of other cardiac tests (such as nuclear imaging or angiography) and clinical events.

Figure 5. Three-year mortality rates in the derivation and validation data sets.

Multivariable model cut-point yields a negative predictive value of 0.97 in the derivation data set.

Despite these limitations, we found that our nomogram-illustrated model works well for predicting all-cause mortality among patients with normal electrocardiograms who are referred for exercise testing to evaluate suspected coronary artery disease. The nomogram outperformed the current standard Duke treadmill score in both discrimination and calibration. Because the variables included in the nomogram are all easily obtained within the stress laboratory, consideration should be given to routinely including them within the standard clinical risk stratification report.

From the Cleveland Clinic Foundation, Cleveland, Ohio; National Heart, Lung, and Blood Institute, Bethesda, Maryland; and Kaiser Permanente Colorado, Denver, Colorado.

Acknowledgment: The authors thank Ethan Katz and Susana Arrigain for their help with statistical analyses.

Grant Support: By the National Heart, Lung, and Blood Institute (grants NHLBI HL66004 and HL072771).

Potential Financial Conflicts of Interest: None disclosed.

Reproducible Research Statement: Statistical code is available by contacting Dr. Lauer at lauerm@nhlbi.nih.gov. The study protocol and data set are not available.

Requests for Single Reprints: Michael S. Lauer, MD, Division of Prevention and Population Science, National Heart, Lung, and Blood Institute, 6701 Rockledge Drive, Bethesda, MD 20892; e-mail, lauerm@nhlbi.nih.gov.

Current author addresses and author contributions are available at www.annals.org.

References

- Gibbons RJ, Balady GJ, Bricker JT, Chaitman BR, Fletcher GF, Froelicher VF, et al. American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Committee to Update the 1997 Exercise Testing Guidelines. ACC/AHA 2002 guideline update for exercise testing: summary article. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to Update the 1997 Exercise Testing Guidelines). *J Am Coll Cardiol*. 2002;40:1531-40. [PMID: 12392846]
- Mark DB, Hlatky MA, Harrell FE Jr, Lee KL, Califf RM, Pryor DB. Exercise treadmill score for predicting prognosis in coronary artery disease. *Ann Intern Med*. 1987;106:793-800. [PMID: 3579066]
- Mark DB, Shaw L, Harrell FE Jr, Hlatky MA, Lee KL, Bengtson JR, et al. Prognostic value of a treadmill exercise score in outpatients with suspected coronary artery disease. *N Engl J Med*. 1991;325:849-53. [PMID: 1875969]
- Kwok JM, Miller TD, Christian TF, Hodge DO, Gibbons RJ. Prognostic value of a treadmill exercise score in symptomatic patients with nonspecific ST-T abnormalities on resting ECG. *JAMA*. 1999;282:1047-53. [PMID: 10493203]
- Hubbard BL, Gibbons RJ, Lapeyre AC 3rd, Zinsmeister AR, Clements IP. Identification of severe coronary artery disease using simple clinical parameters. *Arch Intern Med*. 1992;152:309-12. [PMID: 1739359]
- Ho KT, Miller TD, Hodge DO, Bailey KR, Gibbons RJ. Use of a simple clinical score to predict prognosis of patients with normal or mildly abnormal resting electrocardiographic findings undergoing evaluation for coronary artery disease. *Mayo Clin Proc*. 2002;77:515-21. [PMID: 12059120]
- Kligfield P, Lauer MS. Exercise electrocardiogram testing: beyond the ST segment. *Circulation*. 2006;114:2070-82. [PMID: 17088475]
- Cole CR, Blackstone EH, Pashkow FJ, Snader CE, Lauer MS. Heart-rate recovery immediately after exercise as a predictor of mortality. *N Engl J Med*. 1999;341:1351-7. [PMID: 10536127]

- Nishime EO, Cole CR, Blackstone EH, Pashkow FJ, Lauer MS. Heart rate recovery and treadmill exercise score as predictors of mortality in patients referred for exercise ECG. *JAMA*. 2000;284:1392-8. [PMID: 10989401]
- Watanabe J, Thamilarasan M, Blackstone EH, Thomas JD, Lauer MS. Heart rate recovery immediately after treadmill exercise and left ventricular systolic dysfunction as predictors of mortality: the case of stress echocardiography. *Circulation*. 2001;104:1911-6. [PMID: 11602493]
- Aktas MK, Ozduran V, Pothier CE, Lang R, Lauer MS. Global risk scores and exercise testing for predicting all-cause mortality in a preventive medicine program. *JAMA*. 2004;292:1462-8. [PMID: 15383517]
- Frolkis JP, Pothier CE, Blackstone EH, Lauer MS. Frequent ventricular ectopy after exercise as a predictor of death. *N Engl J Med*. 2003;348:781-90. [PMID: 12606732]
- Morris CK, Morrow K, Froelicher VF, Hideg A, Hunter D, Kawaguchi T, et al. Prediction of cardiovascular death by means of clinical and exercise test variables in patients selected for cardiac catheterization. *Am Heart J*. 1993;125:1717-26. [PMID: 8498316]
- Prakash M, Myers J, Froelicher VF, Marcus R, Do D, Kalisetti D, et al. Clinical and exercise test predictors of all-cause mortality: results from > 6,000 consecutive referred male patients. *Chest*. 2001;120:1003-13. [PMID: 1155539]
- Morise AP, Jalisi F. Evaluation of pretest and exercise test scores to assess all-cause mortality in unselected patients presenting for exercise testing with symptoms of suspected coronary artery disease. *J Am Coll Cardiol*. 2003;42:842-50. [PMID: 12957430]
- Lauer MS, Blackstone EH, Young JB, Topol EJ. Cause of death in clinical research: time for a reassessment? *J Am Coll Cardiol*. 1999;34:618-20. [PMID: 10483939]
- Lauer MS, Topol EJ. Clinical trials—multiple treatments, multiple end points, and multiple lessons [Editorial]. *JAMA*. 2003;289:2575-7. [PMID: 12759328]
- Khan MA, Sinha S, Hayton S, Fynn S, Henderson RA, Bennett DH. A normal electrocardiogram precludes the need for left ventriculography in the assessment of coronary artery disease. *Heart*. 1998;79:262-7. [PMID: 9602660]
- Chen MS, Blackstone EH, Pothier CE, Lauer MS. Heart rate recovery and impact of myocardial revascularization on long-term mortality. *Circulation*. 2004;110:2851-7. [PMID: 15505081]
- Christopher Jones R, Pothier CE, Blackstone EH, Lauer MS. Prognostic importance of presenting symptoms in patients undergoing exercise testing for evaluation of known or suspected coronary disease. *Am J Med*. 2004;117:380-9. [PMID: 15380494]
- Gulati M, Black HR, Shaw LJ, Arnsdorf MF, Merz CN, Lauer MS, et al. The prognostic value of a nomogram for exercise capacity in women. *N Engl J Med*. 2005;353:468-75. [PMID: 16079370]
- Newman TB, Brown AN. Use of commercial record linkage software and vital statistics to identify patient deaths. *J Am Med Inform Assoc*. 1997;4:233-7. [PMID: 9147342]
- Cox DR. Regression models and life tables. *Journal of the Royal Statistical Society B*. 1972;34:187-220.
- Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15:361-87. [PMID: 8668867]
- Harrell FE Jr, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA*. 1982;247:2543-6. [PMID: 7069920]
- Harrell FE Jr. Regression Modeling Strategies with Application to Linear Models, Logistic Regression, and Survival Analysis. New York: Springer; 2001.
- Fearon WF, Gauri AJ, Myers J, Raxwal VK, Atwood JE, Froelicher VF. A comparison of treadmill scores to diagnose coronary artery disease. *Clin Cardiol*. 2002;25:117-22. [PMID: 11890370]
- Goaraya TY, Jacobsen SJ, Pellikka PA, Miller TD, Khan A, Weston SA, et al. Prognostic value of treadmill exercise testing in elderly persons. *Ann Intern Med*. 2000;132:862-70. [PMID: 10836912]
- Gibbons RJ. Abnormal heart-rate recovery after exercise. *Lancet*. 2002;359:1536-7. [PMID: 12047958]
- American Heart Association Council on Clinical Cardiology, Subcommittee on Exercise, Cardiac Rehabilitation, and Prevention. Exercise testing in asymptomatic adults: a statement for professionals from the American Heart Association Council on Clinical Cardiology, Subcommittee on Exercise, Cardiac Rehabilitation, and Prevention. *Circulation*. 2005;112:771-6. [PMID: 15998671]

Current Author Addresses: Dr. Lauer: Division of Prevention and Population Science, National Heart, Lung, and Blood Institute, 6701 Rockledge Drive, Bethesda, MD 20892.

Ms. Pothier: Cleveland Clinic, 9500 Euclid Avenue, JJ5-801, Cleveland, OH 44195.

Dr. Magid: Kaiser Permanente, 10065 East Harvard Avenue, Suite 300, Denver, CO 80231.

Dr. Smith: Kaiser Permanente, 10350 East Dakota Avenue, Denver, CO 80247.

Dr. Kattan: Cleveland Clinic, 9500 Euclid Avenue, Wb4, Cleveland, OH 44195.

Author Contributions: Conception and design: M.S. Lauer.

Analysis and interpretation of the data: M.S. Lauer, D.J. Magid, M.W. Kattan.

Drafting of the article: M.S. Lauer, M.W. Kattan.

Critical revision of the article for important intellectual content: M.S. Lauer, C.E. Pothier, D.J. Magid, S.S. Smith, M.W. Kattan.

Final approval of the article: M.S. Lauer, D.J. Magid, M.W. Kattan.

Provision of study materials or patients: M.S. Lauer.

Statistical expertise: M.S. Lauer, M.W. Kattan.

Obtaining of funding: M.S. Lauer.

Collection and assembly of data: C.E. Pothier.