Journal of the American College of Cardiology
© 2007 by the American College of Cardiology Foundation
Published by Elsevier Inc.

Vol. 50, No. 19, 2007
ISSN 0735-1097/07/$32.00
doi:10.1016/j.jacc.2007.08.003

External Prognostic Validations and Comparisons of Age- and Gender-Adjusted Exercise Capacity Predictions

Esther S. H. Kim, MD, MPH,* Hemant Ishwaran, PhD,† Eugene Blackstone, MD, FACC,†‡ Michael S. Lauer, MD, FACC, FAHA*†§

Cleveland, Ohio; and Bethesda, Maryland

Objectives The purpose of this study was to externally validate the prognostic value of age- and gender-based nomograms and categorical definitions of impaired exercise capacity (EC).

Background Exercise capacity predicts death, but its use in routine clinical practice is hampered by its close correlation with age and gender.

Methods For a median of 5 years, we followed 22,275 patients without known heart disease who underwent symptom-limited stress testing. Models for predicted or impaired EC were identified by literature search. Gender-specific multivariable proportional hazards models were constructed. Four methods were used to assess validity: Akaike Information Criterion (AIC), right-censored c-index in 100 out-of-bootstrap samples, the Nagelkerke Index R^2, and calculation of calibration error in 100 bootstrap samples.

Results There were 646 and 430 deaths in 13,098 men and 9,177 women, respectively. Of the 7 models tested in men, a model based on a Veterans Affairs cohort (predicted metabolic equivalents [METs] = 18 − [0.15 × age]) had the highest AIC and R^2. In women, a model based on the St. James Take Heart Project (predicted METs = 14.7 − [0.13 × age]) performed best. Categorical definitions of fitness performed less well. Even after accounting for age and gender, there was still an important interaction with age, whereby predicted EC was a weaker predictor in older subjects (p for interaction <0.001 in men and 0.003 in women).

Conclusions Several methods describe EC accounting for age and gender-related differences, but their ability to predict mortality differ. Simple cutoff values fail to fully describe EC's strong predictive value. (J Am Coll Cardiol 2007;50:1867–75) © 2007 by the American College of Cardiology Foundation

Exercise capacity (EC) is a strong independent predictor of death among men and women (1–5), but its widespread adoption in exercise test interpretation has been hindered by its well-known correlation with age and gender (6,7). For example, a 35-year-old man who achieves 8 metabolic equivalents (METs) on exercise treadmill testing would not be considered to have the same EC as a 64-year-old woman who achieves the same number of METs. Several age- and gender-specific nomograms and categorical definitions have been proposed to describe normative values for predicted EC (3,6–12), but it is not known whether there are substantial differences in their prognostic power or their ability to adequately adjust for age-related effects.

We sought to externally validate the prognostic ability of previously published age- and gender-based nomograms and categorical definitions of EC and test their ability to fully account for age-related differences in a population of consecutive patients without known coronary artery disease who were referred for exercise testing. We deliberately focused only on externally derived models, and none of the models tested were derived from patients from our own institution. All-cause mortality was used as an unbiased and objective end point (13).

Methods

Patients. Consecutive patients (Table 1) referred for symptom-limited treadmill exercise testing between January 1, 1995, and December 31, 2002, were potentially eligible for study. To minimize possible bias due to training effects, we included only the first test performed for patients who

From the Departments of *Cardiovascular Medicine, †Quantitative Health Sciences, and ‡Cardiothoracic Surgery, Cleveland Clinic, Cleveland, Ohio; and the §Division of Prevention and Population Science, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland. This work was funded by National Institutes of Health grants R01 HL-66004-2, R01 HL-072771-02, P50 HL-77107, and K12 HD049091.

Manuscript received May 18, 2007; revised manuscript received July 19, 2007, accepted August 6, 2007.

JACC Vol. 50, No. 19, 2007
November 6, 2007:1867–75

Abbreviations and Acronyms

AIC = Akaike Information Criterion
EC = exercise capacity
MET = metabolic equivalent
VA = Veterans Affairs

had more than 1 exercise test during this time period. We excluded patients with known coronary artery disease (including silent Q-wave myocardial infarction), heart failure, clinically significant arrhythmias, valvular or congenital heart disease, cardiomyopathy, end-stage renal disease; patients with a history of prior organ transplantation and pacemaker implantation; patients with abnormal resting electrocardiograms (including left bundle branch block, right bundle branch block, intra-ventricular conduction delay, pre-excitation, pathological Q waves, and >1 mm of ST-segment deviation); and patients without a U.S. Social Security number. We only included patients undergoing testing after January 1, 1995, because before then height and weight were not routinely measured.

Table 1 Baseline and Exercise-Related Characteristics of Patients According to Gender

Characteristics	Men (n = 13,098)	Women (n = 9,177)
Demographic characteristics		
Age, yrs	50 [43, 58]	53 [46, 62]
Black	1,384 (11%)	1,706 (19%)
Body mass index, kg/m^2	27.8 [25.4, 31.0]	27.5 [23.8, 31.6]
Clinical history		
Diabetes, insulin-treated	262 (2%)	198 (2%)
Diabetes, not insulin-treated	745 (6%)	560 (6%)
Hypertension	6,279 (48%)	4,636 (51%)
Smoking	2,250 (17%)	1,546 (17%)
Medication use		
Beta-blocker	1,286 (10%)	1,368 (15%)
Diltiazem or verapamil	491 (4%)	526 (6%)
Nifedipine	528 (4%)	489 (5%)
Angiotension-converting enzyme inhibitor	1,326 (10%)	865 (9%)
Aspirin	2,750 (21%)	1,629 (18%)
Nitrates	409 (3%)	425 (5%)
Statin	1,415 (11%)	959 (10%)
Cardiovascular assessment and exercise capacity		
Resting heart rate, beats/min	71 [63, 81]	77 [68, 86]
Resting blood pressure, mm Hg		
Systolic	130 [118, 140]	130 [116, 142]
Diastolic	86 [80, 92]	84 [78, 90]
Peak systolic blood pressure, mm Hg	190 [174, 208]	180 [162, 198]
Peak heart rate, beats/min	166 [153, 176]	160 [148, 171]
Peak exercise capacity, METs	10.5 [9.0, 12.0]	7.8 [6.5, 9.3]
Abnormal heart rate recovery	1,981 (15%)	1,624 (18%)
Ischemic electrocardiographic changes during or after stress		
ST-segment depression of 1–2 mm	1,014 (8%)	879 (10%)
ST-segment depression of >2 mm	504 (4%)	236 (3%)

Values are median [interquartile range] or n (%).
MET = metabolic equivalent.

Permission to analyze the routinely obtained electronic data from our stress laboratory was given by the Cleveland Clinic Institutional Review Board. The requirement for written informed consent was formally waived.

Clinical data. As described in detail elsewhere (14–16), before exercise testing all patients in our laboratory undergo a structured interview and chart review. Type of diabetes was defined according to treatment (i.e., whether or not insulin was being used). Hypertension was defined as a systolic blood pressure ≥140 mm Hg, diastolic blood pressure ≥90 mm Hg, or use of medications specifically for treating hypertension. Patients were considered cigarette abusers if they regularly smoked currently or within the past year. All patients had height and weight directly measured (not self-reported) before testing. Body mass index was calculated as weight in kilograms divided by height in meters squared.

Exercise testing. Methods for exercise treadmill testing in our laboratory have been described in detail elsewhere (14,15). Briefly, standard protocols (usually Bruce, modified Bruce, and Cornell) were chosen with a goal test duration between 8 and 12 min. All patients were exercised to exhaustion irrespective of heart rate achieved; however, tests were terminated in case of severe chest discomfort (≥7/10 on self-rating scale), significant arrhythmia, hypotension with evidence of clinical compromise, severe ST-segment changes, systolic blood pressure >250 mm Hg, or patient request. Patients were explicitly told not to grip handrails.

At rest and during each stage of exercise, data were prospectively recorded on-line regarding heart rate, blood pressure, symptoms, ST-segment changes, rhythm, and rating of perceived exertion (on a 1-to-10 scale, where 10 is maximum exertion).

Exercise capacity in METs (where 1 MET is 3.5 ml/kg/min of oxygen consumption) (17) was estimated on the basis of protocol, speed, and grade (11,17). If patients only achieved a portion of the final stage of exercise, credit for EC was "pro-rated" according to how much of the stage was completed. For example, if a patient exercised 1 min of a 3-min stage, they were credited with one-third of the increment increase. Chronotropic response to exercise was defined as the percent of heart rate reserve used (18,19). Heart rate recovery was defined as the change in heart rate between peak exercise and 1 min of recovery. For patients undergoing standard exercise testing or testing with nuclear imaging, a value of ≤12 beats/min was considered abnormal (14,15). For patients undergoing exercise echocardiography, a value of ≤18 beats/min was considered abnormal (20). Frequent ventricular ectopy in recovery was defined as frequent ventricular premature complexes, frequent couplets, bigeminy, trigeminy, ventricular tachycardia (nonsustained and sustained), torsades de pointes, and ventricular fibrillation (16). The ST-segment changes were considered ischemic if there was at least 1 mm of horizontal or down-sloping depression at least 80 ms after the J-point.

Prediction of EC. We systematically searched the literature for age- and gender-based regression equations of predicted EC (or oxygen consumption) and for dichotomous age- and/or gender-based definitions of impaired EC. Where multiple models were obtained from the same institution, we chose the one based on the largest patient sample. Names and detailed descriptions of these models are given in Tables 2 and 3. For example, among men the "VA [Veterans Affairs] referral model" (7) predicts peak METs as: $18 - 0.15 \times$ age (Table 2, top row). For women, the "St. James model" (5) predicts peak METs as: $14.7 - 0.13 \times$ age (Table 3, top row). The Cooper models (3,10) define low EC as being below certain values for different age groups (Table 2, row 7; Table 3, row 3). The Mayo models (9) define impaired EC as <7 METs in men and <5 METs in women irrespective of age. It should be noted that none of the models we tested were derived from Cleveland Clinic patients; we deliberately focused only on externally derived models.

Primary outcome. The primary outcome was all-cause death up to July 11, 2006. We ascertained deaths by using the Social Security Death Index (21). We have previously shown that this measure has approximately a 97% sensitivity for "detecting death" in our laboratory (15); others have documented a specificity of >99% (21).

Statistical analyses. All analyses were gender-specific. For descriptive purposes, we constructed Kaplan-Meier plots of cumulative mortality according to whether or not 85% of predicted EC was achieved. This 85% cutoff was chosen on the basis of suggestions in prior literature (6).

Table 2 Comparison of Different Measures of Functional Capacity in Men

(Reference) Institution; First Author, Mean Age ± SD (Age Range), Population Description	n	Age-Based Regression Equations or Definitions for Low Exercise Capacity*	AIC	OOB	R2	CE
(7) VA referral; Morris, 57 (21–89 yrs), veterans, no known CAD referred to evaluate possible or probable CAD	1,388	Pr METs = 18 − 0.15(age)	1,145.53	Age 0.2034 % Pr METs 0.0407	0.136	0.0021
(11) University of Washington; Bruce, 48.6 ± 11.1 yrs, sedentary volunteers, no evident cardiac disease	94	Pr METs = $\frac{[57.8 - 0.445(\text{age})]}{3.5}$	1,130.56	Age 0.1871 % Pr METs 0.0456	0.136	0.0036
(7) VA volunteer population; Morris, 45 ± 14 yrs (18–72 yrs), free-living, healthy, non-veteran volunteers	244	Pr METs = 14.7 − 0.11(age)	1,129.58	Age 0.1791 % Pr METs 0.0484	0.136	0.0024
(12) USAFSAM; Froelicher, (20–53 yrs), healthy aircrewmen referred for medical problems not affecting functional capacity or special project candidates	710	Pr METs = $\frac{[45.7 - 0.27(\text{age})]}{3.5}$	1,120.09	Age 0.138 % Pr METs 0.0572	0.135	0.0014
(8) USAFSAM sedentary men; Wolthuis, 37 yrs (median) (26–47 yrs), sedentary aircrewmen referred for medical problems not affecting functional capacity	345	Pr METs = $\frac{[43.2 - 0.17(\text{age})]}{3.5}$	1,107.84	Age 0.0967 % Pr METs 0.0761	0.135	0.0020
(9) Mayo Clinic; McCully, 69 ± 10 yrs, patients (known CAD 14%) with reduced exercise capacity on exercise echo without cardiomyopathy or moderate/severe valvular lesions	136	Pr METs <7	1,071.34	Age 0.166 Low METs 0.0074	0.126	0.0025
(3) Cooper Clinic; Wei, 43.8 ± 10.1 yrs, men who received medical exam 1970–1993 at preventive medicine clinic; prevalent CAD 9.1%, 12.7%, 16.5% in normal weight, overweight, and obese patients, respectively	25,714	Age → Low fitness 20–39 yrs → ≤10.5 METs 40–49 yrs → ≤9.9 METs 50–59 yrs → ≤8.8 METs ≥60 yrs → ≤7.5 METs	1,015.87	Age 0.1559 Low METs 0.0246	0.128	0.0029

Percent predicted (Pr) metabolic equivalents (METs) = (maximum achieved METs/predicted METs by regression equation) × 100. *Regression equations (nomograms) define peak predicted exercise capacity and categorical definitions define the cutoff below which exercise capacity is considered to be impaired.

AIC = Akaike Information Criterion; CAD = coronary artery disease; CE = calibration error; echo = echocardiography; Low METs = low exercise capacity as defined by specified metabolic equivalent cut-points in reference cited; OOB = change in c-index of model when age, percent predicted metabolic equivalents, or the dichotomous measure of low exercise capacity is permuted with out-of-bootstrap estimation in 100 bootstrap resamples; R2 = Nagelkerke index R^2; USAFSAM = U.S. Air Force School of Aerospace Medicine; VA = Veterans Affairs.

Table 3 Comparison of Different Measures of Functional Capacity in Women

(Reference) Institution; First Author, Mean Age ± SD (Age Range), Population Description, Individual Characteristics	n	Age-Based Regression Equations or Definitions for Low Exercise Capacity*	AIC	OOB	R2	CE
(5) St. James Take Heart Project; Gulati, 52 ± 11 yrs (35–86 yrs), women volunteers, no evident cardiac disease	5,721	Pr METs = 14.7–0.13(age)	582.74	Age 0.1933 % Pr METs 0.0326	0.113	0.020
(11) University of Washington; Bruce, 41.4 ± 11.2 yrs, sedentary volunteers, no evident cardiac disease	113	Pr METs = $\frac{[41.2 - 0.343(\text{age})]}{3.5}$	582.73	Age 0.1747 % Pr METs 0.0357	0.113	0.019
(10) Cooper Women; Barlow, normotensive women 42.8 ± 9.9 yrs, hypertensive women 46.9 ± 9.5 yrs, women without known CAD who received a preventive medical exam during 1970–1998	4,884	Age: 20–39 yrs, 40–49 yrs, 50–59 yrs, ≥60 yrs Fitness: <8.2METs, <7.6METs, <6.7METs, <5.8METs	519.71	Age 0.1679 Low METs 0.0159	0.104	0.013
(9) Mayo Clinic; McCully, 69 ± 10 yrs, patients (known CAD 14%) with reduced exercise capacity without cardiomyopathy or moderate/severe valvular lesions	805	METs <5	518.01	Age 0.1599 Low METs 0.0047	0.102	0.013

*Regression equations (nomograms) define peak predicted exercise capacity, and categorical definitions define the cutoff below which exercise capacity is considered to be impaired. Abbreviations as in Table 2.

We constructed nonparsimonious multivariable Cox proportional hazards models for predicting time to death according to percent predicted EC achieved. Covariates included age, race, body mass index, diabetes (insulin-treated and noninsulin-treated), hypertension, current or recent cigarette smoking, medications (beta-blockers, non-dihydropyridine, and dihydropyridine calcium blockers, angiotensin-converting enzyme inhibitors, aspirin, nitrates, and statins), resting heart rate, resting systolic and diastolic blood pressures, chronotropic response, peak systolic blood pressure, heart rate recovery, ST-segment changes, and frequent ventricular ectopy in recovery. Thus, all models had the name number of covariates. For dichotomous models of EC (e.g., Cooper and Mayo) we used a term for low EC rather than percent predicted EC achieved.

The Cox proportional hazards assumption was confirmed by calculation of Schoenfeld residuals. Non-linearity assumptions were relaxed by consideration of restricted cubic splines (22). In supplementary models, we tested for prespecified interactions including age, body mass index, and race. For illustrative purposes, we constructed plots of adjusted, predicted 10-year survival as a function of percent of predicted EC achieved. In these plots age-stratified predictions are shown; all other covariates were held to either median or modal values.

We used 4 methods to compare different models for prediction of time to death. First, we calculated a modified Akaike Information Criterion (AIC) as: LR chi-square − 2p, where LR chi-square is the model likelihood ratio chi-square and p is the number of model parameters (22). By this formulation, higher values imply models that are closer to the truth.

Second, we tested for discrimination by calculating a c-index for right-censored data (23) in 100 out-of-bootstrap resamples. The c-index, which is analogous to the area under the region-of-interest curve for a purely dichotomous outcome, is calculated by comparing outcomes among patients who died with patients who did not die and had at least as much follow-up as those who did (23). It has a potential value of between 0.5 and 1.0, where 1.0 would imply perfect discrimination. To test the discriminative power of percent predicted EC, we calculated c-indexes only among patients who were not included in each bootstrap sample (i.e., "out-of-bootstrap" test sample). We randomly permuted each variable to see what impact this would have on the total model c-index. For a variable that strongly discriminates risk, this value would be large (i.e., converting that variable to noise in that out-of-bootstrap sample would result in a marked decrease in model discrimination). Third, as a measure of calibration we calculated the Nagelkerke Index R^2 (24). Finally, as an arguably better assessment of calibration, we performed 100 bootstrap resamplings in which patients were divided into quintiles of predicted risk. Within each quintile actual versus predicted survival rates were calculated, and the differences were averaged to derive a weighted calibration error (25). An example of a calibration plot is shown in Figure 1; the

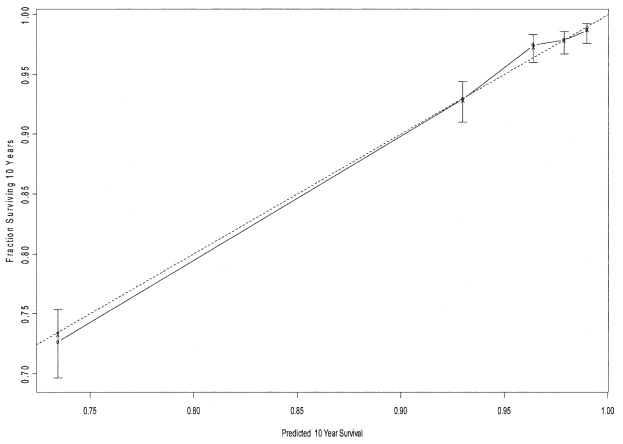


Figure 1 **Actual Versus Predicted 10-Year Survival Rates in Men**

For this calibration plot based on the Veterans Affairs referral nomogram of predicted exercise capacity, patients were divided into quintiles of predicted risk, and 100 bootstrap resamplings were performed. The difference between actual and predicted Kaplan-Meier 10-year death rates was small across all levels of risk.

difference between actual and predicted Kaplan-Meier 10-year death rates was small across all levels of risk.

Statistical analyses were performed with the SAS version 9.1 (SAS Institute, Cary, North Carolina) and R 2.3.1 systems (The R Foundation for Statistical Computing, Vienna, Austria). Regression analyses and plots were performed with Harrell's Design and Hmisc libraries (22).

Results

Baseline characteristics. Baseline and exercise characteristics according to gender are summarized in Table 1. There were 13,098 men and 9,177 women who met inclusion and exclusion criteria. Compared with men, women were older, more likely to be African-American, and somewhat more likely to have hypertension, but there were equivalent frequencies of diabetes and smoking. Women had a higher resting heart rate but similar body mass index. As expected, the median peak EC was lower among women.

Predicted EC and mortality. During follow-up there were 646 and 430 deaths among men and women, respectively. Both for men and women, failure to achieve 85% of predicted EC predicted substantially higher death rates. For example, Figure 2 shows Kaplan-Meier death rates for men according to ability to achieve 85% of predicted EC on the basis of the VA referral model. Similarly, Figure 3 shows that women who failed to achieve 85% of predicted EC on the basis of the St. James model were at markedly increased risk for death.

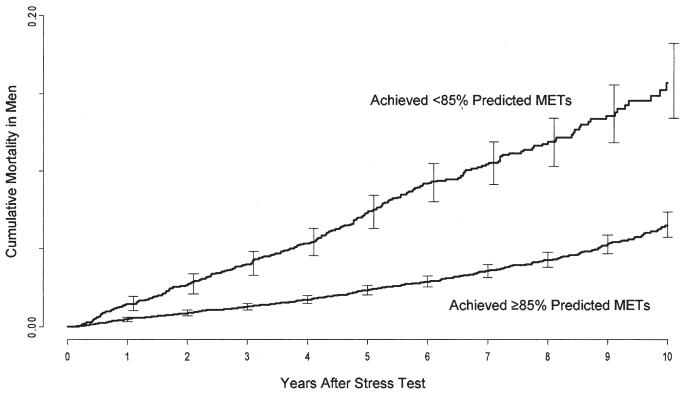
Comparison of EC equations as predictors of death. Tables 2 and 3 show the predictive values of gender-specific multivariable models that used different EC equations. The first column describes the name of the measure, whereas the second column indicates the number of subjects upon which that particular equation was derived. The third column gives the actual equation for peak predicted EC or, in the case of categorical descriptions, the definition of low EC. For all models, percent of predicted EC achieved was a strong independent predictor of death (adjusted $p < 0.0001$).

The fourth column presents the modified AIC for multivariable models according to which measure of predicted EC was used. It is important to note that each model used the exact same covariates (listed in the Methods section). The fifth column presents the relative importance of age and predicted EC on the basis of the change in the right-censored c-index from out-of-bootstrap samples. For all models, age was the strongest predictor of risk, whereas predicted EC was the second strongest ($p < 0.0001$ in all cases). The sixth column presents the Nagelkerke Index R^2. Finally, the right-most column shows the bootstrapped corrected calibration error (that is, the weighted difference [in percent] for predicted versus actual Kaplan-Meier death rates at 10 years).

In men (Table 2), the VA referral model had the highest AIC and the highest Nagelkerke Index R^2. All models showed low calibration error. For discrimination, however, the sedentary Air Force model seemed best. Specifically, for the VA referral model, age had an importance value of 0.20, whereas predicted EC was 0.04. This means that by randomly permuting age in out-of-bootstrap samples, model c-index fell by 0.20 (an enormous change) where, by similarly randomly permuting predicted EC, the c-index fell by a moderate 0.04. In the Air Force model, age was a less important discriminator (change in c-index 0.097), whereas predicted EC became a more important discriminator of risk (0.076); in other words, this model was more successful in using predicted exercise capacity to discriminate risk of death after accounting for age and other confounders.

Corresponding results for women are shown in Table 3. The St. James model performed best by all 4 model validation methods. Again, all models yielded a low calibration error.

Age interactions. For both the VA referral model in men and the St. James model in women, we found an important age interaction (p for interaction <0.001 in men and 0.003 in women) whereby percent of predicted EC achieved behaved differently for predicting death in different age groups (Figs. 4 and 5). These interactions were significant even after adjusting for all confounders. For illustrative purposes, in Figure 4, the multivariable adjusted 10-year survival probability according to percent predicted EC achieved (VA referral model) is shown stratified by different ages. Among the youngest subjects, EC did not predict decreased survival until it had fallen to approximately 60% to 70% of predicted. Below these values, the association between survival and percent predicted EC achieved became fairly steep. In contrast, the association between predicted EC achieved and mortality was less pronounced in older subjects, and there was no clear "hinge point." Among



No. at Risk											
Achieved <85% Predicted METs	2509	2473	2442	2409	2070	1676	1270	932	648	430	187
Achieved ≥85% Predicted METs	10589	10538	10497	10453	9167	7526	5793	4459	3122	2127	1046

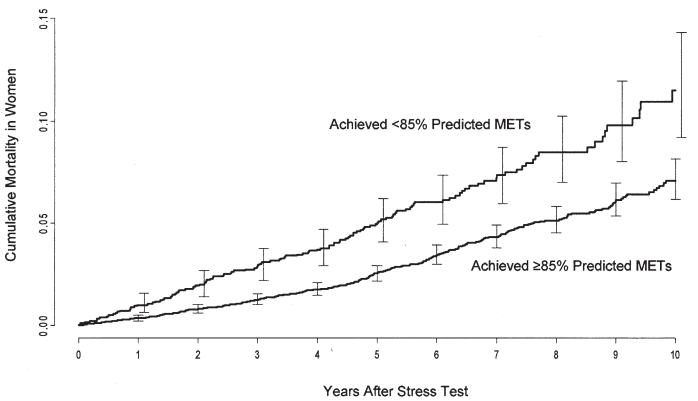
Figure 2 **Mortality in Men by Ability to Achieve ≥85% Predicted METs**

Failure to achieve 85% of predicted exercise capacity with the VA referral nomogram in men predicted substantially higher death rates. MET = metabolic equivalent.

women (Fig. 5), the age-related differences were less pronounced, although the slope of the survival curve is steeper with each decade of life at percent predicted METs <100%.

Discussion

Exercise capacity is known to be one of the most important predictors of death for men and women alike (1–5). In fact, the prognostic ability of the Duke treadmill score might in large part be driven by EC alone (26). Thus, defining normative values for EC is of utmost importance in accurate risk prediction after stress testing. The purpose of our study was not to develop a new nomogram or improve a risk score for predicting death on the basis of exercise testing; we sought to compare previously existing definitions of EC in their prognostic abilities with a large external population



No. at Risk											
Achieved <85% Predicted METs	1796	1779	1761	1744	1524	1237	920	670	531	330	166
Achieved ≥85% Predicted METs	7381	7355	7322	7289	6235	5033	3869	2871	2051	1283	627

Figure 3 **Mortality in Women by Ability to Achieve ≥85% Predicted METs**

Failure to achieve 85% of predicted exercise capacity with the St. James Take Heart Project nomogram in women predicted substantially higher death rates. MET = metabolic equivalent.

through statistical measures of fit (AIC), discrimination (c-index in out-of-bootstrap resamples), and calibration (R^2 and bootstrapped calibration plots).

Although all models performed well, some models clearly predict mortality better than others. Of particular interest is that in both men and women, the categorical descriptors of EC do not predict death as well as age- and gender-based nomograms for predicted EC. An explanation for worse fit of these categorical models in male populations can be derived from inspection of Figure 4, whereby percent predicted METs predicts mortality in different manners on the basis of age by decade. There is a linear relationship between percent predicted METs and mortality in older patients, but in younger patients, there is a "hinge point" where approximately <85% predicted METs predicts increased mortality. In women, however, the slope of the survival curves on the basis of percent predicted METs becomes steeper with increasing age, but there is no value at which there is a "hinge point" for predicting increasing mortality on the basis of EC. This finding indicates that categorical descriptions of EC might correctly predict survival in younger men, but because of the age interaction and linear relationship between EC and survival in other groups, nomograms specific to decade of age might better predict mortality.

Insights from model validations and comparisons. A major strength of our study is that all the models we tested were derived from external data sets. In this regard, our study is the first to perform a series of external validations for previously published descriptions of EC. Model validation is a complex issue, however, because the universal truth is never known and the best investigators can do is identify which models are likely to be closer to the truth (27). Furthermore, validation involves different types of comparisons, with some considering calibration and others discrimination (28). Regarding calibration, which refers to differences between observed and actual event rates at different levels of risk, we found that all multivariable models worked well irrespective of what measure of predicted EC was considered. Regarding discrimination, which refers to the ability to distinguish higher- from lower-risk subjects, there were more marked differences in model performance. It is noteworthy, however, that for nearly all models considered EC was the second strongest discriminator of risk, with only age performing better. This finding stresses the high clinical value of EC in routine risk stratification of patients with suspected coronary disease.

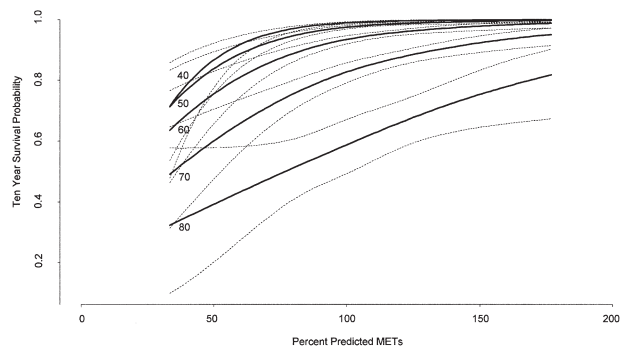


Figure 4 Multivariable Adjusted 10-Year Survival Probability in Men on the Basis of Percent Pr METs and Age

All variables listed in Table 1 were used as covariates to predict survival probability. Continuous variables are considered as medians, whereas categorical variables are considered as modes. Dotted lines refer to 95% confidence intervals. There is a significant age interaction where percent predicted (Pr) metabolic equivalents (METs) achieved during stress testing predicts mortality in a different manner on the basis of age with the VA referral nomogram in men ($p < 0.001$).

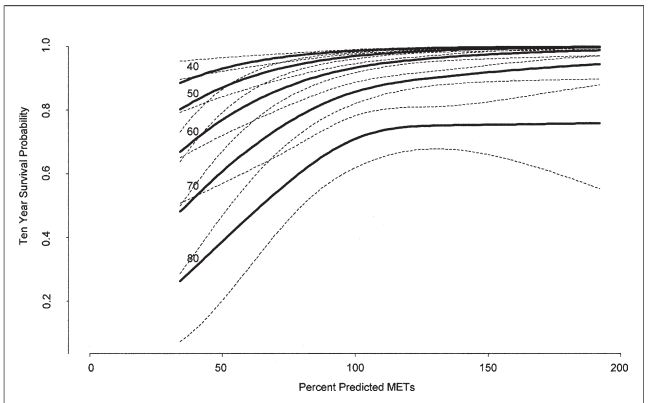


Figure 5 Multivariable Adjusted 10-Year Survival Probability in Women on the Basis of Percent Pr METs and Age

All variables listed in Table 1 were used as covariates to predict survival probability. Continuous variables are considered as medians, whereas categorical variables are considered as modes. Dotted lines refer to 95% confidence intervals. There is a significant age interaction where percent Pr METs achieved during stress testing predicts mortality in a different manner on the basis of age with the Gulati nomogram in women ($p = 0.003$). Abbreviations as in Figure 4.

Study limitations. Our study used METs as an estimate of the EC, because direct measurement of oxygen consumption is not routinely performed during exercise stress testing. Direct measurement of oxygen consumption might have provided a more precise measurement of the effect of EC on mortality; however, the use of METs to describe EC during stress testing is common, and its use has been well established (17). Other potential limitations of our study include the fact that our test population was derived from 1 referral center, patients with known coronary disease were excluded, and nomograms/definitions of impaired EC were derived from diverse populations (healthy U.S. Air Force crewmen in comparison with sedentary veterans referred for stress testing for clinical reasons, for example). Finally, we do not have follow-up data on our study population. Specifically, we do not know whether the clinical management of patients was altered as a result of exercise testing and reporting of EC (e.g., changes in medication prescribing practices or counseling for smoking cessation).

Clinical implications. Although exercise testing is traditionally thought of as a diagnostic test for detection of obstructive coronary lesions, its strength lies in its prognostic ability to identify patients who are at increased risk for death (29). Arguably, EC is the strongest test predictor and should be reported and incorporated into routine clinical practice for risk prediction. We have shown that existing models of EC that account for age and gender are, as expected, strong independent predictors of risk. Recent work has focused on gender-specific nomograms with a clear temptation to define easy-to-remember cutoff values (e.g., 85%) for identifying patients with prognostically important impairment of EC (6). Although this relatively simple approach might work (Figs. 2 and 3), it does not fully capture EC's prognostic value. Continuous measures of percent predicted EC achieved better describe risk than simple dichotomization; this is an observation consistent with that of other risk factors like blood pressure and cholesterol (30). More importantly, however, none of the models fully account for age-related effects. Figures 4 and 5 are illustrations of the observation that even with the age- and gender-based nomograms, EC behaves differently as a risk predictor in older subjects. This behavior is also similar for "classic" risk factors; for example, smoking and cholesterol are weaker predictors of risk in older subjects (30).

Although none of the models completely account for the strong interaction between age and EC in predicting all cause mortality, of the nomograms available for use, we would recommend either the St. James model or the University of Washington model in women and the VA referral model in men. In an age where complex prediction models can be used by clinicians by entering variable fields during a stress test or clinical visit into a computer, we would recommend the routine incorporation of all available clinical and exercise test findings in a global prediction of risk rather than focusing on simplistic normal and abnormal cut-points, even for a prognostic variable as powerful as EC. We have previously derived and validated complex computer-based models for predicting mortality in patients undergoing exercise testing (31). Such comprehensive, integrated, computer-based models need to be improved to better account for age- and gender-related differences in EC as a next critical step for accurately assessing an individual patient's risk and directing preventive care.

Reprint requests and correspondence: Dr. Michael S. Lauer, 6701 Rockledge Drive, Room 10122, Bethesda, Maryland 20892. E-mail: lauerm@nhlbi.nih.gov.

REFERENCES

1. Blair SN, Kohl HW 3rd, Paffenbarger RS Jr., Clark DG, Cooper KH, Gibbons LW. Physical fitness and all-cause mortality. A prospective study of healthy men and women. JAMA 1989;262:2395–401.
2. Ekelund LG, Haskell WL, Johnson JL, Whaley FS, Criqui MH, Sheps DS. Physical fitness as a predictor of cardiovascular mortality in asymptomatic North American men. The Lipid Research Clinics Mortality Follow-up Study. N Engl J Med 1988;319:1379–84.
3. Wei M, Kampert JB, Barlow CE, et al. Relationship between low cardiorespiratory fitness and mortality in normal-weight, overweight, and obese men. JAMA 1999;282:1547–53.
4. Mora S, Redberg RF, Cui Y, et al. Ability of exercise testing to predict cardiovascular and all-cause death in asymptomatic women: a 20-year follow-up of the lipid research clinics prevalence study. JAMA 2003;290:1600–7.
5. Gulati M, Pandey DK, Arnsdorf MF, et al. Exercise capacity and the risk of death in women: the St James Women Take Heart Project. Circulation 2003;108:1554–9.
6. Gulati M, Black HR, Shaw LJ, et al. The prognostic value of a nomogram for exercise capacity in women. N Engl J Med 2005;353:468–75.
7. Morris CK, Myers J, Froelicher VF, Kawaguchi T, Ueshima K, Hideg A. Nomogram based on metabolic equivalents and age for assessing aerobic exercise capacity in men. J Am Coll Cardiol 1993;22:175–82.
8. Wolthuis RA, Froelicher VF Jr., Fischer J, Triebwasser JH. The response of healthy men to treadmill exercise. Circulation 1977;55:153–7.
9. McCully RB, Roger VL, Ommen SR, et al. Outcomes of patients with reduced exercise capacity at time of exercise echocardiography. Mayo Clin Proc 2004;79:750–7.
10. Barlow CE, LaMonte MJ, Fitzgerald SJ, Kampert JB, Perrin JL, Blair SN. Cardiorespiratory fitness is an independent predictor of hypertension incidence among initially normotensive healthy women. Am J Epidemiol 2006;163:142–50.
11. Bruce RA, Kusumi F, Hosmer D. Maximal oxygen intake and nomographic assessment of functional aerobic impairment in cardiovascular disease. Am Heart J 1973;85:546–62.
12. Froelicher VF Jr., Allen M, Lancaster MC. Maximal treadmill testing of normal USAF aircrewmen. Aerosp Med 1974;45:310–5.
13. 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.
14. 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.
15. 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.
16. 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.
17. Jette M, Sidney K, Blumchen G. Metabolic equivalents (METS) in exercise testing, exercise prescription, and evaluation of functional capacity. Clin Cardiol 1990;13:555–65.
18. Wilkoff BL, Miller RE. Exercise testing for chronotropic assessment. Cardiol Clin 1992;10:705–17.
19. Lauer MS, Francis GS, Okin PM, Pashkow FJ, Snader CE, Marwick TH. Impaired chronotropic response to exercise stress testing as a predictor of mortality. JAMA 1999;281:524–9.
20. 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.
21. 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.
22. Harrell FE. Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis. New York, NY: Springer, 2001.
23. Harrell FE Jr., Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. JAMA 1982;247:2543–6.
24. Ash A, Shwartz M. R2: a useful measure of model performance when predicting a dichotomous outcome. Stat Med 1999;18:375–84.
25. Peeters KC, Kattan MW, Hartgrink HH, et al. Validation of a nomogram for predicting disease-specific survival after an R0 resection for gastric carcinoma. Cancer 2005;103:702–7.

26. Gulati M, Arnsdorf MF, Shaw LJ, et al. Prognostic value of the duke treadmill score in asymptomatic women. Am J Cardiol 2005;96:369–75.
27. Burnham KP, Anderson DR. Model Selection and Inference: A Practical Information-Theoretic Approach. New York, NY: Springer-Verlag, 1998.
28. Cook NR, Buring JE, Ridker PM. The effect of including C-reactive protein in cardiovascular risk prediction models for women. Ann Intern Med 2006;145:21–9.
29. Lauer MS. Exercise electrocardiogram testing and prognosis. Novel markers and predictive instruments. Cardiol Clin 2001;19:401–14.
30. D'Agostino RB Sr., Grundy S, Sullivan LM, Wilson P. Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. JAMA 2001;286:180–7.
31. Hesse B, Morise A, Pothier CE, Blackstone EH, Lauer MS. Can we reliably predict long-term mortality after exercise testing? An external validation. Am Heart J 2005;150:307–14.