

1-2017

Use of Sex-Specific Clinical and Exercise Risk Scores to Identify Patients at Increased Risk for All-Cause Mortality

Paul C. Cremer
Cleveland Clinic

Yuping Wu
Cleveland State University, y.wu88@csuohio.edu

Haitham M. Ahmed
Johns Hopkins Hospital

Lee M. Pierson
Cleveland Clinic

Danielle M. Brennan
Henry Ford Hospital

Below this page find additional works that https://engagedscholarship.csuohio.edu/scimath_facpub

 Part of the [Cardiovascular Diseases Commons](#), and the [Mathematics Commons](#)

How does access to this work benefit you? Let us know!

Repository Citation

Cremer, Paul C.; Wu, Yuping; Ahmed, Haitham M.; Pierson, Lee M.; Brennan, Danielle M.; Al-Mallah, Mouaz H.; Brawner, Clinton A.; Ehrman, Jonathan K.; Keteyian, Steven J.; Blumenthal, Roger S.; Blaha, Michael J.; and Cho, Leslie, "Use of Sex-Specific Clinical and Exercise Risk Scores to Identify Patients at Increased Risk for All-Cause Mortality" (2017). *Mathematics Faculty Publications*. 257.
https://engagedscholarship.csuohio.edu/scimath_facpub/257

This Article is brought to you for free and open access by the Mathematics and Statistics Department at EngagedScholarship@CSU. It has been accepted for inclusion in Mathematics Faculty Publications by an authorized administrator of EngagedScholarship@CSU. For more information, please contact library.es@csuohio.edu.

Authors

Paul C. Cremer, Yuping Wu, Haitham M. Ahmed, Lee M. Pierson, Danielle M. Brennan, Mouaz H. Al-Mallah, Clinton A. Brawner, Jonathan K. Ehrman, Steven J. Keteyian, Roger S. Blumenthal, Michael J. Blaha, and Leslie Cho

Use of Sex-Specific Clinical and Exercise Risk Scores to Identify Patients at Increased Risk for All-Cause Mortality

Paul C. Cremer, MD; Yuping Wu, PhD; Haitham M. Ahmed, MD, MPH; Lee M. Pierson, MD; Danielle M. Brennan, MS; Mouaz H. Al-Mallah, MD, MSc; Clinton A. Brawner, PhD; Jonathan K. Ehrman, PhD; Steven J. Keteyian, PhD; Roger S. Blumenthal, MD; Michael J. Blaha, MD, MPH; Leslie Cho, MD

IMPORTANCE Risk assessment tools for exercise treadmill testing may have limited external validity. Cardiovascular mortality has decreased in recent decades, and women have been underrepresented in prior cohorts.

OBJECTIVES To determine whether exercise and clinical variables are associated with differential mortality outcomes in men and women and to assess whether sex-specific risk scores better estimate all-cause mortality.

DESIGN, SETTING, AND PARTICIPANTS This retrospective cohort study included 59 877 patients seen at the Cleveland Clinic Foundation (CCF cohort) from January 1, 2000, through December 31, 2010, and 49 278 patients seen at the Henry Ford Hospital (FIT cohort) from January 1, 1991, through December 31, 2009. All patients were 18 years or older and underwent exercise treadmill testing. Data were analyzed from January 1, 2000, to October 27, 2011, in the CCF cohort and from January 1, 1991, to April 1, 2013, in the FIT cohort.

MAIN OUTCOMES AND MEASUREMENTS The CCF cohort was divided randomly into derivation and validation samples, and separate risk scores were developed for men and women. Net reclassification, C statistics, and integrated discrimination improvement were used to compare the sex-specific risk scores with other tools that have all-cause mortality as the outcome. Discrimination and calibration were also evaluated with these sex-specific risk scores in the FIT cohort.

RESULTS The CCF cohort included 59 877 patients (59.4% men; 40.5% women) with a median (interquartile range [IQR]) age of 54 (45-63) years and 2521 deaths (4.2%) during a median follow-up of 7 (IQR, 4.1-9.6) years. The FIT cohort included 49 278 patients (52.5% men; 47.4% women) with a median (IQR) age of 54 (46-64) years and 6643 deaths (13.5%) during a median (IQR) follow-up of 10.2 (7-13.4) years. C statistics for the sex-specific risk scores in the CCF validation sample were higher (0.79 in women and 0.81 in men) than C statistics using other tools in women (0.70 for Duke Treadmill Score; 0.74 for Lauer nomogram) and men (0.72 for Duke Treadmill Score; 0.75 for Lauer nomogram). Net reclassification and integrated discrimination improvement were superior with the sex-specific risk scores, mostly owing to correct reclassification of events. The sex-specific risk scores in the FIT cohort demonstrated similar discrimination (C statistic, 0.78 for women and 0.79 for men), and calibration was reasonable.

CONCLUSIONS AND RELEVANCE Sex-specific risk scores better estimate mortality in patients undergoing exercise treadmill testing. In particular, these sex-specific risk scores help to identify patients at the highest residual risk in the present era.

JAMA Cardiol. 2017;2(1):15-22. doi:10.1001/jamacardio.2016.3720
Published online October 26, 2016.

← Invited Commentary page 22

+ Supplemental content at
jamacardiology.com

Author Affiliations: Heart and Vascular Institute, Cleveland Clinic, Cleveland, Ohio (Cremer, Pierson, Cho); Department of Mathematics, Cleveland State University, Cleveland, Ohio (Wu); Ciccarone Center for the Prevention of Heart Disease, Johns Hopkins Hospital, Baltimore, Maryland (Ahmed, Blumenthal, Blaha); C5 Research, Cleveland Clinic, Cleveland, Ohio (Brennan); King Abdulaziz Cardiac Center, King Saud bin Abdulaziz University for Health Sciences, King Abdullah International Medical Research Center, Ministry of National Guard, Health Affairs, Riyadh, Saudi Arabia (Al-Mallah); Division of Cardiovascular Medicine, Henry Ford Hospital, Detroit, Michigan (Al-Mallah, Brawner, Ehrman, Keteyian).

Corresponding Author: Leslie Cho, MD, Heart and Vascular Institute, Cleveland Clinic, 9500 Euclid Ave, Ste JB-1, Cleveland, OH 44124 (chol@ccf.org).

Exercise testing is recommended to assess prognosis in patients with known or suspected coronary artery disease (CAD) who present with symptoms suggestive of worsening ischemic heart disease.¹ For nearly 3 decades, the Duke Treadmill Score (DTS) has been the standard to assess prognosis in these patients.² However, during the past few decades, advances in therapy have dramatically reduced cardiovascular mortality,³ and the validity of this score in a contemporary population is unclear. The DTS was also developed in a predominantly male population, and a paucity of data regarding risk stratification with this score exists in women.⁴ Finally, the DTS incorporates only exercise duration, ST-segment depression, and exercise-induced chest pain.² Other exercise variables are associated with prognosis,⁵⁻⁷ and many of these patients have cardiovascular comorbidities that also affect their overall risk.

To address this latter concern, a nomogram was created by Lauer and colleagues⁸ that is superior to the DTS at predicting all-cause mortality. The broader clinical use of this nomogram is unclear and in part has been limited by exclusion of certain patient populations, including those with known CAD, valvular heart disease, heart failure, and end-stage renal disease (ESRD). Furthermore, sex-related differences in the prognostic impact of exercise test variables and cardiovascular comorbidities should be accounted for.^{9,10} We therefore aimed to develop comprehensive sex-specific risk scores to estimate all-cause mortality in a more inclusive and contemporary population. We then determined whether these sex-specific risk scores better estimated mortality when compared with the DTS and Lauer nomogram. Finally, we validated these sex-specific risk scores by assessing discrimination and calibration in an external cohort.

Methods

Cleveland Clinic Cohort

From a cohort of 60 895 consecutive patients undergoing symptom-limited treadmill testing at the Cleveland Clinic Foundation (CCF) from January 1, 2000, to December 31, 2010, 1018 were excluded owing to atrial fibrillation, a resting electrocardiogram that precluded interpretation of the ST segment, digoxin use, being younger than 18 years, or having no Social Security number available. The final CCF cohort included 59 877 patients.

At the time of stress testing, patient demographic characteristics, comorbidities, and medications were prospectively entered into a stress database. Known CAD was defined as a previous myocardial infarction, previous percutaneous coronary intervention, or a history of coronary artery bypass grafting. Heart failure was defined by self-reported history and review of the medical record. Hypertension was defined as self-reported history or use of antihypertensives. Hyperlipidemia was defined as an abnormal fasting lipid panel according to Adult Treatment Panel III guidelines, self-reported history, or use of medications to lower lipid levels. Diabetes was defined as a fasting blood glucose level of at least 126 mg/dL (to convert to millimoles per liter, multiply by 0.0555), self-reported

Key Points

Question Do sex-specific risk scores better estimate all-cause mortality for patients undergoing exercise treadmill testing?

Findings In a retrospective cohort study of nearly 110 000 patients, sex-specific risk scores better estimated mortality. Exercise capacity had the greatest effect on prognosis in both sexes, and all risk factors had a differential effect on prognosis in women compared with men.

Meaning Risk stratification is improved with sex-specific risk scores, and in particular, patients at the highest risk are more readily identified.

history, or use of medication to lower glucose levels. Patients who were actively smoking cigarettes or who had smoked within the past year were considered current smokers, and patients who had smoked less recently were considered to have a history of smoking. We defined ESRD as receiving dialysis.

Patients underwent symptom-limited treadmill testing using a protocol based on a pretest estimation of exercise capacity and designed to have the patient reach maximal exertion within 8 to 12 minutes, as suggested by exercise testing guidelines.¹ Standard exercise protocols were used, and most patients performed a Bruce protocol (61%). Other protocols included Cornell, Naughton, modified Naughton, and modified Bruce.¹ Heart rate targets were not used as an end point or to judge the adequacy of the test. The ST segment was measured 80 milliseconds after the J point, and the magnitude of ST depression was recorded as the greatest horizontal or down-sloping ST-segment depression in any lead except aVR during the test or in recovery.

Blood pressure was measured during every stage of the test. Heart rate was recorded from an electrocardiogram printed every minute during the test. Peak estimated metabolic equivalents of task (METs) were calculated from treadmill speed and grade at peak exercise. Chest discomfort during the test was recorded as none, nonlimiting chest pain, or test-limiting chest pain. Rate-pressure product (RPP) was calculated as the product of heart rate and systolic blood pressure. A Δ RPP was calculated as RPP at peak exercise minus RPP at rest. Heart rate recovery (HRR) was calculated as peak exercise heart rate minus heart rate at 1 minute after exercise.

Patients were given a standard walking recovery for tests involving electrocardiography only, technetium imaging, or metabolic stress testing. For patients undergoing stress echocardiography, a supine recovery immediately after exercise was used. Therefore, HRR at 1 minute was classified as abnormal if 12 or fewer beats/min for patients undergoing upright recovery and abnormal if 18 or fewer beats/min in patients undergoing stress echocardiography.^{5,11,12} Chronotropic reserve index was calculated as (Peak heart rate - resting heart rate) / [(220 - age) - resting heart rate] and was considered abnormal if no greater than 0.8 for patients not taking a β -blocker and abnormal if at least 0.62 for patients taking a β -blocker. In patients who did not undergo a Bruce protocol, the estimated METs achieved by each patient were converted to minutes per the Bruce protocol before calculation of the DTS.

The DTS was calculated as Exercise time - (5 × maximum ST-segment depression) - (4 × treadmill chest pain index). Treadmill chest pain was scored from 0 to 2, with 0 representing no chest pain; 1, nonlimiting chest pain; and 2, chest pain for which the exercise test was terminated.²

Validation Cohort

The Henry Ford Exercise Testing (FIT) cohort is from a registry of 69 885 consecutive patients who had physician-referred exercise treadmill tests at the Henry Ford Health System from January 1, 1991, through December 31, 2009. Methodologic details have been reported previously.¹³ In brief, patients older than 18 years who underwent exercise treadmill tests were included. All testing used the standard Bruce protocol. Exercise test, medical history, and medication data were collected at the time of testing, and supporting clinical data were derived from the electronic medical record and administrative databases. For external validation purposes, the 3880 patients without recorded weight and the 16 727 without glomerular filtration rate data were excluded. A final sample size of 49 278 patients was included for external validation.

All data in both cohorts were deidentified. The institutional review boards at CCF, Henry Ford Health System, and Johns Hopkins Hospital approved this study with an exemption for individual patient consent.

Outcome

The primary outcome was all-cause mortality and was determined from the Social Security Death Index Master File. Previous work¹⁴ demonstrated that more than 95% of the time the Social Security Death Index correctly identifies patients who have died. The final censoring date was October 27, 2011, in the CCF cohort, and April 1, 2013, in the FIT cohort.

Statistical Analysis

Data were analyzed from January 1, 2000, to October 27, 2011, in the CCF cohort and from January 1, 1991, to April 1, 2013, in the FIT cohort. For men and women, the CCF sample was divided randomly, with 50% of patients in the derivation cohort and 50% in the validation cohort. All data analysis to develop risk scores was performed in the derivation cohorts. Data are summarized as median and interquartile range (IQR) for continuous data and number (percentage) of nonmissing data for categorical variables. Comparisons across age categories and survival status used 2-tailed unpaired *t* tests for continuous variables and χ^2 tests for categorical variables. Cox proportional hazards regression models were used to create separate multivariable models for men and women to determine independent risk factors for all-cause mortality.

All variables that were significantly associated with all-cause mortality on univariable analysis ($P < .05$) were considered for multivariable adjustment. Bootstrapping methods were used to identify variables for inclusion in the final models. Two hundred bootstrapped models were generated for men and women; variables that were entered into the models at least 50% of the time were then entered into a backward stepwise selection modeling process to create separate Cox models for

men and women. These Cox models were validated for calibration accuracy to estimate overall survival (eMethods in the Supplement).

The Cox models from the derivation cohort were then used to develop sex-specific risk scores for estimating mortality. To assign value to each variable in creating a risk score, categories were created for the continuous variables in the model. Creation of these categories was based on the distribution of each variable. Linearity was tested with restricted cubic splines. Continuous variables were divided into quartiles with the exception of age for women and weight for men. The β coefficients across quartiles were similar for these variables; thus, age for women was divided into older than 65 years or 65 years or younger, and weight for men was divided into more than 80 kg or 80 kg or less. The β coefficients of each covariate in these categories were then used to assign points for every risk factor (eMethods in the Supplement). The points were then added together to obtain a total score. Overall, 7% of data were missing. To reduce bias in estimates and uncertainty related to the imputation model, multiple imputation of missing variables was performed with a regression-based method. In the CCF validation cohorts, discrimination was assessed with receiver operating characteristic curves and Harrell C statistics, category-free net reclassification improvement, and integrated discrimination improvement.¹⁵

The FIT cohort served as an external validation cohort. The risk scores were modified because certain data were not available in both cohorts. Heart rate recovery was not available in the FIT database and was excluded from the models using this cohort. A history of smoking was not available in the FIT cohort and was replaced with current smoking. Finally, ESRD was not categorized in the FIT cohort and was replaced by glomerular filtration rate of less than 15 mL/min/1.73 m². Discrimination with these modifications to the risk scores was assessed with receiver operating characteristic curves and Harrell C statistics. Calibration was assessed by dividing the risk scores into deciles for men and women and plotting observed vs predicted mortality.

All analyses were performed using SAS (version 9.2; SAS Institute Inc), R (R Core Team 2015 [<http://www.R-project.org/>]), and STATA (version 14.0; StataCorp) statistical software. Two-tailed $P < .05$ was considered statistically significant.

Results

Patient Characteristics

Owing to known differences in the prevalence and impact of comorbidities and exercise variables between men and women,¹⁰ the 59 877 patients in the CCF cohort were divided by sex (59.4% men; 40.5% women) for all analyses. Overall, the median age was 54 (IQR, 45-63) years, and 66.4% of the population was white. Cardiovascular comorbidities were more common in men, especially a history of CAD (24.2% vs 9.2%). Exercise capacity was generally preserved, and men had higher exercise capacity (10 [IQR, 8.3-11.5] vs 8 [IQR, 6.6-10] METs) and DTSs (8.5 [IQR, 5.5-10.2] vs 6.5 [IQR, 4-8.2]) compared with women (Table 1). In the FIT cohort (52.5% men; 47.4%

Table 1. Baseline Clinical, Medication, and Exercise Data for the CCF Cohort

Data	CCF Cohort, No. (%)			P Value
	Overall (n = 59 877)	Women (n = 24 292)	Men (n = 35 585)	
Clinical				
Age, median (IQR), y	54 (45-63)	53 (45-62)	54 (45-63)	.007
White race	39 753 (66.4)	15 201 (62.6)	24 552 (69)	<.001
BMI, median (IQR)	28 (24.9-31.7)	27.5 (23.7-32.2)	28.2 (25.6-31.4)	<.001
Body weight, median (IQR), kg	84 (72-97)	73 (63-86)	89 (80-101)	<.001
Diabetes	6924 (11.6)	2635 (10.8)	4289 (21.1)	<.001
Hypertension	27 142 (45.3)	10 610 (43.7)	16 532 (46.4)	<.001
Hyperlipidemia	31 594 (52.8)	11 610 (47.8)	19 984 (56.2)	<.001
History of smoking	26 894 (44.9)	9735 (40.1)	17 159 (48.2)	<.001
Current smoker	7361 (12.3)	2988 (12.3)	4373 (12.3)	.82
Family history of CAD	19 927 (33.3)	8655 (35.6)	11 272 (31.7)	<.001
CAD	10 855 (18.1)	2225 (9.2)	8630 (24.2)	<.001
Previous				
MI	4766 (8)	957 (3.9)	3809 (10.7)	<.001
PCI	4370 (7.3)	853 (3.5)	3517 (9.9)	<.001
CABG	3181 (5.3)	466 (1.9)	2715 (7.6)	<.001
COPD	900 (1.5)	349 (1.4)	551 (1.5)	.14
Stroke or TIA	1350 (2.3)	532 (2.2)	818 (2.3)	.17
CHF	918 (1.5)	333 (1.4)	585 (1.6)	.002
ESRD	264 (0.4)	84 (0.3)	180 (0.5)	.002
PVD	684 (1.1)	221 (0.9)	463 (1.3)	<.001
Medications				
Aspirin	22 692 (37.9)	7228 (29.8)	15 464 (43.4)	<.001
Clopidogrel	3339 (5.6)	856 (3.5)	2483 (7)	<.001
β-Blocker	14 848 (24.8)	5683 (23.4)	9165 (25.8)	<.001
ACEI or ARB	14 964 (25)	5110 (21)	9854 (27.7)	<.001
Statin	18 990 (31.7)	5950 (24.5)	13 040 (36.6)	<.001
Exercise data				
Rest				
Heart rate, median (IQR), beats/min	69 (62-78)	73 (65-82)	67 (60-76)	<.001
Systolic BP, median (IQR), mm Hg	128 (118-140)	126 (112-140)	130 (118-142)	<.001
RPP ^a	8.9 (7.6-10.5)	9.2 (7.8-10.8)	8.7 (7.4-10.2)	<.001
No ST depression	47 444 (79.2)	19 628 (81)	27 816 (78.2)	<.001
Chest pain	1623 (2.7)	552 (2.3)	1071 (3)	<.001
Termination owing to chest pain	927 (1.5)	362 (1.5)	565 (1.6)	.36
Maximum heart rate, median (IQR), beats/min	157 (144-171)	157 (144-170)	159 (144-171)	<.001
Maximum RPP, median (IQR) ^a	28.1 (24.2-31.8)	26.8 (23.3-30.3)	29 (25-32.8)	<.001
Maximum ΔRPP, median (IQR) ^b	18.8 (14.9-22.6)	17.3 (13.9-20.6)	20.1 (16-23.7)	<.001
Peak METs, median (IQR)	9.3 (7.2-11)	8 (6.6-10)	10 (8.3-11.5)	<.001
Abnormal HRR ^c	9704 (16.2)	3948 (16.3)	5756 (16.2)	.81
Abnormal CRI ^d	11 156 (18.6)	4993 (20.6)	6163 (17.3)	<.001
DTS, median (IQR)	7.5 (4.5-9.5)	6.5 (4-8.2)	8.5 (5.5-10.2)	<.001

Abbreviations: ACEI, angiotensin converting enzyme inhibitor; ARB, angiotensin receptor blocker; BMI, body mass index (calculated as weight in kilograms divided by height in meters squared); BP, blood pressure; CABG, coronary artery bypass graft; CAD, coronary artery disease; CCF, Cleveland Clinic Foundation; CHF, congestive heart failure; COPD, chronic obstructive pulmonary disease; CRI, chronotropic response index; DTS, Duke Treadmill Score; ESRD, end-stage renal disease; HRR, heart rate recovery; IQR, interquartile range; METs, estimated metabolic equivalents; MI, myocardial infarction; PCI, percutaneous coronary intervention; PVD, peripheral vascular disease; RPP, rate-pressure product; TIA, transient ischemic attack.

^a Calculated as (heart rate × systolic BP)/1000.

^b Calculated as maximal RPP minus resting RPP.

^c Calculated as peak exercise heart rate minus heart rate at 1 minute after exercise.

^d Calculated as (peak heart rate – resting heart rate)/[(220 – age) – resting heart rate].

women), the median (IQR) age was also 54 (46-64) years, 63.9% were white, and cardiovascular comorbidities were common (eTable 1 in the Supplement).

Univariable Associations With Increased Mortality

In the CCF cohort, during a median follow-up of 7 (IQR, 4.1-9.6) years, 2521 deaths occurred (4.2% mortality), with 742

deaths in women (3.1% mortality) and 1779 deaths in men (5% mortality). In both sexes, death was associated with increased age, lower body weight, diabetes, hypertension, hyperlipidemia, current or former smoking, CAD, myocardial infarction, percutaneous coronary intervention, coronary artery bypass grafting, chronic obstructive pulmonary disease, stroke or transient ischemic attack, heart failure, ESRD, and periph-

eral arterial disease. In men, death was also associated with a lower body mass index and a family history of coronary disease. With regard to exercise variables, death was associated with ST-segment depression, lower maximal heart rate, lower maximal RPP and Δ RPP, lower peak METs, an abnormal HRR, and an abnormal chronotropic reserve index. A lower DTS was also associated with increased mortality, but no association was found between nonlimiting or limiting chest pain and mortality (eTable 2 in the [Supplement](#)).

Developing the Sex-Specific Risk Scores

In multivariable Cox models, lower peak estimated METs, abnormal HRR, increasing age, lower body weight, current or former smoking, and ESRD were all associated with mortality in men and women (eTable 3 in the [Supplement](#)). In addition, a history of diabetes was associated with mortality in women, whereas a history of heart failure and hypertension were associated with mortality in men. These Cox models showed good calibration at predicting mortality at 10 years (eFigure 1 in the [Supplement](#)). The β coefficients from these Cox proportional hazards regression models were then used to assign points for each covariate (Table 2). The final risk scores for women and men showed high discrimination in estimating mortality in the derivation cohorts (C statistic for women, 0.82; C statistic for men, 0.81).

Kaplan-Meier Survival Analysis

Although the DTS was associated with mortality when assessed as a continuous variable, differentiation of risk was limited when assessed according to the typical DTS categories. In particular, few patients had high-risk DTSs that resulted in wide and overlapping 95% CIs. Similar results were obtained in men and women (Figure 1A and C). Of note, 78 women (0.3%) and 221 men (0.6%) had high-risk DTSs. Conversely, survival curves using the sex-specific risk scores effectively identified patients at highest risk for all-cause mortality (Figure 1B and D), as is also evident in the estimate of 10-year mortality risk according to the sex-specific risk scores (eTables 4 and 5 in the [Supplement](#)).

Validating the Risk Scores

For the sex-specific risk scores, C statistics were similar in the CCF validation cohorts (0.79 for women and 0.81 for men). These sex-specific risk scores also performed better at estimating mortality when compared with the other models in women (C statistic for DTS, 0.70; C statistic for Lauer nomogram, 0.74) (eFigure 2A in the [Supplement](#)) and in men (C statistic for DTS, 0.72; C statistic for Lauer nomogram, 0.75) (eFigure 2B in the [Supplement](#)). Category-free net reclassification improvement and integrated discrimination improvement were also significantly improved with the sex-specific risk scores when compared with the DTS and the Lauer nomogram (Table 3). This improved discrimination primarily corresponded to the correct reclassification of patients who died.

Finally, sex-specific risk scores were calculated for patients in the FIT cohort to assess external validation. At a median follow-up of 10.2 years, there were 6643 deaths in a population of 49 278 (13.5%). The C statistics for the sex-specific risk

Table 2. Sex-Specific Exercise and Clinical Risk Scores for Estimating Mortality

Variable	Points
Women	
Peak treadmill METs	
<5	10
5-7	8
7-10	4
ESRD	5
Weight, kg	
<50	4
50-70	3
70-90	2
Age >65 y	1
Abnormal HRR	2
Former/current smoking	2
Diabetes	1
Men	
Peak treadmill METs	
<8	10
8-10	6
10-12	4
ESRD	7
Weight <80 kg	1
Age, y	
55-65	1
65-75	4
>75	7
Abnormal HRR	3
Former/current smoking	1
History	
Hypertension	2
Heart failure	5

Abbreviations: ESRD, end-stage renal disease; HRR, heart rate recovery; METs, metabolic equivalents.

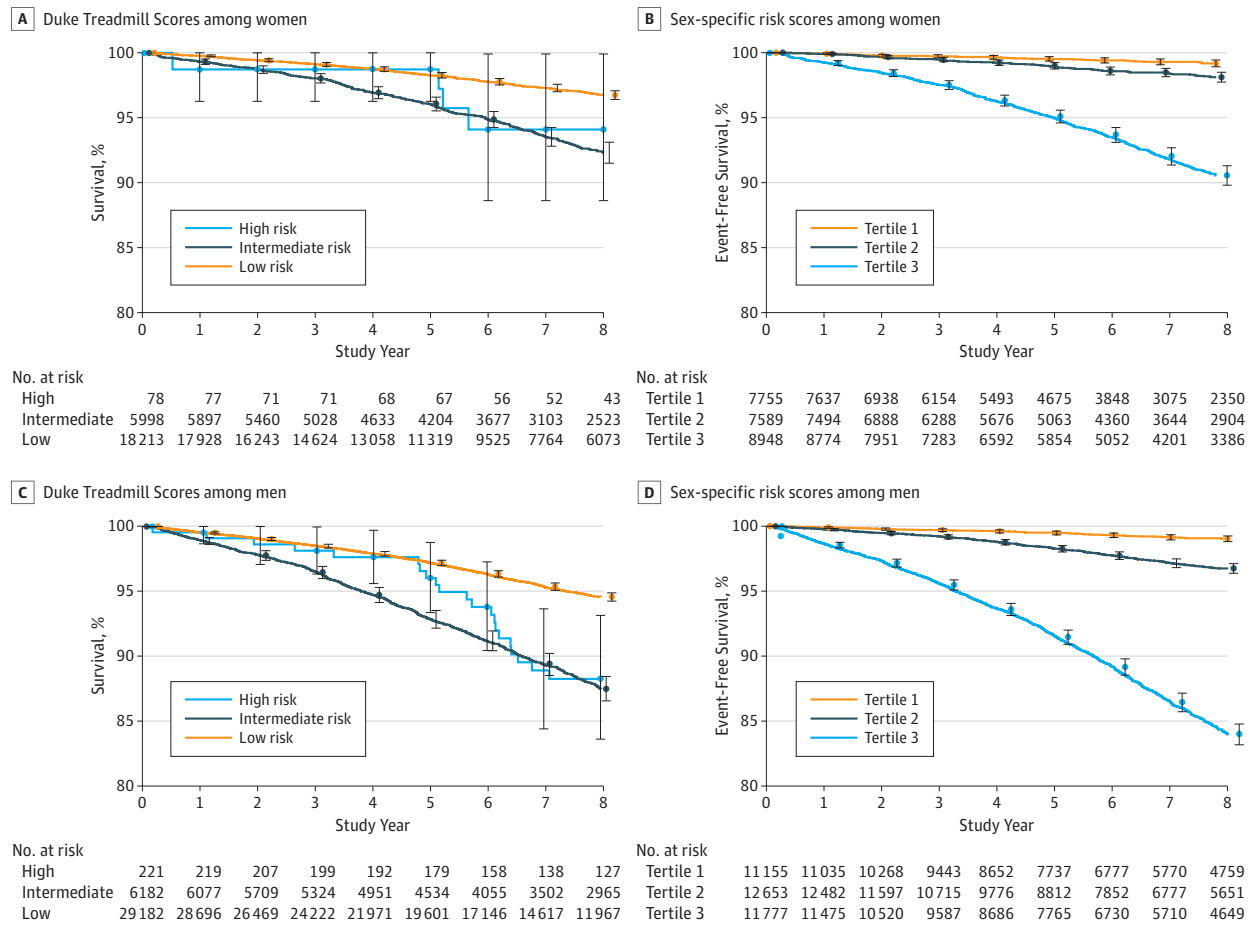
scores were similar in women (0.78) and men (0.79). Regarding calibration, good tracking of observed vs predicted mortality was found (Figure 2).

Discussion

The present study is, to our knowledge, the largest to date to develop sex-specific prognostic risk scores using treadmill testing data. We have demonstrated excellent discrimination and calibration for estimating mortality in the CCF derivation and validation cohorts. Moreover, in the CCF validation cohort, C statistics, net reclassification improvement, and integrated discrimination improvement with these new sex-specific risk scores were improved compared with the DTS and Lauer nomogram. Finally, discrimination and calibration were also reasonable when the sex-specific risk scores were tested externally in the FIT cohort.

A few important observations from our study should be highlighted. First, in patients undergoing treadmill testing, our

Figure 1. Kaplan-Meier Survival Analysis for Women and Men in the Cleveland Clinic Foundation (CCF) Cohort



The CCF cohort included 24 292 women and 35 585 men. Data points indicate median; error bars, 95% CI. Duke Treadmill Scores are stratified as low, intermediate, and high risk. The sex-specific risk scores are stratified into tertiles.

Table 3. Discrimination Analysis for the CCF Cohort Comparing the NRSs With the DTS and Lauer Nomogram in Women and Men

Discrimination Analysis	% (95% CI)	
	Women	Men
IDI: NRSs vs DTS	20 (18.7-21.3)	19 (18.2-19.8)
Category-free NRI		
NRSs vs DTS	75.8 (68.1-83.6)	81.6 (76.3-86.8)
Events correctly reclassified	66.5 (58.9-74)	63.2 (58.1-68.2)
Nonevents correctly reclassified	9.4 (7.5-11.2)	18.4 (16.9-19.9)
IDI: NRSs vs Lauer nomogram	23 (21.7-24.4)	24 (23.2-24.8)
Category-free NRI		
NRSs vs Lauer nomogram	80.1 (72.6-87.6)	87.3 (82.5-92)
Events correctly reclassified	69.2 (61.9-76.5)	72 (67.5-76.5)
Nonevents correctly reclassified	10.9 (9.1-12.7)	15.3 (13.8-16.8)

Abbreviations: CCF, Cleveland Clinic Foundation; DTS, Duke Treadmill Score; IDI, integrated discrimination improvement; NRI, net reclassification improvement; NRSs, sex-specific new risk scores.

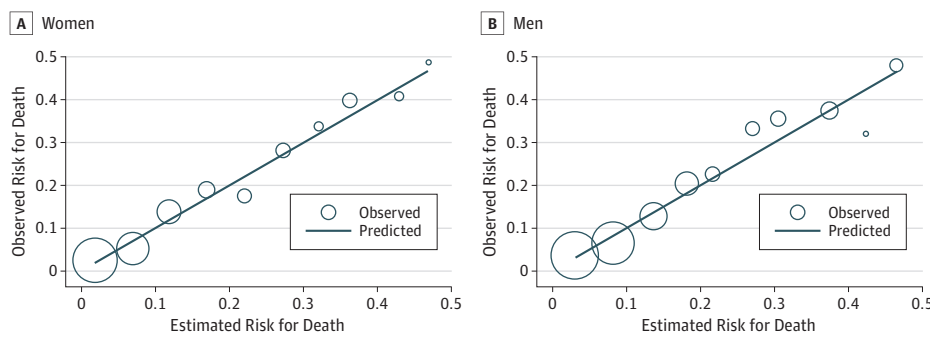
data support separate risk scores according to sex. Certain variables are present or absent in the models for men vs women, and the hazard ratios for risk factors common to both models

differ according to sex. Therefore, rather than simply adjusting for sex, we argue that a sex-specific approach should be considered when assessing the prognosis for patients who undergo exercise testing. In addition, although many variables are associated with mortality and refine risk stratification, decreased exercise capacity is the most important risk factor for men and women. Finally, as shown in our category-free net reclassification improvement, the major advantage of our risk scores is the identification of patients who are likely to have a fatal event. As cardiovascular mortality continues to decline,³ identification of patients with the highest residual risk is increasingly important.

Previous Studies

The DTS remains the most common method to assess prognosis in patients with exercise testing, although this score was developed in higher-risk patients—predominantly middle-aged men—who all had chest pain and invasive coronary angiography.² In a lower-risk and more diverse patient population, our study demonstrates that the prognostic value of the DTS is related solely to the importance of exercise capacity. Chest pain and ST depression with exertion were not associ-

Figure 2. Estimated vs Observed Mortality for Women and Men in the Henry Ford Hospital Exercise Testing (FIT) Cohort



In the FIT cohort (23 386 women, 25 892 men), the sex-specific risk score was slightly altered to exclude abnormal heart rate recovery and to define end-stage renal disease as a glomerular filtration rate of less than 15 mL/min/1.73 m².

ated with mortality in multivariable models. The Lauer nomogram improved on the DTS by incorporating other exercise variables and comorbidities.⁸ However, notable patient populations were excluded. Our objective was to create more comprehensive risk scores with a more inclusive patient population. This approach facilitates a broader clinical use for our sex-specific risk scores.

Limitations

Our study has several notable limitations. First, we assessed all-cause instead of cardiac death, although all-cause mortality may be preferred because it is an unbiased end point. Second, imaging data were not included in our analysis and have been shown to have prognostic importance.^{16,17} However, the focus of this study was to develop risk scores based on clinical and exercise variables alone. Third, all patients in the study underwent evaluation at large referral centers, and generalizability to smaller hospitals may be limited. Finally, because of differences in data collection, the risk scores tested in the FIT cohort were similar, but not identical, to the scores used in the CCF cohort. Discrimination was good in both cohorts, but this observation should not lead to the exclusion of certain variables in risk assess-

ment, especially because HRR has emerged as a risk factor in several studies.^{7,11,12,18} In fact, in a well-developed model, little change may occur in the C statistic when an additional variable is added, even if that variable improves risk stratification.¹⁹

Conclusions

In a large cohort of patients who underwent treadmill testing, we have demonstrated a differential effect of exercise variables and clinical risk factors on overall mortality according to sex. The sex-specific risk scores outperform previous risk stratification tools and help to identify patients at the highest risk for death. To facilitate clinical use of these sex-specific risk scores, we have developed an online calculator to estimate 10-year mortality (<http://www.clevelandclinic.org/lp/hvi-tools/10YearMortality.html>). Even when accounting for multiple comorbidities, exercise capacity was still the predominant risk factor in men and women. This online calculator can be used by physicians and patients to not only assess prognosis but also emphasize the importance of exercise, even in the presence of other cardiovascular risk factors.

ARTICLE INFORMATION

Accepted for Publication: August 11, 2016.

Published Online: October 26, 2016.
doi:10.1001/jamacardio.2016.3720

Author Contributions: Drs Cremer and Cho had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.
Study concept and design: Cremer, Ahmed, Pierson, Al-Mallah, Cho.

Acquisition, analysis, or interpretation of data: Cremer, Wu, Ahmed, Pierson, Brennan, Al-Mallah, Brawner, Ehrman, Keteyian, Blumenthal, Blaha.
Drafting of the manuscript: Cremer, Ahmed, Pierson, Keteyian, Cho.

Critical revision of the manuscript for important intellectual content: Cremer, Wu, Ahmed, Pierson, Brennan, Al-Mallah, Brawner, Ehrman, Blumenthal, Blaha.

Statistical analysis: Cremer, Wu, Ahmed, Brennan, Cho.

Administrative, technical, or material support:

Al-Mallah, Keteyian, Blaha.

Study supervision: Al-Mallah, Blumenthal, Blaha.

Conflict of Interest Disclosures: All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest and none were reported.

Funding/Support: This study was supported by the Women's Cardiovascular Center, Cleveland Clinic, and Karo's Chair for Women's Cardiovascular Research.

Role of the Funder/Sponsor: The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

REFERENCES

1. Gibbons RJ, Balady GJ, Bricker JT, et al. ACC/AHA 2002 guideline update for exercise testing: a report of the American College of Cardiology/American

Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol.* 2002;40(8):1531-1540.

2. 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(6):793-800.

3. Mozaffarian D, Benjamin EJ, Go AS, et al; American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics—2015 update: a report from the American Heart Association. *Circulation.* 2015; 131(4):e29-e322.

4. Gulati M, Arnsdorf MF, Shaw LJ, et al. Prognostic value of the Duke Treadmill Score in asymptomatic women. *Am J Cardiol.* 2005;96(3):369-375.

5. 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(11): 1392-1398.

6. Sadrzadeh Rafie AH, Dewey FE, Sungar GW, et al. Age and double product (systolic blood pressure × heart rate) reserve-adjusted modification of the Duke Treadmill Score nomogram in men. *Am J Cardiol.* 2008;102(10):1407-1412.
7. Maddox TM, Ross C, Ho PM, et al. The prognostic importance of abnormal heart rate recovery and chronotropic response among exercise treadmill test patients. *Am Heart J.* 2008;156(4):736-744.
8. Lauer MS, Pothier CE, Magid DJ, Smith SS, Kattan MW. An externally validated model for predicting long-term survival after exercise treadmill testing in patients with suspected coronary artery disease and a normal electrocardiogram. *Ann Intern Med.* 2007;147(12):821-828.
9. Daugherty SL, Magid DJ, Kikla JR, et al. Gender differences in the prognostic value of exercise treadmill test characteristics. *Am Heart J.* 2011;161(5):908-914.
10. Mosca L, Barrett-Connor E, Wenger NK. Sex/gender differences in cardiovascular disease prevention: what a difference a decade makes. *Circulation.* 2011;124(19):2145-2154.
11. 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(18):1351-1357.
12. 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(16):1911-1916.
13. Al-Mallah MH, Keteyian SJ, Brawner CA, Whelton S, Blaha MJ. Rationale and design of the Henry Ford Exercise Testing Project (the FIT project). *Clin Cardiol.* 2014;37(8):456-461.
14. Newman TB, Brown AN. Use of commercial record linkage software and vital statistics to identify patient deaths. *J Am Med Inform Assoc.* 1997;4(3):233-237.
15. Pencina MJ, D'Agostino RB Sr, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med.* 2011;30(1):11-21.
16. Hachamovitch R, Berman DS, Shaw LJ, et al. Incremental prognostic value of myocardial perfusion single photon emission computed tomography for the prediction of cardiac death: differential stratification for risk of cardiac death and myocardial infarction. *Circulation.* 1998;97(6):535-543.
17. Marwick TH, Case C, Vasey C, Allen S, Short L, Thomas JD. Prediction of mortality by exercise echocardiography: a strategy for combination with the Duke Treadmill Score. *Circulation.* 2001;103(21):2566-2571.
18. Dhoble A, Lahr BD, Allison TG, Kopecky SL. Cardiopulmonary fitness and heart rate recovery as predictors of mortality in a referral population. *J Am Heart Assoc.* 2014;3(2):e000559.
19. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation.* 2007;115(7):928-935.

Invited Commentary

Can We Improve Mortality Estimation in Women After Treadmill Testing by Using Sex-Specific Scores?

Ashok Krishnaswami, MD, MAS; Jeff W. Christle, MA; Victor Froelicher, MD

The use of predictive analytics in modern cardiology has had a significant impact in decreasing the subjectivity of forecasting cardiovascular events. The abundance of currently available clinical prediction models (CPMs) has been demonstrated by a recent systematic review.¹

This review unearthed 796 scientific articles on the topic of CPMs and cardiovascular disease published from 1990 to 2012, with 90% being novel and the remainder recalibration or other adaptations of prior CPMs. Although utilization of CPMs is currently low, it promises to decrease use of the routine subjective eyeball test.²

To establish an anatomical diagnosis of coronary artery disease and to estimate the probability of future cardiac events is a difficult task. This concept is not dissimilar to the Heisenberg “uncertainty principle” in which a particle’s position and momentum cannot be known with complete precision.³ Therefore, working in this arena requires an acceptance of some uncertainty. The exercise treadmill test has been recommended as a routine first-line test for individuals capable of exercising with no contraindications and without an abnormal resting electrocardiographic finding.⁴ However, concern about its low sensitivity and specificity (leading to false-positive findings), owing to an undue focus on the ST segment often results in the use of additional imaging modalities (myocardial perfusion or echocardiography). Advanced CPMs have come to the rescue and have been shown to be superior to subjective prognostication.⁵

They have improved the accuracy of diagnosis of coronary artery disease and the prognosis of future cardiovascular events, reduced costs, and aided the decision-making process.^{6,7}

In this issue of *JAMA Cardiology*, Cremer et al⁸ have tied together 3 important areas in cardiovascular medicine: the use of CPMs, sex-specific research,⁹ and treadmill testing. The investigators used the Cleveland Clinic Foundation (CCF) cohort for derivation and internal validation and the Henry Ford Exercise Testing (FIT) cohort for external validation. The study examined the hypothesis that sex-specific CPMs are superior to general models that have been risk adjusted for sex, such as the Lauer score.¹⁰ The models were tested using standard discrimination and calibration statistics. Patients in the CCF cohort underwent symptom-limited treadmill testing using multiple treadmill protocols. Comorbidities were obtained at the time of the baseline treadmill test. Subsequent prognostic variables, such as the Duke Treadmill Score, chronotropic reserve index, and rate-pressure product, were then calculated using standard treadmill variables that were also obtained at the time of baseline testing. The primary outcome of all-cause mortality was determined by the Social Security Death Index Master File. Covariate β coefficients were obtained using time-to-event analysis (Cox proportional hazards regression model) with a backward-stepwise covariate selection that was subsequently used to develop a linear score.

The study findings demonstrated that the women referred for exercise testing were younger, were less often white,