Importance of Treadmill Exercise Time as an Initial Prognostic Screening Tool in Patients With Systolic Left Ventricular Dysfunction

Eileen Hsich, Eiran Z. Gorodeski, Randall C. Starling, Eugene H. Blackstone, Hemant Ishwaran and Michael S. Lauer

Circulation published online Jun 15, 2009;
DOI: 10.1161/CIRCULATIONAHA.109.848382
Circulation is published by the American Heart Association. 7272 Greenville Avenue, Dallas, TX 72514
 Print ISSN: 0009-7322. Online ISSN: 1524-4539

Importance of Treadmill Exercise Time as an Initial Prognostic Screening Tool in Patients With Systolic Left Ventricular Dysfunction

Eileen Hsich, MD; Eiran Z. Gorodeski, MD, MPH; Randall C. Starling, MD, MPH; Eugene H. Blackstone, MD; Hemant Ishwaran, PhD; Michael S. Lauer, MD

Background—We sought to determine whether treadmill exercise time may be of value as an initial prognostic screening tool in ambulatory patients with impaired systolic function who are referred for cardiopulmonary exercise testing.

Methods and Results—We studied 2231 adult systolic heart failure patients (27% of whom were women) who underwent cardiopulmonary stress testing using a modified Naughton protocol. We assessed the value of treadmill exercise time for prediction of all-cause death and a composite of death or United Network for Organ Sharing status 1 heart transplantation. During a mean follow-up of 5 years, 742 patients (33%) died. There were 249 United Network for Organ Sharing status 1 heart transplants (11%). Treadmill exercise time was predictive of death and the composite outcome in both women and men, even after accounting for peak oxygen consumption and other clinical covariates (adjusted hazard ratio of lowest versus high sex-specific quartile for prediction of death 1.70, 95% confidence interval 1.05 to 2.75, $P=0.03$; for prediction of the composite outcome, 1.75, 95% confidence interval 1.15 to 2.66, $P=0.009$). For a 1-minute change in exercise time, there was a 7% increased hazard of death (eg, comparing 480 to 540 seconds, hazard ratio $=1.07$, 95% confidence interval 1.02 to 1.12, $P=0.004$).

Conclusions—Because cardiopulmonary stress testing is not available in every hospital, treadmill exercise time with a modified Naughton protocol may be of value as an initial prognostic screening tool. (***Circulation*. 2009;119:3189-3197.**)

Key Words: heart failure ■ exercise ■ sex ■ prognosis

Peak oxygen consumption ($\dot{V}O_2$) remains one of the most powerful single predictors of mortality for heart failure patients with severe systolic left ventricular dysfunction.[1–3] Current guidelines suggest ambulatory patients be considered for transplantation when the peak $\dot{V}O_2$ is ≤ 14 mL · kg^{-1}min^{-1}, or ≤ 12 mL · kg^{-1}min^{-1} in the setting of β-blockade.[3] Although cardiopulmonary exercise testing routinely is used to determine candidacy for heart transplantation, it is not known whether a simple measurement of treadmill exercise time may be of comparable value as an initial prognostic screening tool. We sought to evaluate whether (1) treadmill exercise time predicts survival in heart failure patients with systolic dysfunction, (2) treadmill exercise time adds incremental prognostic value beyond established risk factors, including peak $\dot{V}O_2$, and (3) a model with treadmill exercise time in addition to established risk factors better classifies risk than a model with established risk factors alone.

Editorial see p 3165
Clinical Perspective on p 3197

Methods

We performed an observational prospective cohort study at the Cleveland Clinic involving consecutive patients with left ventricular ejection fraction <40% who underwent cardiopulmonary stress testing between August 1997 and April 2007. We focused specifically on patients who were tested according to the modified Naughton protocol, which was the most common protocol used in our laboratory for this type of patient. Patients were excluded if they were <18 years old or had no US Social Security number. The study was approved by the institutional review board at the Cleveland Clinic, and because all data were collected and recorded as part of routine clinical care, the requirement for informed consent was waived.

Clinical Data

All demographic information, medications, medical and surgical history, heart rate, blood pressure, directly measured height and

Continuing medical education (CME) credit is available for this article. Go to http://cme.ahajournals.org to take the quiz.
Received January 2, 2009; accepted April 15, 2009.

From the Departments of Cardiovascular Medicine (E.H., E.Z.G., R.C.S.), Quantitative Health Sciences (H.I.), and Cardiothoracic Surgery (E.H.B.) at the Cleveland Clinic and Case Western Reserve University School of Medicine (E.H., R.C.S.), Cleveland, Ohio; and the Division of Prevention and Population Sciences (M.S.L.), National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Md.

The online-only Data Supplement is available with this article at http://circ.ahajournals.org/cgi/content/full/CIRCULATIONAHA.109.848382/DC1.

Reprint requests to Michael S. Lauer, MD, FACC, FAHA, Director, Division of Prevention and Population Science, National Heart, Lung, and Blood Institute, National Institutes of Health, Rockledge Center II, 6701 Rockledge Dr, Room 10122, Bethesda, MD 20892. E-mail lauerm@nhlbi.nih.gov

Circulation **is available at http://circ.ahajournals.org** **DOI: 10.1161/CIRCULATIONAHA.109.848382**

weight, and stress results were obtained at the time of stress testing and prospectively recorded in our electronic database. Biventricular pacemakers were not reported separately from other pacemakers during database entry. Glomerular filtration rate was estimated with the Cockcroft-Gault equation.[4] Left ventricular ejection fraction was determined either by echocardiography, left ventriculography, or ECG-gated myocardial perfusion imaging. If a patient underwent >1 metabolic stress test with a modified Naughton protocol, only the first test was considered. Serum laboratory tests within 3 months of the cardiopulmonary study were included, and only the laboratory tests closest to the stress test date were used.

Cardiopulmonary Stress Testing

Cardiopulmonary stress testing was symptom limited, and patients were strongly advised to not use the handrails for support. Details on our laboratory's protocols have been published previously.[5] Briefly, the test was performed with a modified Naughton protocol that increases workload ≈1 metabolic equivalent every 2 minutes.[6] Results were recorded on a MedGraphics cardiopulmonary system (Medical Graphics Corp, St Paul, Minnesota). Oxygen consumption, carbon dioxide production, minute ventilation, tidal volume, heart rate, blood pressure, and respiratory rate were obtained at rest, every 30 seconds during exercise, and during recovery. Total duration of exercise, henceforth referred to as treadmill exercise time, was recorded to the nearest second.

End Points

The primary end points were all-cause mortality and a composite of all-cause mortality or United Network for Organ Sharing (UNOS) status 1 heart transplantation during a mean follow-up of 5 years (maximum for survivors, 11 years). Mortality data were obtained by linking our database with the US Social Security Administration Death Index. We validated this approach previously, yielding a sensitivity of 97%,[7] which is equivalent to a 97% follow-up rate. UNOS status 1 heart transplantation was based on Organ Procurement and Transplantation Network data as of June 18, 2008. All events (death or heart transplantation) were censored as of April 1, 2008.

Statistical Analysis

Baseline characteristics were reported as sex-specific quartiles of treadmill exercise time. Continuous variables were expressed as means with SDs, and categorical variables were expressed as frequencies. Serum laboratory tests before October 1999 were systematically unavailable on our electronic database. We used informed imputation to fill in these missing values by constructing regression models for each laboratory measure on all other characteristics of our existing database except for outcome (10% of creatinine clearance, blood urea nitrogen, glucose, and sodium values were imputed, and 15% of serum hemoglobin values were imputed).

We generated Kaplan–Meier plots relating treadmill exercise time and cumulative mortality. We constructed unadjusted and adjusted Cox proportional hazards models to examine the association of treadmill exercise time and outcomes. Possible nonlinear associations were tested with restricted cubic splines. Potentially important interactions were tested. The proportional hazards assumption was tested by scaled Schoenfeld residuals and inspection of hazard ratio plots.

Model discrimination was assessed by calculating the out-of-bag (OOB) c-index for time to event outcomes, an approach that we have used previously.[8] The OOB method involves obtaining bootstrap samples from the original cohort and using each sample to compute a prediction model. Each bootstrap sample left out ≈37% of the data, which were referred to as the OOB data. The prediction model was applied to the OOB data to calculate the OOB c-index, a measure that is conceptually similar to the area under a receiver operating curve. The purpose of this method is to perform multiple internal validations with an internal cohort that has been randomly sampled multiple times. This is a more conservative measure than what we have reported in previous work, which was based on 1 assessment of the original data.[9] Hence, we would expect that these OOB c-index values would be lower than what we reported previously, because they represent more conservative estimates.

To determine the change in prediction error attributable to each variable, we recalculated the prediction error after random permutation of that variable in the OOB data; a variable with a high degree of importance would be expected to yield a greater change in the OOB c-index. The process was repeated 100 times for each variable.

To address clinical utility, we constructed risk-reclassification figures. We compared 1-year predicted risk estimates based on models of established risk factors with and without the variables in question. We selected categories of risk of 15% based on the current 1-year mortality outcomes after cardiac transplantations.[10] We used 2 objective methods to quantify improvement in categories as suggested by Pencina et al[11]: Net reclassification improvement and integrated discrimination improvement. The net reclassification improvement and integrated discrimination improvement were calculated by comparing an individual patient's predicted risks at 1 year based on a Cox model 1 versus another Cox model 2. These predicted risks were calculated with a modification of the R function *predict*(), with censoring at 1 year.

Statistical analysis was performed with SAS version 9.1.3 (SAS Institute Inc, Cary, NC) and R version 2.6.2 (www.R-project.org). We used Harrell's Design and Hmisc libraries for model construction and assessment, an R macro written by 1 of the authors (HI) for the OOB and change in prediction error analyses, and another R macro written by Lauer and colleagues[11a] to calculate the observed risks in the reclassification figures. We used an SAS macro written by Pencina et al to calculate the net reclassification improvement and integrated discrimination improvement.[11]

The authors had full access to and take full responsibility for the integrity of the data. All authors have read and agree to the manuscript as written.

Results

The study cohort consisted of 602 women (27%) and 1629 men (73%). During a mean follow-up of 5 years (maximum for survivors, 11 years), 155 women (26% of the female cohort) and 587 men (36% of the male cohort) died. There were 249 patients (16% women) who underwent heart transplantation as UNOS status 1. Only 4 patients underwent 2 transplantations.

Table 1 shows the baseline characteristics of the 2231 patients according to sex-specific quartiles for treadmill exercise time. Compared with those in the highest quartile, those in the lowest quartile were older; were more likely to have coronary artery disease, diabetes mellitus, implantable cardioverter-defibrillators (ICDs), or pacemakers; and had lower systolic blood pressures, with higher resting heart rates. They also had lower peak $\dot{V}O_2$ and lower peak respiratory exchange ratio. They were less likely to be taking a β-blocker or an ACE inhibitor and more likely to be taking nitrates, hydralazine, angiotensin receptor blockers, diuretics, and antiarrhythmic agents. Laboratory test results were similar, except those in the lowest quartile had higher blood urea nitrogen and lower creatinine clearance.

Outcomes

Figure 1 shows the association of sex-specific quartiles of treadmill exercise time with all-cause mortality. The composite of death and UNOS status 1 heart transplantation yielded similar sex-specific results (online-only Data Supplement Figure I). In both women and men, treadmill exercise time predicted survival (death or death/UNOS status 1 transplantation), with the worst survival in the quartile with the lowest

Table 1. Baseline Characteristics According to Sex-Specific Quartiles for Treadmill Exercise Time

	Quartile 1 (n=553)	Quartile 2 (n=588)	Quartile 3 (n=524)	Quartile 4 (n=566)	*P*
Treadmill time: Women (n=602), range in seconds	21–315	317–478	480–600	603–1320	
Treadmill time: Men (n=1629), range in seconds	35–357	360–480	484–657	660–1415	
Age, y	57±10	56±10	54±11	49±11	<0.001
Body mass index, kg/m^2	29±6	28±6	29±6	28±5	0.20
Current smokers, n (%)	112 (20)	117 (20)	108 (21)	122 (22)	0.91
Diabetes mellitus: insulin treated, n (%)	74 (13)	78 (13)	42 (8)	21 (4)	<0.001
Diabetes mellitus: not insulin treated, n (%)	111 (20)	107 (18)	84 (16)	48 (8)	<0.001
Coronary artery disease, n (%)	267 (48)	278 (47)	210 (40)	151 (27)	<0.001
Previous Q-wave MI, n (%)	68 (12)	63 (11)	76 (15)	72 (13)	0.30
Previous CABG, n (%)	187 (34)	192 (33)	131 (25)	84 (15)	<0.001
Previous PCI, n (%)	135 (24)	147 (25)	123 (23)	71 (13)	<0.001
ICD, n (%)	209 (38)	180 (31)	159 (30)	99 (17)	<0.001
Pacemaker, n (%)	160 (29)	149 (25)	117 (22)	76 (13)	<0.001
Medication use, n (%)					
β-Blocker	349 (63)	357 (61)	345 (66)	378 (67)	0.13
ACE inhibitor	378 (68)	443 (75)	422 (81)	468 (83)	<0.001
Angiotensin receptor blocker	86 (16)	92 (16)	56 (11)	56 (10)	0.003
Potassium-sparing diuretics	181 (33)	186 (32)	152 (29)	130 (23)	0.001
Antiarrhythmic	159 (29)	152 (26)	117 (22)	81 (14)	<0.001
Anticoagulation	253 (46)	255 (43)	219 (42)	172 (30)	<0.001
Aspirin	261 (47)	270 (46)	251 (48)	256 (45)	0.81
Digoxin	376 (68)	425 (72)	361 (69)	408 (72)	0.28
Nitrates	212 (38)	223 (38)	162 (31)	142 (25)	<0.001
Hydralazine	44 (8)	40 (7)	29 (6)	23 (4)	0.041
Loop diuretics	502 (91)	530 (90)	447 (85)	401 (71)	<0.001
Thiazide diuretics	90 (16)	98 (17)	53 (10)	38 (7)	<0.001
Statin	210 (38)	266 (45)	187 (36)	187 (33)	<0.001
Calcium channel blocker: Nondihydropyridine	6 (1)	2 (0)	4 (1)	4 (1)	0.52
Calcium channel blocker: dihydropyridine	24 (4)	33 (6)	31 (6)	11 (2) 0.005	0.005
Resting heart rate, bpm	80±14	78±14	75±14	72±14	<0.001
Resting systolic blood pressure, mm Hg	108±18	110±18	111±18	112±17	<0.001
Left ventricular ejection fraction, %	19±8	20±7	21±8	21±7	<0.001
Peak oxygen consumption, mL · kg^{-1} · min^{-1}	11±3	14±2	17±3	22±4	<0.001
Peak respiratory exchange ratio	1.03±0.16	1.07±0.11	1.10±0.09	1.11±0.07	<0.001
Serum sodium, mmol/L	139±4	139±3	140±3	140±3	<0.001
Creatinine clearance, mL/min	78±37	82±41	94 +/−43	110±44	<0.001
Serum BUN, mg/dL	31±16	28±13	23±10	19±7	<0.001
Serum hemoglobin, g/dL	13±2	13±1	14±1	14±1	<0.001
Serum glucose, mg/dL	112±44	115±46	111±47	100±30	<0.001

n indicates total number of women and men per quartile; MI, myocardial infarction; CABG, coronary artery bypass grafting; PCI, percutaneous coronary intervention; and BUN, blood urea nitrogen.

treadmill exercise time. After adjustment for all the variables listed in Table 1, including peak $\dot{V}O_2$, in a Cox proportional hazards model, a shorter exercise time remained associated with an increase hazard of death (Table 2) and of death and UNOS status 1 heart transplantation (online-only Data Supplement Table I). For a 1-minute change in exercise time, there was a 7% increased hazard of death (eg, for a comparison of 480 to 540 seconds, hazard ratio 1.07, 95% confidence interval [CI] 1.02 to 1.12, P=0.004). To determine whether the imputed laboratory values affected our data, we performed a sensitivity analysis leaving out all laboratory variables and found similar results (hazard ratio 1.08, 95% CI 1.03 to 1.13, P=0.002).

We further evaluated the association of treadmill exercise time and survival for the subgroup of patients with peak $\dot{V}O_2$ $\geq$14 mL · kg^{-1}min^{-1}. Even among these lower-risk patients, lower treadmill exercise time predicted worse outcomes in both women and men (Figure 2; online-only Data Supple-

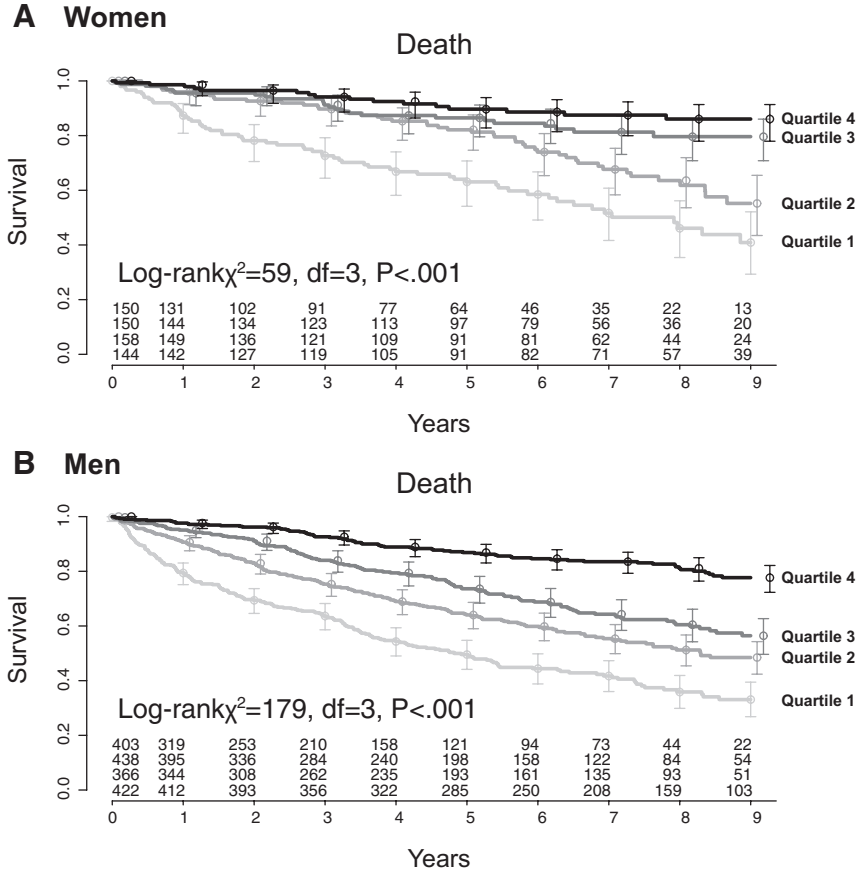


Figure 1. Kaplan–Meier plots for (A) women and (B) men, stratified by quartiles of treadmill exercise time. A, Women (n=602): Quartile 1, 21 to 315 seconds; quartile 2, 317 to 478 seconds; quartile 3, 480 to 600 seconds; quartile 4, 603 to 1320 seconds. B, Men (n=1629): Quartile 1, 35 to 357 seconds; quartile 2, 360 to 480 seconds; quartile 3, 484 to 657 seconds; quartile 4, 660 to 1415 seconds.

ment Figure II). In this subset of patients with high peak $\dot{V}O_2$, a 1-minute change in exercise time also yielded a 7% increased hazard of death in a fully adjusted model (ie, for a comparison of 480 to 540 seconds, hazard ratio 1.07, 95% CI 1.00 to 1.14, $P=0.04$).

The c-indices of the models that contained all variables including exercise time were 0.723 for all-cause death and 0.744 for the composite of death and UNOS status 1 heart transplantation. The OOB c-indices, a more conservative measure, of the models that contained all variables including treadmill exercise time were 0.698 (95% CI 0.670 to 0.726) for all-cause death (Table 3) and 0.727 (95% CI 0.703 to 0.751) for the composite outcome of death and UNOS status 1 heart transplantation (online-only Data Supplement Table II), which indicates moderate discriminatory ability. There were no notable differences between the models that contained clinical variables and either peak $\dot{V}O_2$, treadmill exercise time, or both. Treadmill exercise time was among the top 3 most important contributors to discriminative prediction in both sets of models (Figure 3; online-only Data Supplement Figure III).

Table 2. Treadmill Exercise Time and Outcome: Cox Proportional Hazards Analyses

Model: Death	Hazard Ratio (95% CI)	*P*
Treadmill exercise time as a continuous variable*		
Unadjusted	2.23 (2.00–2.49)	<0.0001
Adjusted for age and sex	2.23 (1.98–2.50)	<0.0001
Adjusted for age, sex, history of CAD, and peak $\dot{V}O_2$	1.54 (1.23–1.92)	0.0005
Multivariable adjusted	1.40 (1.12–1.76)	0.004
Treadmill exercise time as a dichotomous variable†		
Unadjusted	4.91 (3.86–6.24)	<0.0001
Multivariable adjusted	1.70 (1.05–2.75)	0.03

CAD indicates coronary artery disease.

*Comparisons are between the 25th percentile (317 seconds in women, 360 seconds in men) and 75th percentile (600 seconds in women, 657 seconds in men).

†Comparisons between quartile 1 (<316 seconds in women, <358 seconds in men) and quartile 4 (>602 seconds in women, >659 seconds in men).

To better compare model performance within clinical categories, we classified patients into low-risk (<15%) and high-risk (≥15%) categories of 1-year risk, with these cut points based on the observed national 1-year survival after heart transplantation. We compared models with and without treadmill exercise time by cross-classifying predicted risks, stratified by whether or not the patients died (Figure 4) or developed the composite end point of death or transplantation (online-only Data Supplement Figure IV). If an additional variable were to add no predictive value, all points would fall on the line of identity. A spread around the line indicates modulation of predicted risk; if the variable correctly modulates predicted risk, there should be a greater preponderance of events above the line of identity. Clinical variables and

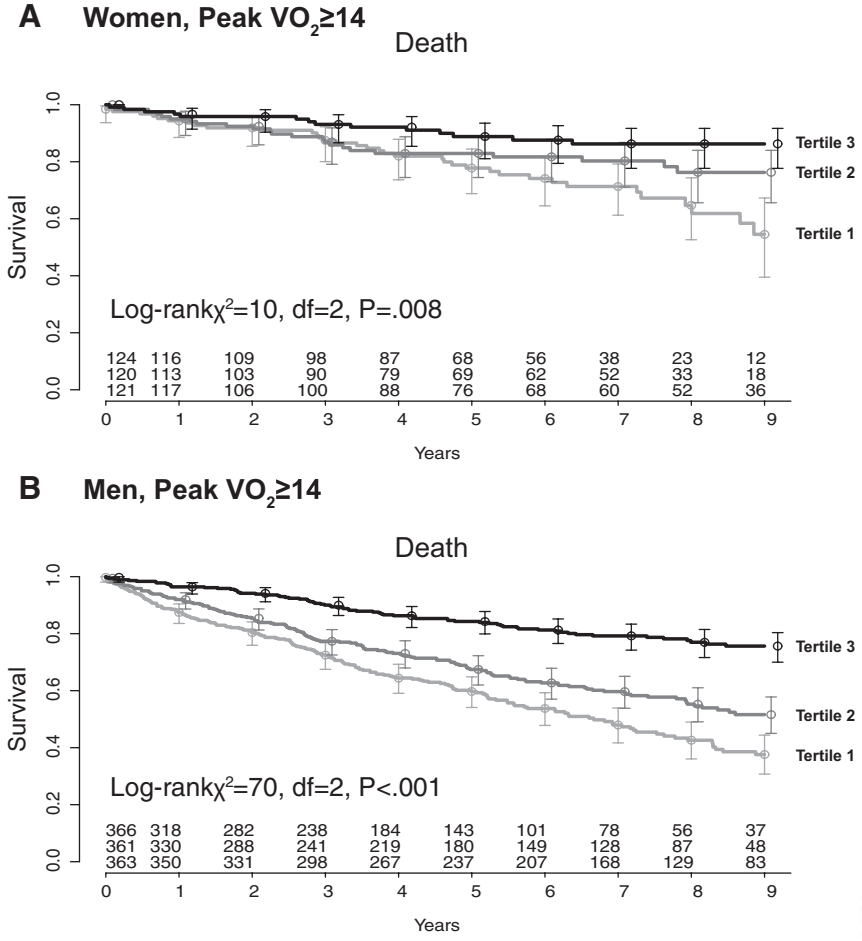


Figure 2. Kaplan–Meier plots for (A) women and (B) men with peak $\dot{V}O_2 \geq 14$, stratified by tertiles of increasing treadmill exercise time. A, Women: Tertile 1, <518 seconds; tertile 2, 518 to 639 seconds; tertile 3, ≥640 seconds. B, Men: Tertile 1, <511 seconds; tertile 2, 511 to 682 seconds; tertile 3, ≥683 seconds.

usage of peak $\dot{V}O_2$ (Figure 4A) or treadmill exercise time (Figure 4B) improved net reclassification for the outcome of death and the composite outcome (online-only Data Supplement Figures IV-A and IV-B). However, there were no significant differences in net reclassification between a model with clinical variables and treadmill exercise time versus clinical variables and peak $\dot{V}O_2$ (Figure 4C; online-only Data Supplement Figure IV-C).

The integrated discrimination improvement (a reclassification measure that is not limited to predetermined categories of risk) for a model with treadmill exercise time was 1.9% ($P<0.0001$; Figure 4B) for the outcome of death and 3.4% ($P<0.001$; online-only Data Supplement Figure IV-B) for the composite outcome. There were no significant differences in integrated discrimination improvement between a model with clinical variables and treadmill exercise time versus clinical variables and peak $\dot{V}O_2$ (Figure 4C; online-only Data Supplement Figure IV-C).

Table 3. OOB Concordance Index Values for Cox Regression Models

	All-Cause Death: OOB Concordance Index (95% CI)
Clinical variables only	0.672 (0.646–0.698)
Peak $\dot{V}O_2$ only	0.675 (0.662–0.688)
Time only	0.666 (0.651–0.681)
Peak $\dot{V}O_2$+time only	0.679 (0.666–0.692)
Clinical variables+peak $\dot{V}O_2$	0.699 (0.673–0.725)
Clinical variables+time	0.697 (0.671–0.723)
Clinical variables+peak $\dot{V}O_2$+time	0.698 (0.670–0.726)

Discussion

In a large cohort of patients with impaired left ventricular systolic function who underwent the same cardiopulmonary stress testing protocol, we found that treadmill exercise time predicted survival and yielded similar prognostic value to peak oxygen consumption (peak $\dot{V}O_2$). Even among low-risk patients with a peak $\dot{V}O_2 > 14$ mL · kg^{-1}min^{-1}, those who had a lower treadmill exercise time had markedly worse outcomes. The present findings are consistent with the hypothesis that exercise capacity is a valuable initial prognostic screening tool in patients with systolic left ventricular dysfunction.

The present study expands on previous work demonstrating the prognostic power of exercise capacity.[12,13] Exercise capacity (ie, treadmill time or "estimated" metabolic equivalents) is predictive of survival in both healthy adults and those with known coronary artery disease.[14–16] Exercise tolerance reflects a number of prognostically important factors, including cardiac function, endothelial function, pulmonary function, oxygen-carrying capacity, and autonomic nervous system balance; however, its usefulness as a prognostic

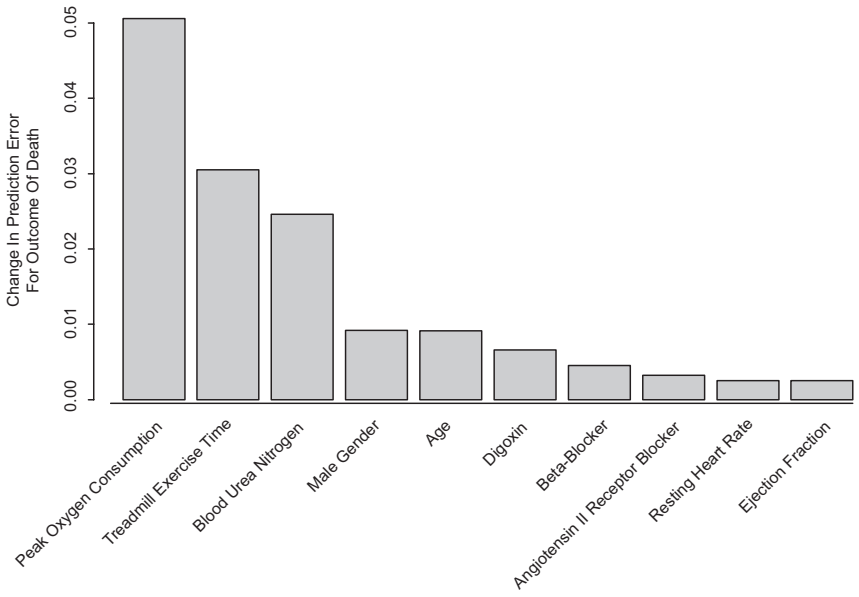


Figure 3. Change in OOB-determined prediction error. Only the 10 most important variables are shown. Results are based on 100 bootstrapped samples.

screening tool in patients with systolic left ventricular dysfunction is less established.

Treadmill exercise time has not been used in most heart failure survival models,[1,17–19] but it has been explored previously in at least 3 heart failure studies. In the Captopril-Digoxin Research Group report, exercise time on a modified Naughton protocol was deemed to have no significant prognostic value; however, the cohort was relatively healthy, because patients who exercised less than 240 seconds were excluded, and the average exercise time was quite high at 558 seconds.[20] In a small study of 60 heart failure patients,[21] there was a high degree of correlation between treadmill exercise time and peak $\dot{V}O_2$; however, owing to the small sample size, that study did not have the power to determine prognostic significance. More recently, a cohort of 31 children with idiopathic cardiomyopathy who underwent cardiopulmonary stress testing with a modified Naughton protocol were followed up prospectively for a median of 1282 days, and 20 of them died or underwent heart transplantation.[22] Both univariable and multivariable analyses showed treadmill exercise time to be predictive of outcome.[22]

Can a simple exercise stress test replace cardiopulmonary stress testing for patients being evaluated for heart transplantation? Possibly, but we do not advocate this approach. By both OOB change in c-index and integrated discrimination improvement, we found that peak $\dot{V}O_2$ is a more powerful predictor of mortality than treadmill exercise time. Other researchers have also found measured peak $\dot{V}O_2$ to be a better predictor than either estimated peak $\dot{V}O_2$ or work rate.[23,24] Another advantage of cardiopulmonary stress testing is the ability to measure patient effort (ie, respiratory exchange ratio). When patients achieve an adequate level of effort, peak $\dot{V}O_2$ is a reliable and reproducible measurement and does not depend on the use of only 1 exercise protocol. Exercise time may not be as reproducible, based on the results of HF-ACTION (Heart Failure and A Controlled Trial Investigation Outcomes of exercise traiNing), which repeated within 7 days a baseline exercise study in 401 patients with systolic heart failure (87% of whom used a modified Naughton protocol); the study noted that the peak $\dot{V}O_2$ was unchanged, but exercise time increased by 26 seconds on average.[25] Cardiopulmonary stress testing can also help identify noncardiac causes for shortness of breath.

However, cardiopulmonary stress testing is not available in every hospital, and therefore, treadmill exercise time may be a reasonable screening tool for physicians to determine prognosis and need for referral to advanced heart failure treatment centers. On the basis of the present data, treadmill exercise times of less than 5 minutes 17 seconds for women and less than 6 minutes for men may be useful cutoff values when a modified Naughton protocol is used, because in the present study cohort, this corresponded to approximately a 15% 1-year mortality rate. Future research in other cohorts will be needed to verify our findings.

The strengths of the present study include a large sample size with a large number of hard events, prospective data collection, adequate power to present sex-specific results,[26,27] and the recording of treadmill exercise time to the nearest second. All end points were determined by query of databases outside our center. Deaths were determined by Social Security files, whereas UNOS 1 transplantation was determined by direct query of UNOS files. Of note, 20% of patients who experienced UNOS 1 transplantation underwent transplantation outside our center; thus, an ascertainment bias based on consideration of only events known to Cleveland Clinic information systems was avoided. Body mass index was calculated on the basis of direct measures of height and weight. We adjusted for >35 clinical/demographic variables, which included sex, peak $\dot{V}O_2$, and coronary artery disease, and still found a longer exercise time interval to be associated with a lower hazard of death. We also used both traditional (c-index) and contemporary statistical discrimination and reclassification methods[11] to assess model performance.

The c-index values reported here are lower than those we reported previously for a similar cohort.[9] We now have adopted a more sophisticated and conservative method (OOB

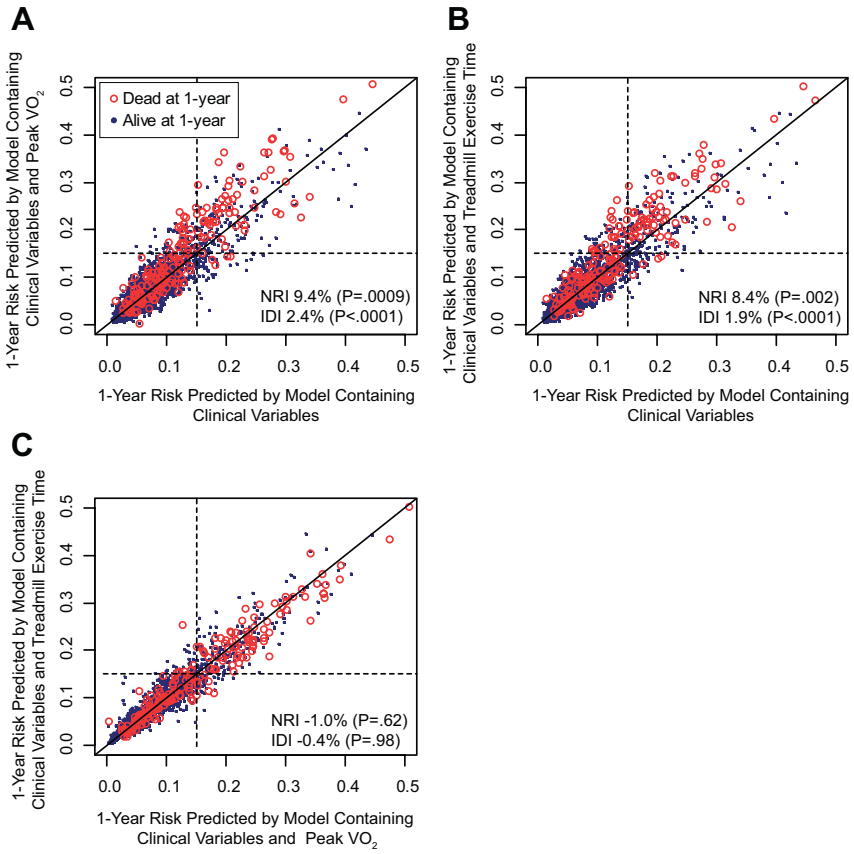


Figure 4. Reclassification: One-year low-risk (<15%) and high-risk (≥15%) survival categories are cutoffs based on the observed national 1-year survival rate after heart transplantation. The end point is death, and improvement of classification is expressed as net reclassification improvement (NRI), which is dependent on the predefined cutoffs, and integrated discrimination improvement (IDI), which is not limited to predetermined risk categories. If a variable adds no predictive value to the model, all points fall on the dark line of identity within each figure. A spread around the line indicates modulation of predicted risk; if the variable correctly modulates predicted risk, there should be a greater preponderance of events (red open circles) above the line of identify. A, Model with clinical variables and peak $\dot{V}O_2$ compared with model with only clinical variables. Peak $\dot{V}O_2$ with cardiac risk factors improved classification in 74 patients (23 who died, 51 who survived) but worsened it in 71 patients (4 who died, 67 who survived), with an IDI of 2.4%. B, Model with clinical variables and treadmill exercise time compared with model with only clinical variables. Treadmill exercise time with cardiac risk factors improved classification in 70 patients (21 who died, 49 who survived) but worsened it in 68 patients (4 who died, 64 who survived), with an IDI of 1.9%. C, Model with clinical variables and treadmill exercise time compared with model with clinical variables and peak $\dot{V}O_2$ improved classification in 43 patients (6 who died, 37 who survived) but worsened it in 46 patients (8 who died, 38 who survived), with an IDI of −0.4%, which was not statistically significant.

c-index) than what we used previously. This approach effectively simulates numerous internal validations, and as expected, model performance would be worse if applied to the entire original data set.

Limitations include that this was a single-center study with only patients who could ambulate on a treadmill. Because the cohort included patients from 1997, there were only a few patients who used aldosterone antagonists, there was modest β-blocker usage, and ≈25% of patients had an ICD. However, these proportions all compare favorably to what is seen in registries that have been published recently.[28–31] Biventricular pacemakers were not reported separately during database entry, but most are identified in the ICD category, because at our institution, biventricular pacemakers are almost always implanted with an ICD. Therefore, when adjusting for all possible confounding variables, we included those with an ICD alone and with an ICD combined with a biventricular pacemaker. Laboratory tests such as measurement of B-type natriuretic peptides were not included in the baseline characteristics because they were not routinely obtained at our center between 1997 and 2007. Although B-type natriuretic peptide has prognostic value, it has not been used in other heart failure survival models.[17,32–35] Like all epidemiological studies, we also cannot account for variables that change with time that may have an impact on death, such as patients receiving an ICD after the cardiopulmonary stress test or changes in medication with time. However, this limitation is inherent in all studies that have confirmed the value of cardiopulmonary stress testing as a prognostic tool. We also did not compare treadmill exercise time to indices of ventilatory inefficiency such as the $\dot{V}E/\dot{V}CO_2$ slope, $PETCO_2$, oscillatory ventilation, and oxygen uptake efficiency slope, which are more powerful risk predictors than peak $\dot{V}O_2$. These indices require minute-by-minute measurements that were not available in the present database.

Despite these limitations, we found that treadmill exercise time is a powerful and independent predictor of outcomes in both women and men with impaired left ventricular systolic

function, even after accounting for peak oxygen consumption. These findings have potentially important implications for our understanding of the clinical pathophysiology of systolic cardiac dysfunction, as well as for finding optimal pathways to screen heart failure patients for possible transplantation. Because cardiopulmonary stress testing is not available in every hospital, treadmill exercise time with a modified Naughton protocol may be of value as an initial prognostic screening tool.

Sources of Funding

This work was supported in part by Health Resources and Services Administration contract 234-2005-370011C. It was also supported in part by American Heart Association Scientist Development Grant 0730307N and National Heart, Lung, and Blood Institute CAN #8324207.

Disclosures

None.

References

1. Mancini DM, Eisen H, Kussmaul W, Mull R, Edmunds LH Jr, Wilson JR. Value of peak exercise oxygen consumption for optimal timing of cardiac transplantation in ambulatory patients with heart failure. *Circulation*. 1991;83:778–786.
2. Hunt SA, Abraham WT, Chin MH, Feldman AM, Francis GS, Ganiats TG, Jessup M, Konstam MA, Mancini DM, Michl K, Oates JA, Rahko PS, Silver MA, Stevenson LW, Yancy CW, Antman EM, Smith SC Jr, Adams CD, Anderson JL, Faxon DP, Fuster V, Halperin JL, Hiratzka LF, Jacobs AK, Nishimura R, Ornato JP, Page RL, Riegel B. ACC/AHA 2005 Guideline update for the diagnosis and management of chronic heart failure in the adult: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Writing Committee to Update the 2001 Guidelines for the Evaluation and Management of Heart Failure): developed in collaboration with the American College of Chest Physicians and the International Society for Heart and Lung Transplantation: endorsed by the Heart Rhythm Society. *Circulation*. 2005;112:e154–e235.
3. Mehra MR, Kobashigawa J, Starling R, Russell S, Uber PA, Parameshwar J, Mohacsi P, Augustine S, Aaronson K, Barr M. Listing criteria for heart transplantation: International Society for Heart and Lung Transplantation guidelines for the care of cardiac transplant candidates–2006. *J Heart Lung Transplant*. 2006;25:1024–1042.
4. Cockcroft DW, Gault MH. Prediction of creatinine clearance from serum creatinine. *Nephron*. 1976;16:31–41.
5. Robbins M, Francis G, Pashkow FJ, Snader CE, Hoercher K, Young JB, Lauer MS. Ventilatory and heart rate responses to exercise: better predictors of heart failure mortality than peak oxygen consumption. *Circulation*. 1999;100:2411–2417.
6. Kleber FX, Waurick P, Winterhalter M. CPET in heart failure. *Eur Heart J Suppl*. 2004;6:D1–D4.
7. 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;284:1392–1398.
8. Lauer MS, Martino D, Ishwaran H, Blackstone EH. Quantitative measures of electrocardiographic left ventricular mass, conduction, and repolarization, and long-term survival after coronary artery bypass grafting. *Circulation*. 2007;116:888–893.
9. O'Neill JO, Young JB, Pothier CE, Lauer MS. Peak oxygen consumption as a predictor of death in patients with heart failure receiving beta-blockers. *Circulation*. 2005;111:2313–2318.
10. UNOS Heart Kaplan–Meier graft survival rates for transplants performed 1997–2004. June 27, 2008. Available at: http://optn.transplant.hrsa.gov/latestData/step2.asp? Accessed July 3, 2008.
11. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med*. 2008;27:157–172.

11a. 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;147:821–828.

12. Myers J, Prakash M, Froelicher V, Do D, Partington S, Atwood JE. Exercise capacity and mortality among men referred for exercise testing. *N Engl J Med*. 2002;346:793–801.
13. Gulati M, Pandey DK, Arnsdorf MF, Lauderdale DS, Thisted RA, Wicklund RH, Al-Hani AJ, Black HR. Exercise capacity and the risk of death in women: the St James Women Take Heart Project. *Circulation*. 2003;108:1554–1559.
14. Mora S, Redberg RF, Cui Y, Whiteman MK, Flaws JA, Sharrett AR, Blumenthal RS. 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;290:1600–1607.
15. Blair SN, Kampert JB, Kohl HW III, Barlow CE, Macera CA, Paffenbarger RS Jr, Gibbons LW. Influences of cardiorespiratory fitness and other precursors on cardiovascular disease and all-cause mortality in men and women. *JAMA*. 1996;276:205–210.
16. Adabag AS, Grandits GA, Prineas RJ, Crow RS, Bloomfield HE, Neaton JD. Relation of heart rate parameters during exercise test to sudden death and all-cause mortality in asymptomatic men. *Am J Cardiol*. 2008;101:1437–1443.
17. Levy WC, Mozaffarian D, Linker DT, Sutradhar SC, Anker SD, Cropp AB, Anand I, Maggioni A, Burton P, Sullivan MD, Pitt B, Poole-Wilson PA, Mann DL, Packer M. The Seattle Heart Failure Model: prediction of survival in heart failure. *Circulation*. 2006;113:1424–1433.
18. Fonarow GC, Adams KF Jr, Abraham WT, Yancy CW, Boscardin WJ. Risk stratification for in-hospital mortality in acutely decompensated heart failure: classification and regression tree analysis. *JAMA*. 2005;293:572–580.
19. Aaronson KD, Schwartz JS, Chen TM, Wong KL, Goin JE, Mancini DM. Development and prospective validation of a clinical index to predict survival in ambulatory patients referred for cardiac transplant evaluation. *Circulation*. 1997;95:2660–2667.
20. Gradman A, Deedwania P, Cody R, Massie B, Packer M, Pitt B, Goldstein S; Captopril-Digoxin Study Group. Predictors of total mortality and sudden death in mild to moderate heart failure. *J Am Coll Cardiol*. 1989;14:564–570.
21. Clark AL, Coats AJ. Exercise endpoints in patients with chronic heart failure. *Int J Cardiol*. 2000;73:61–66.
22. Guimaraes GV, d'Avila VM, Camargo PR, Moreira LF, Luces JR, Bocchi EA. Prognostic value of cardiopulmonary exercise testing in children with heart failure secondary to idiopathic dilated cardiomyopathy in a non-beta-blocker therapy setting. *Eur J Heart Fail*. 2008;10:560–565.
23. Myers J, Gullestad L, Vagelos R, Do D, Bellin D, Ross H, Fowler MB. Clinical, hemodynamic, and cardiopulmonary exercise test determinants of survival in patients referred for evaluation of heart failure. *Ann Intern Med*. 1998;129:286–293.
24. Myers J, Gullestad L, Vagelos R, Do D, Bellin D, Ross H, Fowler MB. Cardiopulmonary exercise testing and prognosis in severe heart failure: 14 mL/kg/min revisited. *Am Heart J*. 2000;139:78–84.
25. Bensimhon DR, Leifer ES, Ellis SJ, Fleg JL, Keteyian SJ, Pina IL, Kitzman DW, McKelvie RS, Kraus WE, Forman DE, Kao AJ, Whellan DJ, O'Connor CM, Russell SD. Reproducibility of peak oxygen uptake and other cardiopulmonary exercise testing parameters in patients with heart failure (from the Heart Failure and A Controlled Trial Investigating Outcomes of exercise traiNing). *Am J Cardiol*. 2008;102:712–717.
26. Hsich E, Chadalavada S, Krishnaswamy G, Starling RC, Pothier CE, Blackstone EH, Lauer MS. Long-term prognostic value of peak oxygen consumption in women versus men with heart failure and severely impaired left ventricular systolic function. *Am J Cardiol*. 2007;100:291–295.
27. Elmariah S, Goldberg LR, Allen MT, Kao A. Effects of gender on peak oxygen consumption and the timing of cardiac transplantation. *J Am Coll Cardiol*. 2006;47:2237–2242.
28. Yancy CW, Lopatin M, Stevenson LW, De Marco T, Fonarow GC. Clinical presentation, management, and in-hospital outcomes of patients admitted with acute decompensated heart failure with preserved systolic function: a report from the Acute Decompensated Heart Failure National Registry (ADHERE) Database. *J Am Coll Cardiol*. 2006;47:76–84.
29. Gheorghiade M, Abraham WT, Albert NM, Greenberg BH, O'Connor CM, She L, Stough WG, Yancy CW, Young JB, Fonarow GC. Systolic blood pressure at admission, clinical characteristics, and outcomes in patients hospitalized with acute heart failure. *JAMA*. 2006;296:2217–2226.

30. Fonarow GC, Stough WG, Abraham WT, Albert NM, Gheorghiade M, Greenberg BH, O'Connor CM, Sun JL, Yancy CW, Young JB. Characteristics, treatments, and outcomes of patients with preserved systolic function hospitalized for heart failure: a report from the OPTIMIZE-HF Registry. *J Am Coll Cardiol*. 2007;50:768–777.
31. Ezekowitz JA, Lee DS, Tu JV, Newman AM, McAlister FA. Comparison of one-year outcome (death and rehospitalization) in hospitalized heart failure patients with left ventricular ejection fraction >50% versus those with ejection fraction <50%. *Am J Cardiol*. 2008;102:79–83.
32. Brophy JM, Dagenais GR, McSherry F, Williford W, Yusuf S. A multivariate model for predicting mortality in patients with heart failure and systolic dysfunction. *Am J Med*. 2004;116:300–304.
33. Zugck C, Kruger C, Kell R, Korber S, Schellberg D, Kubler W, Haass M. Risk stratification in middle-aged patients with congestive heart failure: prospective comparison of the Heart Failure Survival Score (HFSS) and a simplified two-variable model. *Eur J Heart Fail*. 2001;3:577–585.
34. Stempfle HU, Alt A, Stief J, Siebert U. The Munich score: a clinical index to predict survival in ambulatory patients with chronic heart failure in the era of new medical therapies. *J Heart Lung Transplant*. 2008;27:222–228.
35. Lund LH, Aaronson KD, Mancini DM. Validation of peak exercise oxygen consumption and the Heart Failure Survival Score for serial risk stratification in advanced heart failure. *Am J Cardiol*. 2005;95:734–741.

CLINICAL PERSPECTIVE

Physicians often refer patients with advanced heart failure for exercise testing with metabolic gas exchange measurements to assess suitability for cardiac transplantation. Peak oxygen consumption and other metabolic measures are known to be powerful predictors of mortality. We asked whether treadmill exercise time according to a standardized protocol might be comparable to peak oxygen consumption for assessment of prognosis. We analyzed the outcomes of 2231 patients with systolic heart failure who all underwent metabolic exercise stress testing on a modified Naughton protocol. During a mean follow-up of 5 years, 742 patients (33%) died, and 249 patients (11%) underwent heart transplantation urgently. We found that after accounting for baseline clinical characteristics, treadmill exercise time performed similarly to peak oxygen consumption for predicting poor outcome. For a 1-minute change in exercise time, there was a 7% increased hazard of death (eg, comparing 480 to 540 seconds, hazard ratio =1.07, 95% confidence interval 1.02 to 1.12, $P=0.004$). Mortality rates were particularly high for women who exercised less than 5 minutes and 17 seconds and for men who exercised less than 6 minutes. Our findings suggest that treadmill exercise time may be valuable as an initial prognostic screening tool in patients with advanced heart failure.

Appendix Figure Legends

Appendix Figure 1. Kaplan-meier plots for (A) women and (B) men, stratified by quartiles of treadmill exercise time. (A) Women (N=602) Quartile 1: 21-315 seconds, Quartile 2: 317-478 seconds, Quartile 3: 480-600 seconds, Quartile 4: 603-1320 seconds. (B) Men (N=1629) Quartile 1: 35-357 seconds, Quartile 2: 360-480 seconds, Quartile 3: 484-657 seconds, Quartile 4: 660-1415 seconds.

Appendix Figure 2. Kaplan-meier plots for (A) women and (B) men with peak $VO_2 \geq 14$, stratified by tertiles of increasing treadmill exercise time. (A) Women: Tertile 1: < 518 seconds, Tertile 2: 518- 639 seconds , Tertile 3: $\geq$ 640 seconds. (B) Men: Tertile 1: < 511 seconds, Tertile 2: 511-682 seconds, Tertile 3: $\geq$ 683 seconds. Peak $VO_{2=}$ peak oxygen consumption

Appendix Figure 3. Change in out-of-bagging determined prediction error. Only the 10 most important variables are shown. Results are based on 100 bootstrapped samples

Appendix Figure 4. Reclassification: One year low-risk (<15%) and high-risk (≥15%) survival categories are cut-offs based on the observed national 1-year survival after heart transplantation. The improvement of classification to predict death or need for UNOS status 1 heart transplantion is expressed as net reclassification improvement (NRI) which is dependent on the pre-defined cut-offs and integrated discrimination improvement (IDI) which is not limited to pre-determined risk categories. If a variable adds no predictive value to the model, all points fall on the dark line of identity within each figure. Spread around the line indicates

modulation of predicted risk; if the variable correctly modulates predicted risk, there should be a greater preponderance of events (red open circles) above the line of identify. A. Model with clinical variables and peak VO_2 compared to model with only clinical variables. Peak VO_2 with cardiac risk factors improved classification in 150 patients (14 who died, 136 who survived) but worsened it in 120 patients (10 who died, 110 who survived) with an IDI of 4.7% B. Model with clinical variables and treadmill exercise time compared to model with only clinical variables. Treadmill exercise time with cardiac risk factors improved classification in 143 patients (15 who died, 128 who survived) but worsened it in 120 patients (8 who died, 112 who survived) with an IDI of 3.8% C. Model with clinical variables and treadmill exercise time compared to model with clinical variables and peak VO_2 improved classification in 85patients (11 who died, 74 who survived) but worsened it in 72 patients (9who died, 63 who survived) with an IDI of -0.8% which was not statistically significant.

Appendix Table 1. Treadmill Exercise Time and Outcome: Cox Proportional Hazards Analyses

Model: *Death or UNOS 1 Transplant*	**Hazard Ratio (95% CI)**	**P**
Treadmill exercise time as a continuous variable*		
Unadjusted	2.26 (2.05 to 2.49)	<.0001
Adjusted for age and sex	2.33 (2.11 to 2.58)	<.0001
Adjusted for age, sex, history of CAD, and peak VO_2	1.37 (1.13 to 1.67)	.002
Multivariable adjusted	1.34 (1.10 to 1.63)	.004
Treadmill exercise time as a dichotomous variable^		
Unadjusted	5.20 (4.20 to 6.43)	<.0001
Multivariable adjusted	1.75 (1.15 to 2.66)	.009

* Comparisons are between the 25th percentile (317 seconds in women, 360 seconds in men) and 75th percentile (600 seconds in women, 657 seconds in men)

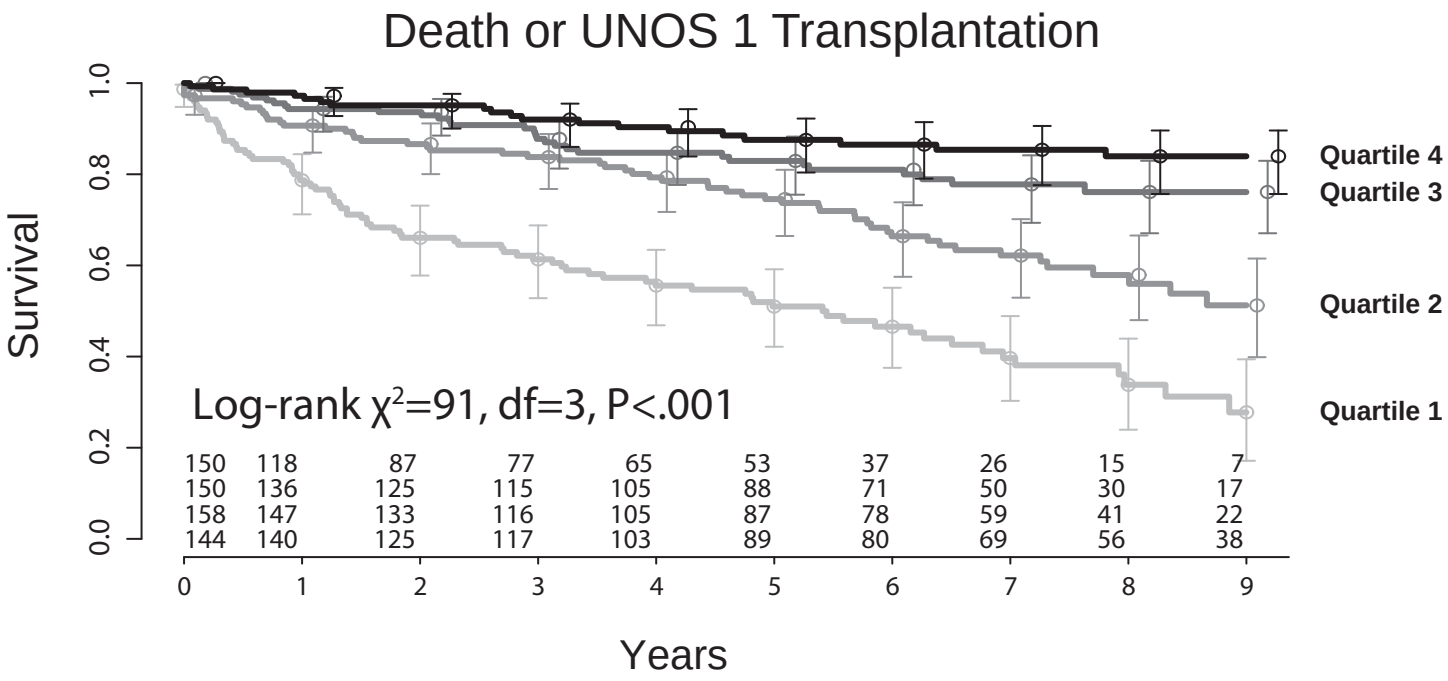
^ Comparisons between quartile 1 (<316 seconds in women, <358 seconds in men) and quartile 4 (>602 seconds in women, >659 seconds in men)

Appendix Table 2. Out of Bag (OOB) Concordance Index Values For Cox Regression Models

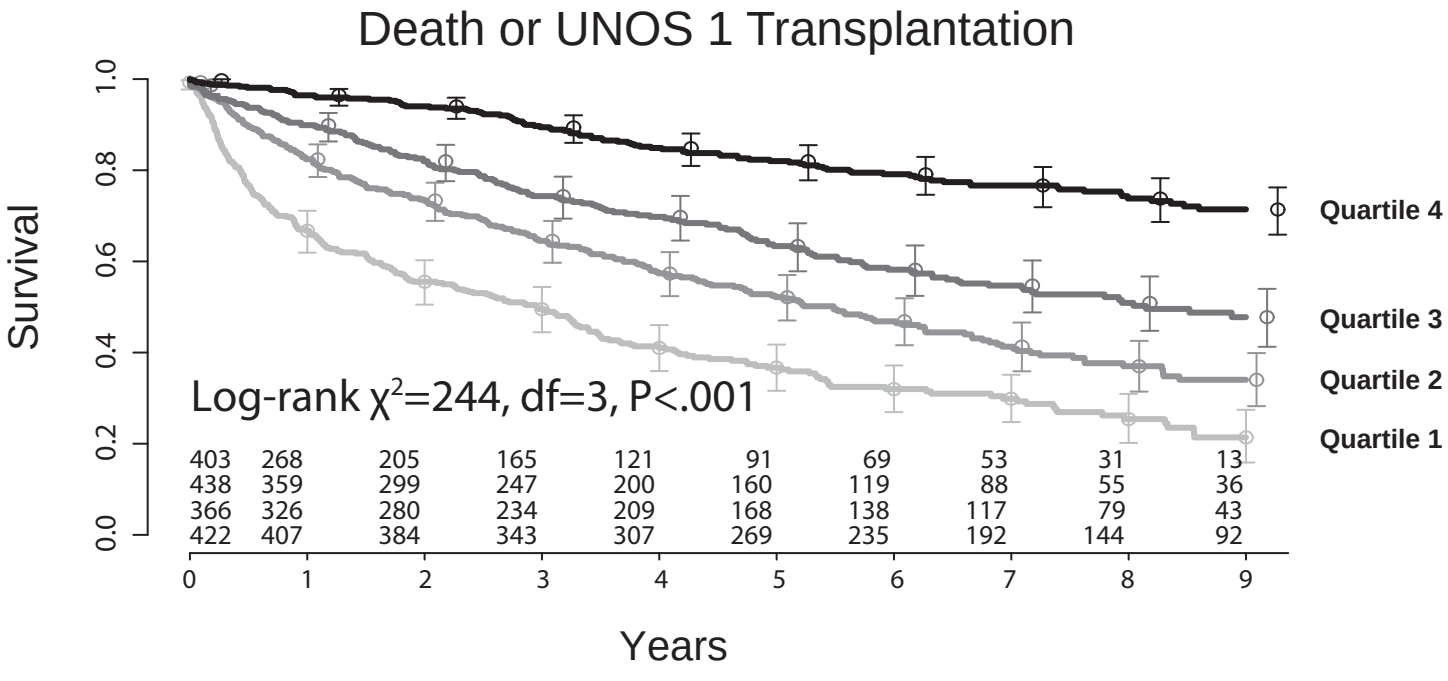
	Death or UNOS 1 transplantation	
	Concordance Index (95% Confidence Intervals)	
Clinical variables only	0.699	(0.679 to 0.719)
Clinical variables + Peak VO_2	0.726	(0.702 to 0.750)
Clinical variables + Time	0.722	(0.698 to 0.746)
Clinical variables + Peak VO_2 + Time	0.727	(0.703 to 0.751)

Appendix Figure 1.

A. **Women**

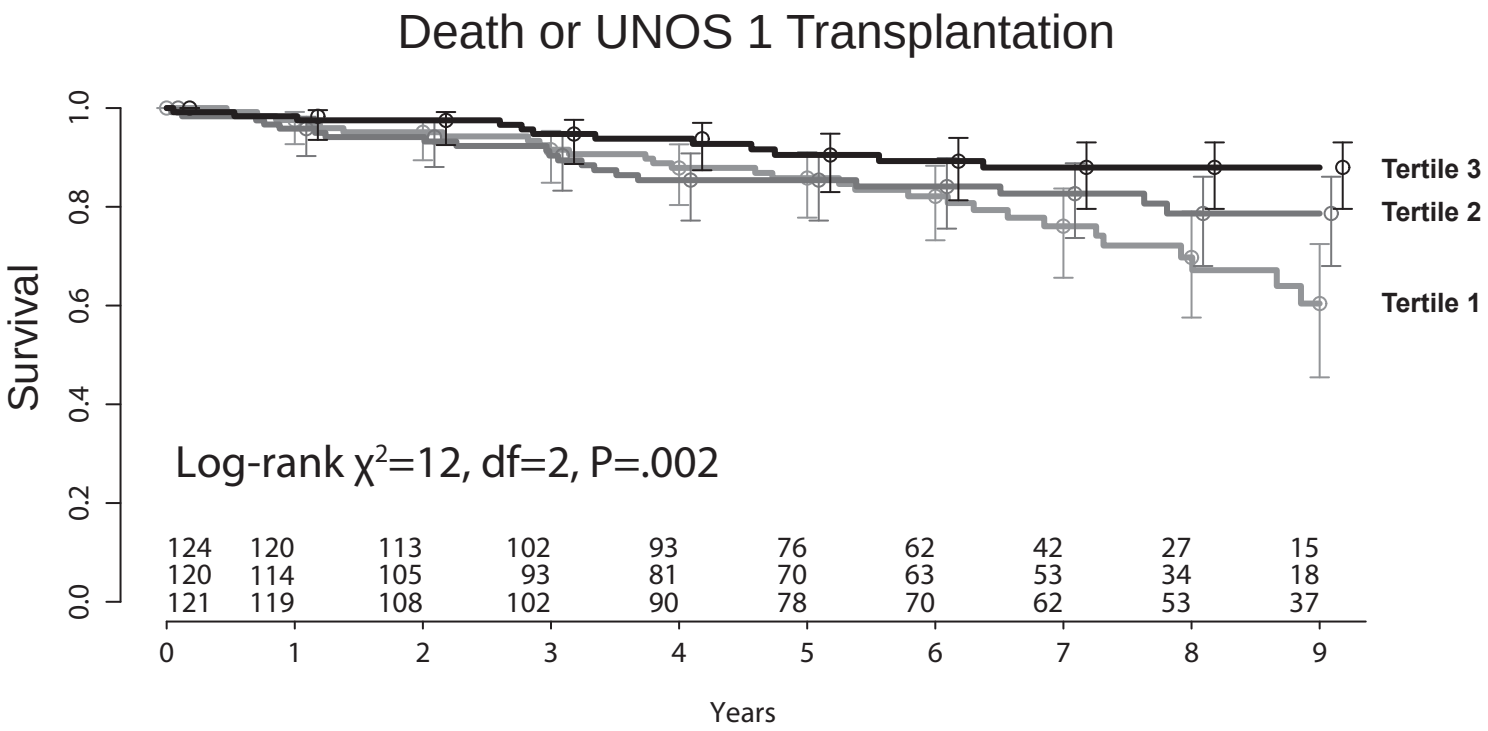


B. **Men**

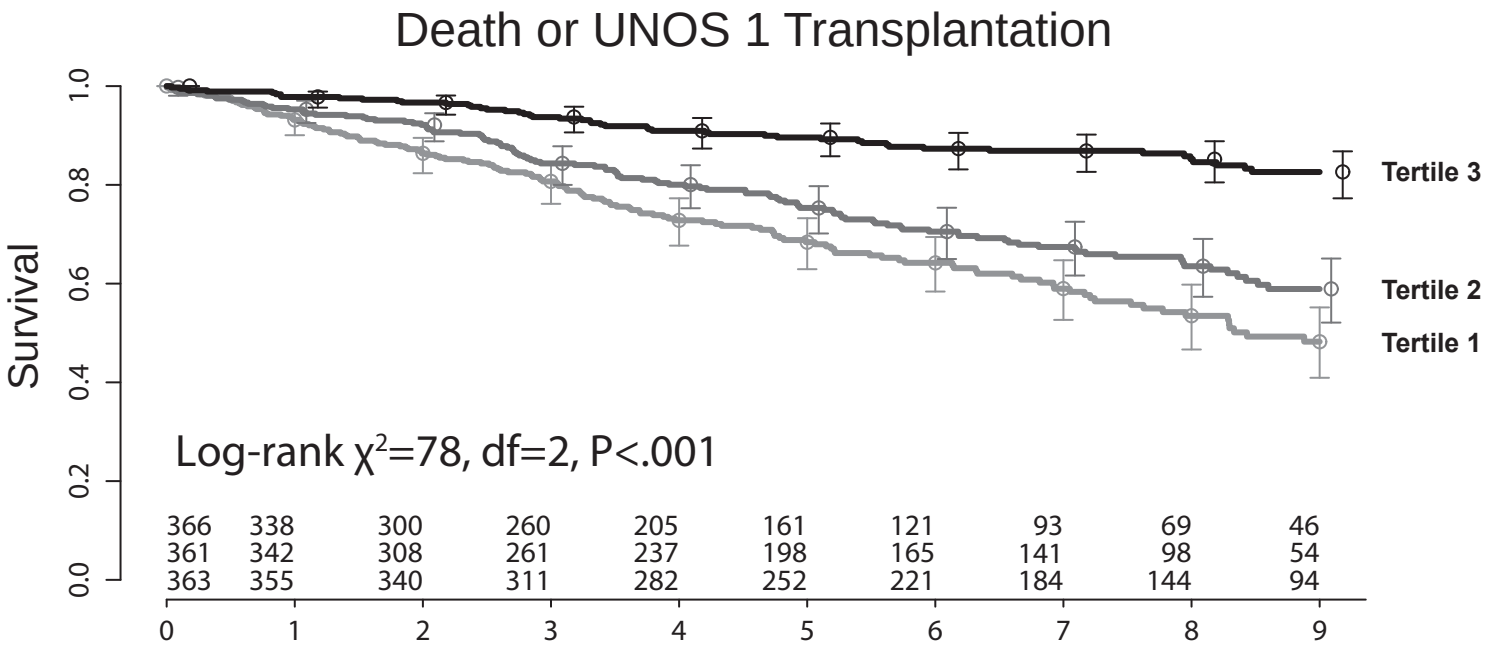


Appendix Figure 2.

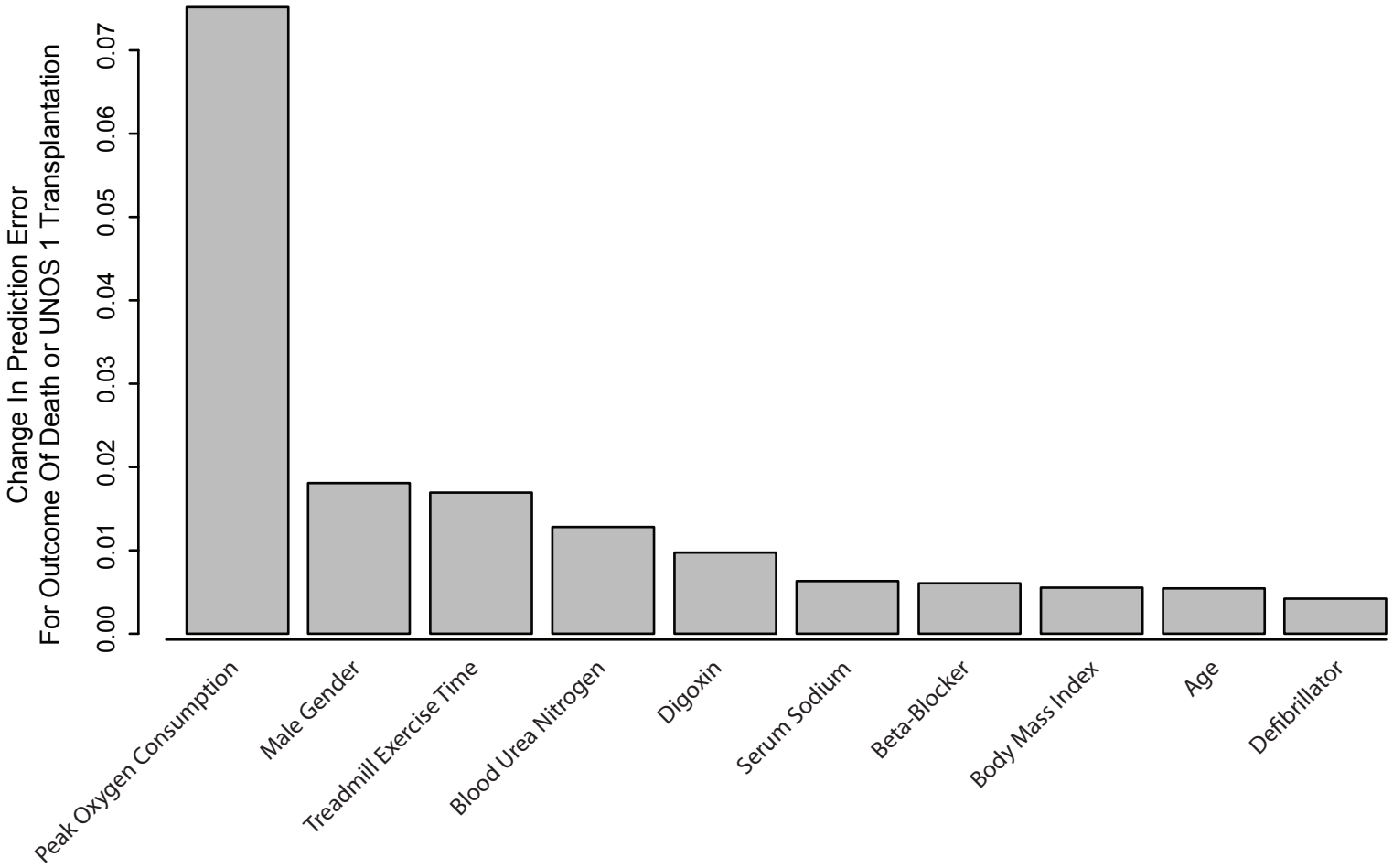
A. Women, Peak $VO_2 \geq 14$



B. Men, Peak $VO_2 \geq 14$

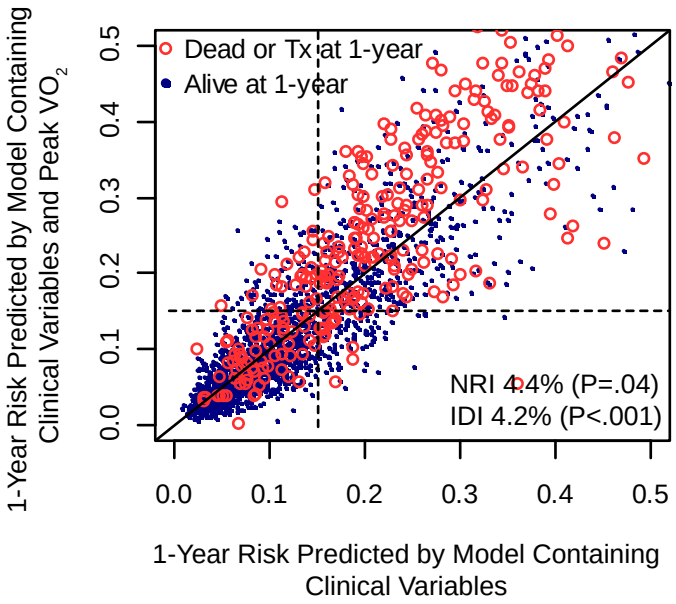


Appendix Figure 3.

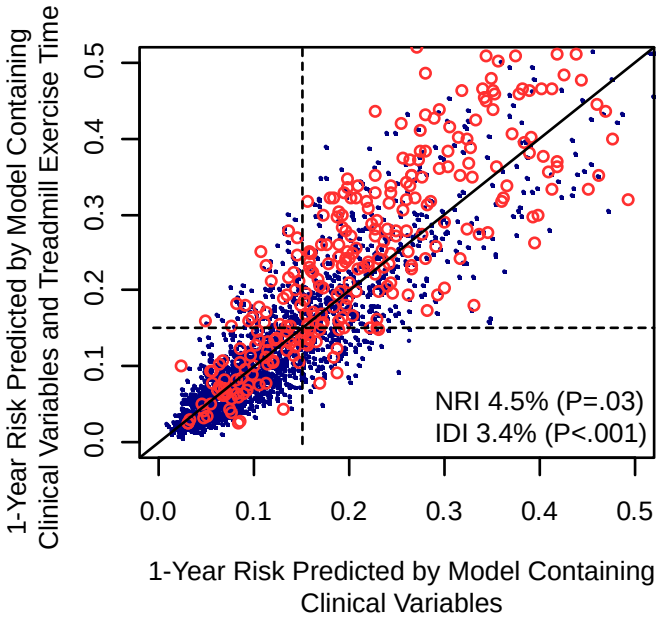


Appendix Figure 4.

A.



B.



C.