HOSTED BY

Available online at www.sciencedirect.com

ScienceDirect

Journal of Sport and Health Science 13 (2024) 611–620

Original article

Development, validation, and transportability of several machine-learned, non-exercise-based VO_{2max} prediction models for older adults

Benjamin T. Schumacher [a,*], Michael J. LaMonte [b], Andrea Z. LaCroix [a], Eleanor M. Simonsick [c], Steven P. Hooker [d], Humberto Parada Jr. [e,f], John Bellettiere [a], Arun Kumar [g]

[a] *Herbert Wertheim School of Public Health and Human Longevity Science, University of California San Diego, La Jolla, CA 92093, USA*
[b] *Department of Epidemiology and Environmental Health, School of Public Health and Health Professions, University at Buffalo–State University of New York, Buffalo, NY 14214, USA*
[c] *Translational Gerontology Branch, Intramural Research Program, National Institute on Aging, National Institutes of Health, Baltimore, MD 21225, USA*
[d] *College of Health and Human Services, San Diego State University, San Diego, CA 92182, USA*
[e] *Division of Epidemiology and Biostatistics, School of Public Health, San Diego State University, San Diego, CA 92182, USA*
[f] *University of California San Diego Moores Cancer Center, La Jolla, CA 92093, USA*
[g] *Computer Science and Engineering and Halicioglu Data Science Institute, University of California San Diego, La Jolla, CA 92093, USA*

Received 1 September 2023; revised 30 November 2023; accepted 18 January 2024
Available online 29 February 2024

Abstract

Background: There exist few maximal oxygen uptake (VO_{2max}) non-exercise-based prediction equations, fewer using machine learning (ML), and none specifically for older adults. Since direct measurement of VO_{2max} is infeasible in large epidemiologic cohort studies, we sought to develop, validate, compare, and assess the transportability of several ML VO_{2max} prediction algorithms.

Methods: The Baltimore Longitudinal Study of Aging (BLSA) participants with valid VO_{2max} tests were included ($n = 1080$). Least absolute shrinkage and selection operator, linear- and tree-boosted extreme gradient boosting, random forest, and support vector machine (SVM) algorithms were trained to predict VO_{2max} values. We developed these algorithms for: (a) the overall BLSA, (b) by sex, (c) using all BLSA variables, and (d) variables common in aging cohorts. Finally, we quantified the associations between measured and predicted VO_{2max} and mortality.

Results: The age was 69.0 ± 10.4 years (mean ± SD) and the measured VO_{2max} was 21.6 ± 5.9 mL/kg/min. Least absolute shrinkage and selection operator, linear- and tree-boosted extreme gradient boosting, random forest, and support vector machine yielded root mean squared errors of 3.4 mL/kg/min, 3.6 mL/kg/min, 3.4 mL/kg/min, 3.6 mL/kg/min, and 3.5 mL/kg/min, respectively. Incremental quartiles of measured VO_{2max} showed an inverse gradient in mortality risk. Predicted VO_{2max} variables yielded similar effect estimates but were not robust to adjustment.

Conclusion: Measured VO_{2max} is a strong predictor of mortality. Using ML can improve the accuracy of prediction as compared to simpler approaches but estimates of association with mortality remain sensitive to adjustment. Future studies should seek to reproduce these results so that VO_{2max}, an important vital sign, can be more broadly studied as a modifiable target for promoting functional resiliency and healthy aging.

Keywords: Cardiorespiratory fitness; Prediction algorithms; Epidemiology; Mortality

1. Introduction

Cardiorespiratory fitness (CRF) refers to the circulatory and respiratory systems' capacity to provide oxygen to skeletal muscles for engaging in physical activity.[1] While factors such as age, sex, health status, and genetics are strong determinants of CRF, one's level of habitual physical activity is the principal modifiable determinant of this attribute.[1] Scientific evidence from clinical, epidemiologic, and exercise science studies has consistently shown higher CRF to have strong, independent, and beneficial associations with several clinical outcomes. Higher CRF predicts lower incidence and mortality from coronary heart disease/cardiovascular disease,[2–4] longer survival times,[3,5–7] and lower rates of loss of independence for older adults.[8]

Peer review under responsibility of Shanghai University of Sport.
* Corresponding author.
E-mail address: BTS70@pitt.edu (B.T. Schumacher).

https://doi.org/10.1016/j.jshs.2024.02.004
Cite this article: Schumacher BT, LaMonte MJ, LaCroix AZ, et al. Development, validation, and transportability of several machine-learned, non-exercise-based VO_{2max} prediction models for older adults. *J Sport Health Sci* 2024;13:611–20.

Maximal oxygen uptake (VO_{2max}) is the gold standard measure of CRF and is recognized as a hallmark biomarker of healthy aging.[1,9] VO_{2max} measurements in research settings involve maximal graded exercise tests, usually conducted on a treadmill or stationary cycle ergometer. Such assessments typically require highly trained personnel, specialized testing equipment and, in most instances, direct physician supervision to reduce the risk of adverse events. Because VO_{2max} testing involves strenuous activity to the point of absolute exhaustion, it is often contraindicated for older adults. These features make direct measurement of VO_{2max} infeasible in large epidemiologic cohort studies. To provide alternatives, researchers have published non-exercise VO_{2max} prediction equations that can be used to approximate laboratory-measured VO_{2max} in large epidemiologic cohorts. However, few of these equations were designed for use specifically in older adults.[10,11] A recent systematic review of the published VO_{2max} prediction equations utilizing machine learning (ML) algorithms determined few equations could be applied to epidemiologic cohorts that do not have exercise testing data, and none of these ML models were developed in older adult populations.[12] By the year 2060, nearly one-fourth of the U.S. population will be ≥65 years of age. Given the associations of higher CRF with beneficial health outcomes, the ability to precisely estimate VO_{2max} in older adults is a growing and critical need as we continue to investigate the effects of CRF on healthy aging.[13]

Thus, we aimed to develop, validate, and compare multiple machine-learned, non-exercise based VO_{2max} prediction algorithms for older adults using laboratory-measured VO_{2max} in the Baltimore Longitudinal Study of Aging (BLSA). We aimed to develop these algorithms for the BLSA sample overall and by sex, to assess the association of measured and predicted VO_{2max} with all-cause mortality, and to assess the feasibility of transporting these algorithms to an external epidemiologic cohort.

2. Methods

2.1. Study participants

The analytic sample was drawn from the BLSA, which is conducted by the National Institute on Aging Intramural Research Program.[14] Established in 1958, the BLSA is the longest on-going scientific study of aging.[15,16] Participants visit the BLSA testing facility every 1−4 years for health, cognitive, and functional evaluations lasting 3 days. Since its inception, over 3500 individuals have participated in the BLSA, and more than 1300 remain active.[15] Extensive details about BLSA design, recruitment, and measurements are available elsewhere.[16] All participants provided written informed consent, which was approved by the applicable Institutional Review Boards (IRB protocol number: 03-AG-0325).

2.2. Measures

2.2.1. VO_{2max} measurement

Using a modified Balke treadmill testing protocol,[17,18] VO_{2max} was measured as milliliters of oxygen uptake per kilogram (kg) of body weight per minute (mL/kg/min). The participants walked on a treadmill at a set pace (3.0 miles per hour for women and 3.5 miles per hour for men) and the incline of the treadmill increased by 3% every 2 min until the participant indicated having reached exhaustion. Standard safety measures were observed and are detailed in the Technical Appendix. During this test, a gas meter (Parkinson-Cowan, Waitsfield, VA, USA) was used to measure expired gas volumes. A medical mass spectrometer (Perkin-Elmer MGA-1110; Milwaukee, WI, USA; calibrated daily using standard gases) was used to measure oxygen and carbon dioxide concentrations. Every 30 s during the test, average expired gas concentrations were calculated by a programmed interface between the gas meter and mass spectrometer, and VO_{2max} was defined as the highest 30-s oxygen uptake value.

Maximal effort on the treadmill test was specified as a respiratory exchange ratio of greater than 1.0. Of 52 participants with a respiratory exchange ratio value just below the cutoff when the treadmill was stopped, 11 achieved 85% or more of their age-predicted maximal heart rate in beats per minute (computed as 220 − age in years) and had a value greater than 17 on the 20-point Borg rating of perceived exertion (RPE) scale. These test results were considered to reflect their maximal effort and were included in the present analysis. Of the remaining 41 participants with a respiratory exchange ratio of less than 1.0 when the treadmill was stopped, 31 had no other VO_{2max} test meeting the aforementioned criteria and were excluded from the present analysis, and 10 provided a subsequent VO_{2max} test that satisfied these maximal test criteria and so were included in the analysis, resulting in a final analytic sample of 1080 participants. For participants having more than one VO_{2max} measurement, only the first measurement meeting the maximal effort criteria was included.

2.2.2. Outcome ascertainment

Participant information was linked to the National Death Index[19] to ascertain vital status and, for those deceased, their date of death. Follow-up occurred from the participant's VO_{2max} test date (VO_{2max} measurements ranged from January 1, 2007, to January 21, 2020) until April 15, 2021. Vital status classification was obtained for 96% of participants. There were 141 participant deaths from any cause during a median follow-up of 9.6 years (range: 0.6−14.1 years).

2.2.3. Covariates

2.2.3.1. Demographics and physical attributes. Demographic variables included self-identified sex (male or female), race and ethnicity (non-Hispanic White, non-Hispanic Black, Hispanic, non-Hispanic Asian/other Pacific Islander, or non-Hispanic other/not classifiable), education (non-college graduate, college graduate, or post-college), age, height (cm) measured using a stadiometer, weight (kg) measured using a calibrated scale, body mass index (BMI) calculated as weight (kg) divided by height (m) squared, and waist circumference (cm) using a tape measure.

2.2.3.2. Health status/health history/functional capacity. Health status variables included the 12-item Short Form Survey of self-rated health scale[20] and its physical and mental health composite scores, hand grip muscle strength scores (kg) in both hands using a Jamar Hydraulic Hand Dynamometer (Lafayette Instrument Company, Lafayette, IN, USA),[21] and Short Physical Performance Battery (SPPB) physical function score (0−12, higher is better) and its 3 components.[22,23] Additional timed walk tests included the number of meters walked at usual pace for 2.5 min,[24] the number of seconds to walk 400 m at a fast pace,[24] and a walking capacity summary score. Details about the derivation of the walking capacity summary score have been published elsewhere.[25] Health history variables included dichotomous indicators (yes/no) for a physician diagnosis of myocardial infarction, congestive heart failure, stroke, diabetes, glucose intolerance, high blood sugar, and breast cancer. Additionally, measurements were taken of seated, resting systolic and diastolic blood pressure from both arms, resting heart rate, and heart rate at the end of the 2.5-min usual pace walk.

2.2.3.3. Health behaviors. BLSA participants reported their time spent performing 97 activities over the last 2 years, and each activity was assigned a metabolic equivalent value[26] to estimate calories expended in all activity, calories expended in all activities per kg of body weight, calories expended in exercise-related activity, minutes of any exercise per week (0−29 min, >29−74 min, >74−149 min, or >149 min), continuous minutes of: any walking per week, brisk walking per week, vigorous activity per week. Self-reported health behavior variables were included: smoking history (never, current, or former smoker), beta blocker use (yes/no), and blood sugar medication use (yes/no).

2.2.3.4. Objective Physical Activity and Cardiovascular Health Study (OPACH) covariates. We sought to assess the feasibility of transporting the machine-learned algorithms from BLSA to an external epidemiologic aging cohort containing variables common in population studies on aging. To accomplish this, the algorithms were re-trained using only predictors that exist in both BLSA and the OPACH, an ancillary study of the Women's Health Initiative. Extensive details about OPACH have been published elsewhere.[27] For the purposes of the present study, the OPACH dataset contained all the BLSA covariates except for measures of rapid gait speed, 2.5-min usual pace walk, 400-m fast walk, walking capacity summary score, and heart rate measures during and after the 2.5-min walk.

2.3. ML algorithms

Using measured VO_{2max} as the ground truth, we trained 5 ML algorithms to predict VO_{2max}: least absolute shrinkage and selection operator (LASSO), extreme gradient boosting (XGBoost) with a linear booster, XGBoost with a tree booster, random forest, and Support Vector Regression, a specific application of support vector machine (SVM). Details about these algorithms' processes, hyperparameter specifications, and packages can be found in the Supplementary Technical Appendix. When the LASSO algorithm was applied to OPACH, approximately 38% of OPACH participants' predicted VO_{2max} values were missing due to missing covariate data, so LASSO was not used in the regression modeling.

These ML algorithms were trained using all BLSA participants combined and separately for BLSA men and women. The total sample and sex-stratified algorithms were trained using all the aforementioned variables within the BLSA and, to assess whether the results are transportable to an external cohort, using only the variables common between BLSA and OPACH.

2.4. Statistical analysis

Analysis of variance tests for continuous variables and χ^2 tests for categorical variables were used to compare covariates by sex-specific quartiles of measured VO_{2max}.

Next, Cox proportional hazards regression models were used to estimate the associations between quartiles of VO_{2max} (measured and predicted VO_{2max}; independent variables) and all-cause mortality (dependent variable). Model 1 was unadjusted, and Model 2 was adjusted for age, sex, race and ethnicity, and education. To test the linear trends across quartiles and obtain a p value for trend (p_{trend}), we specified the indicator for quartile in the model as a continuous variable. Using the same modeling approach, we also assessed VO_{2max} as a continuous variable estimating adjusted hazard ratios (HRs) for all-cause mortality associated with a 1-standard deviation (SD) increase in VO_{2max}. The p values for mean-centered, SD-scaled VO_{2max} variable for Models 1 and 2 are presented. The concordance statistic (C-statistic), a measure of discrimination for time-to-event models that gives the proportion of participant pairs for which the model correctly predicts the participant in the pair who experiences a mortality event first, is also presented.[28]

All analyses were conducted in R version 3.6.3 (R Foundation for Statistical Computing, Vienna, Austria). All R codes and trained algorithms are available at: https://github.com/benschumacher12/VO2maxPredicitionAlgos.

3. Results

3.1. Sample characteristics

For the 565 women and 515 men with laboratory measures of VO_{2max}, age was 69.0 ± 10.4 years (mean ± SD), BMI was 27.0 ± 4.4 kg/m^2, and the measured VO_{2max} was 21.6 ± 5.9 mL/kg/min (median ± SD) (Table 1). The median VO_{2max} for the men was 23.7 ± 6.1 mL/kg/min (range: 9.5−48.9 mL/kg/min) and the median VO_{2max} for the women was 19.9 ± 5.1 mL/kg/min (range: 6.2−42.1 mL/kg/min). Two-thirds of the participants were non-Hispanic White, 25.8% non-Hispanic Black, 4.6% Asian, 3.2% Hispanic, while the remaining 0.7% belonged to other race/ethnicity categories. The majority of participants (61.9%) had a post-college education. The prevalence of current smoking was 1.8%. Mean systolic and diastolic

blood pressure was 114.1 ± 14.1 mmHg and 66.7 ± 8.8 mmHg, respectively (Table 1).

3.2. Performance of machine-learned VO_{2max} prediction algorithms

The first algorithm, LASSO, yielded a root mean squared error (RMSE; lower values indicated better prediction) of 3.4 mL/kg/min for VO_{2max} prediction in the total sample using all predictors (Table 2). For the subgroups (sex-stratified in combination with the BLSA-predictor and OPACH-predictor algorithms), predicted VO_{2max} RMSEs ranged from 2.8 to 3.8 mL/kg/min for the women's BLSA-predictor and men's OPACH-predictor, respectively. The linear XGBoost yielded an RMSE of 3.6 mL/kg/min for VO_{2max} prediction in the total sample using all predictors and OPACH predictors. For the

Table 1
Characteristics of BLSA participants overall and according to quartiles of measured VO_{2max} ($n = 1080$).

Characteristics	Total ($n = 1080$)	Measured VO_{2max}				p[b]
		Quartile 1[a] ($n = 270$)	Quartile 2[a] ($n = 277$)	Quartile 3[a] ($n = 265$)	Quartile 4[a] ($n = 268$)	
Death	141 (13.1)	81 (30.0)	38 (13.7)	14 (5.3)	8 (3.0)	**<0.01**
Age (year)	69.0 ± 10.4	75.5 ± 8.8	72.1 ± 9.7	67.3 ± 8.9	60.9 ± 8.2	**<0.01**
Race and ethnicity						**<0.01**
Non-Hispanic, White	708 (65.6)	169 (62.6)	177 (63.9)	177 (66.8)	185 (69.0)	
Non-Hispanic, Black	279 (25.8)	87 (32.2)	82 (29.6)	60 (22.6)	50 (18.7)	
Non-Hispanic, Asian/other Pacific Islander	50 (4.6)	8 (3.0)	9 (3.2)	14 (5.3)	19 (7.1)	
Hispanic	35 (3.2)	4 (1.5)	6 (2.2)	11 (4.2)	14 (5.2)	
Non-Hispanic, other/not classifiable	8 (0.7)	2 (0.7)	3 (1.1)	3 (1.1)	0 (0.0)	
Highest attained education						**<0.01**
Post college	669 (61.9)	152 (56.3)	168 (60.6)	169 (63.8)	180 (67.2)	
College	225 (20.8)	51 (18.9)	53 (19.1)	57 (21.5)	64 (23.9)	
Non-college graduate	183 (16.9)	67 (24.8)	56 (20.2)	39 (14.7)	21 (7.8)	
Missing	3 (0.3)	0 (0.0)	0 (0.0)	0 (0.0)	3 (1.1)	
BMI (kg/m^2)	27.0 ± 4.4	28.9 ± 4.7	27.4 ± 4.6	26.6 ± 4.1	24.9 ± 3.4	**<0.01**
Beta blocker use	152 (14.1)	78 (28.9)	39 (14.1)	22 (8.3)	13 (4.9)	**<0.01**
Minutes of exercise						**<0.01**
0−29	465 (43.1)	171 (63.3)	127 (45.8)	93 (35.1)	74 (27.6)	
>29−74	169 (15.6)	36 (13.3)	48 (17.3)	33 (12.5)	52 (19.4)	
>74−149	165 (15.3)	25 (9.3)	42 (15.2)	52 (19.6)	46 (17.2)	
>149	272 (25.2)	36 (13.3)	59 (21.3)	84 (31.7)	93 (34.7)	
Missing	9 (0.8)	2 (0.7)	1 (0.4)	3 (1.1)	3 (1.1)	
Self-rated health						**<0.01**
Excellent	339 (31.4)	43 (15.9)	84 (30.3)	90 (34.0)	122 (45.5)	
Very good/good	715 (66.2)	219 (81.1)	185 (66.8)	170 (64.2)	141 (52.6)	
Fair/poor	14 (1.3)	5 (1.9)	6 (2.2)	2 (0.8)	1 (0.4)	
Missing	12 (1.1)	3 (1.1)	2 (0.7)	3 (1.1)	4 (1.5)	
Systolic BP (mmHg)	114.1 ± 14.1	117.3 ± 14.8	116 ± 13.3	113 ± 13.7	110.2 ± 13.3	**<0.01**
Diastolic BP (mmHg)	66.7 ± 8.8	65.0 ± 8.4	66.3 ± 9.3	66.9 ± 8.6	68.5 ± 8.5	**<0.01**
Smoking status						**<0.01**
Never	682 (63.1)	149 (55.2)	169 (61.0)	180 (67.9)	184 (68.7)	
Former	372 (34.4)	112 (41.5)	103 (37.2)	83 (31.3)	74 (27.6)	
Current	19 (1.8)	7 (2.6)	4 (1.4)	1 (0.4)	7 (2.6)	
Missing	7 (0.6)	2 (0.7)	1 (0.4)	1 (0.4)	3 (1.1)	
Maximal exercise test						
VO_{2max} (mL/kg/min) (median ± SD)	21.6 ± 5.9	15.5 ± 2.5	19.8 ± 2.1	23.5 ± 2.2	28.8 ± 4.5	**<0.01**
Respiratory exchange ratio	1.2 ± 0.1	1.2 ± 0.1	1.2 ± 0.1	1.2 ± 0.1	1.2 ± 0.1	0.67
Borg score	16.5 ± 1.7	16.1 ± 1.7	16.2 ± 1.7	16.7 ± 1.7	17 ± 1.6	**<0.01**
Percent of maximum predicted HR[c]	98.8 ± 50.2	89.6 ± 13.3	97.5 ± 9.3	100.3 ± 8.5	107.8 ± 98.3	**<0.01**

Notes: Data are shown as mean ± SD or n (%) unless otherwise noted. Percentages may not add up to 100% due to rounding.
[a] Sex-specific quartile definitions were as follows:
Quartile 1: Men: <19.9; $n = 129$; Women: <16.5; $n = 141$.
Quartile 2: Men: ≥19.9 and ≤23.7; $n = 131$; Women: ≥16.5 and ≤19.9; $n = 146$.
Quartile 3: Men: >23.7 and ≤27.4; $n = 128$; Women: >19.9 and ≤23.7; $n = 137$.
Quartile 4: Men: >27.4; $n = 127$; Women: >23.7; $n = 141$.
[b] p value for continuous variables from the 1-way analysis of variance and χ^2 goodness of fit test for categorical variables across VO_{2max} quartiles.
[c] Maximum predicted HR = 220 − age.
* Bold indicates significance at the $p < 0.05$ level.
Abbreviations: BLSA = Baltimore Longitudinal Study of Aging; BMI = body mass index; BP = blood pressure; HR = heart rate; VO_{2max} = maximal oxygen uptake.

Table 2
RMSE for various ML algorithms.

Sample, universe of predictors	LASSO	XGBoost, linear	XGBoost, tree	Random forest	SVM
Total BLSA, all BLSA predictors	3.4	3.6	3.4	3.6	3.5
Total BLSA, OPACH predictors	3.5	3.6	3.6	3.7	3.6
BLSA men, all BLSA predictors	3.7	4.0	3.8	4.0	4.0
BLSA men, OPACH predictors	3.8	4.0	4.0	4.2	4.1
BLSA women, all BLSA predictors	2.8	3.2	3.0	2.9	2.8
BLSA women, OPACH predictors	2.9	3.2	3.1	3.1	3.0

Note: RMSE in units of VO_{2max} (mL/kg/min).
Abbreviations: BLSA = Baltimore Longitudinal Study of Aging; LASSO = least absolute shrinkage and selection operator; ML = machine learning; OPACH = Objective Physical Activity and Cardiovascular Health Study; RMSE = root mean squared errors; SVM = support vector machine; XGBoost = extreme gradient boosted.

subgroups, RMSEs ranged from 3.2 to 4.0 mL/kg/min for both women's algorithms and both men's algorithms (i.e., BLSA-predictor and OPACH-predictor algorithms), respectively. The tree-boosted XGBoost algorithm yielded an RMSE of 3.4 mL/kg/min for VO_{2max} prediction in the total sample using all BLSA predictors. For the subgroups, RMSEs ranged from 3.0 to 4.0 mL/kg/min for the women's BLSA-predictor and men's OPACH-predictor algorithms, respectively. The random forest algorithm yielded an RMSE of 3.6 mL/kg/min for the total sample using all predictors. For the subgroups, RMSEs ranged from 2.9 to 4.2 mL/kg/min for the women's BLSA-predictor and men's OPACH-predictor algorithms, respectively. The SVM algorithm yielded an RMSE of 3.5 mL/kg/min for the total sample using all predictors. For the subgroups, RMSEs ranged from 2.8 to 4.1 mL/kg/min for the women's BLSA-predictor and men's OPACH-predictor algorithms, respectively.

To summarize the performance of each algorithm, the LASSO and tree-boosted XGBoost algorithms had the lowest RMSE for the entire sample using the BLSA predictors (3.4 mL/kg/min). LASSO had the best RMSE for the entire sample when using the OPACH predictors (3.5 mL/kg/min). Further details about the combination of subgroups can be found in Table 2. Finally, for all algorithms the RMSE values for the women were lower than the RMSE values for the men.

3.3. Correlations of measured and predicted VO_{2max} with selected covariates

Correlations between measured VO_{2max}, all predicted VO_{2max} estimates, age, BMI, and SPPB are shown in Supplementary Table 1. In short, the correlations between predicted VO_{2max} and measured VO_{2max} ranged from 0.80 (OPACH-predictor linear-boosted XGBoost) to 0.93 (BLSA-predictor tree-boosted XGBoost). All predicted VO_{2max} estimates were more strongly associated with age, BMI, and SPPB than measured VO_{2max}.

3.4. Associations of measured and predicted VO_{2max} with all-cause mortality

When assessing the associations between quartiles of measured VO_{2max} and all-cause mortality, a steep inverse gradient in mortality risk across incremental VO_{2max} quartiles was evident in all models. Adjusting for Model 2 covariates, the HRs (95% confidence intervals (95%CIs)) were 0.55 (0.37−0.82), 0.30 (0.17−0.54), and 0.34 (0.15−0.75) for quartile 2 (Q2)−Q4 relative to Q1 of measured VO_{2max}, respectively, $p_{trend} < 0.01$ (Table 3). When evaluated in continuous format, every 1-SD increment (5.9 mL/kg/min) in measured VO_{2max} was associated with a 50% percent lower risk of all-cause mortality ($p < 0.01$) controlling for Model 2 covariates. The C-statistic for this model (95%CI) was 0.79 (0.75−0.83).

In the unadjusted models, every VO_{2max} prediction algorithm demonstrated patterns that were similar to those seen for measured VO_{2max}—that is, an inverse gradient in mortality risk across incremental predicted VO_{2max} quartiles (Q4 HRs ranged 0.09−0.17). However, adjusting for the Model 2 covariates attenuated the HRs for Q2−Q4, and while the majority of the 95%CIs widened to include 1.0, the significant trend across quartiles persisted except for the SVM-OPACH algorithm. After adjusting for Model 2 covariates, the HRs for a 1-SD increment in predicted VO_{2max} were similar to that seen for measured VO_{2max} (HRs ranged from 0.48 to 0.61). The C-statistics for all predicted VO_{2max} models were 0.78 and 0.79 after adjustment for Model 2 covariates (see Table 3 for the C-statistics' 95%CIs).

Among the BLSA men, there were 91 deaths: 53, 27, 8, and 3 in Q1−Q4 of measured VO_{2max}, respectively. Among the BLSA women, there were 50 deaths: 28, 11, 6, and 5 in Q1−Q4 of measured VO_{2max}, respectively. Sex-specific associations for measured and predicted VO_{2max} with all-cause mortality can be found in Supplementary Table 2 (men) and Supplementary Table 3 (women). In the unadjusted and adjusted models, higher measured VO_{2max} values are more strongly, inversely associated with risk of death in men than in women (Model 2 Q4 *vs.* Q1: men HR = 0.20 (0.06−0.70), $p_{trend} < 0.01$; women HR = 0.63 (0.21−1.90), $p_{trend} = 0.14$). This pattern of stronger inverse associations with mortality among men than women held for every predicted VO_{2max} estimate. In both the BLSA- and OPACH-predictor models, inverse trends were observed between increasing quartiles and mortality risk, with most HRs and trends achieving significance in men but fewer significant HRs and trends in women. Model 2 C-statistics were somewhat stronger for the men than the women.

Table 3
HRs of all-cause mortality by measured and predicted VO_{2max} in the BLSA ($n = 1080$).

Sample, universe of predictors	Model	Quartiles of VO_{2max} (mL/kg/min) or HR (95%CI)				p_{trend}	HR for 1-SD increase	p	C-statistic
		Q1	Q2	Q3	Q4				
		<17.8	≥17.8 and <21.6	≥21.6 and <25.7	≥25.7		5.9		
Measured VO_{2max}	1	1.00 (ref.)	0.43 (0.29−0.63)	0.16 (0.09−0.29)	0.10 (0.05−0.20)	<**0.01**	0.46 (0.38−0.57)	<**0.01**	0.71 (0.67−0.75)
Measured VO_{2max}	2	1.00 (ref.)	0.55 (0.37−0.82)	0.30 (0.17−0.54)	0.34 (0.15−0.75)	<**0.01**	0.50 (0.38−0.66)	<**0.01**	0.79 (0.75−0.83)
Total BLSA, all BLSA predictors									
Quartiles		<18.7	≥18.7 and <22.1	≥22.1 and <25.2	≥25.2		4.6		
XGBoost, linear	1	1.00 (ref.)	0.39 (0.26−0.58)	0.18 (0.10−0.30)	0.14 (0.07−0.26)	<**0.01**	0.53 (0.44−0.63)	<**0.01**	0.69 (0.65−0.73)
XGBoost, linear	2	1.00 (ref.)	0.62 (0.40−0.94)	0.44 (0.24−0.81)	0.72 (0.33−1.55)	**0.02**	0.61 (0.46−0.80)	<**0.01**	0.78 (0.74−0.82)
Quartiles		<18.5	≥18.5 and <22.1	≥22.1 and <25.4	≥25.4		5.0		
XGBoost, tree	1	1.00 (ref.)	0.36 (0.24−0.53)	0.18 (0.11−0.30)	0.09 (0.05−0.19)	<**0.01**	0.48 (0.39−0.58)	<**0.01**	0.71 (0.67−0.75)
XGBoost, tree	2	1.00 (ref.)	0.46 (0.30−0.69)	0.36 (0.20−0.63)	0.40 (0.17−0.91)	<**0.01**	0.49 (0.37−0.66)	<**0.01**	0.79 (0.75−0.83)
Quartiles		<18.6	≥18.6 and <22.3	≥22.3 and <25.5	≥25.5		5.0		
Random forest	1	1.00 (ref.)	0.47 (0.29−0.77)	0.19 (0.10−0.38)	0.11 (0.05−0.27)	<**0.01**	0.49 (0.38−0.62)	<**0.01**	0.69 (0.63−0.75)
Random forest	2	1.00 (ref.)	0.65 (0.39−1.07)	0.38 (0.18−0.78)	0.40 (0.15−1.07)	<**0.01**	0.52 (0.37−0.72)	<**0.01**	0.79 (0.73−0.85)
Quartiles		<18.7	≥18.6 and <22.2	≥22.2 and <25.5	≥25.5		4.6		
SVM	1	1.00 (ref.)	0.38 (0.23−0.64)	0.18 (0.09−0.34)	0.14 (0.06−0.30)	<**0.01**	0.54 (0.43−0.67)	<**0.01**	0.70 (0.64−0.76)
SVM	2	1.00 (ref.)	0.51 (0.30−0.87)	0.36 (0.17−0.76)	0.52 (0.20−1.37)	<**0.01**	0.57 (0.40−0.80)	<**0.01**	0.78 (0.72−0.84)
Total BLSA, OPACH predictors									
Quartiles		<18.7	≥18.7 and <22.2	≥22.2 and <25.2	≥25.2		4.5		
XGBoost, linear	1	1.00 (ref.)	0.40 (0.27−0.60)	0.23 (0.14−0.38)	0.11 (0.06−0.23)	<**0.01**	0.54 (0.45−0.64)	<**0.01**	0.69 (0.65−0.73)
XGBoost, linear	2	1.00 (ref.)	0.56 (0.37−0.86)	0.60 (0.34−1.05)	0.62 (0.27−1.45)	**0.04**	0.60 (0.45−0.81)	<**0.01**	0.78 (0.74−0.82)
Quartiles		<18.6	≥18.6 and <22.1	≥22.1 and <25.2	≥25.2		4.8		
XGBoost, tree	1	1.00 (ref.)	0.36 (0.24−0.54)	0.18 (0.11−0.31)	0.11 (0.05−0.22)	<**0.01**	0.50 (0.41−0.60)	<**0.01**	0.71 (0.67−0.75)
XGBoost, tree	2	1.00 (ref.)	0.50 (0.33−0.76)	0.37 (0.21−0.65)	0.58 (0.25−1.33)	<**0.01**	0.51 (0.38−0.69)	<**0.01**	0.79 (0.75−0.83)
Quartiles		<18.6	≥18.6 and <22.1	≥22.1 and <25.2	≥25.2		4.8		
Random forest	1	1.00 (ref.)	0.35 (0.23−0.55)	0.16 (0.09−0.30)	0.10 (0.04−0.21)	<**0.01**	0.46 (0.37−0.57)	<**0.01**	0.71 (0.67−0.75)
Random forest	2	1.00 (ref.)	0.50 (0.32−0.79)	0.33 (0.17−0.63)	0.36 (0.15−0.89)	<**0.01**	0.48 (0.35−0.67)	<**0.01**	0.78 (0.74−0.82)
Quartiles		<18.8	≥18.8 and <22.2	≥22.2 and <25.5	≥25.5		4.5		
SVM	1	1.00 (ref.)	0.52 (0.32−0.85)	0.19 (0.10−0.38)	0.17 (0.08−0.37)	<**0.01**	0.55 (0.44−0.69)	<**0.01**	0.68 (0.62−0.74)
SVM	2	1.00 (ref.)	0.75 (0.45−1.25)	0.44 (0.21−0.95)	1.00 (0.37−2.69)	0.16	0.59 (0.41−0.86)	<**0.01**	0.79 (0.73−0.85)

Notes: Model 1 = VO_{2max} quartiles, crude; Model 2 = Model 1 + age + race and ethnicity + education. Data in bold indicate significant.
Abbreviations: 95%CI = 95% confidence interval; BLSA = Baltimore Longitudinal Study of Aging; C-statistic = concordance statistic; HR = hazard ratio; OPACH = Objective Physical Activity and Cardiovascular Health Study; ref. = reference; SVM = support vector machine; VO_{2max} = maximum oxygen uptake; XGBoost = extreme gradient boosted.

3.5. Variable importance scores

Variable importance scores were obtained for the random forest, tree-boosted XGBoost, and linear-boosted XGBoost algorithms. The 5 most important variables in the linear-boosted all BLSA-predictor XGBoost algorithm were: (a) non-Hispanic other race, (b) usual gait speed in the 2.5-min walk, (c) history of myocardial infarction, (d) usual gait speed, and (e) being a former smoker. The 5 most important variables in the tree-boosted all BLSA-predictor XGBoost algorithm were, in order from more to less important: (a) number of seconds to complete the 400-m walk, (b) caloric expenditure from all activity, (c) caloric expenditure from exercise, (d) right-hand grip muscle strength, and (e) diastolic blood pressure. The 5 most important variables in the random forest all BLSA-predictor XGBoost algorithm were: (a) number of seconds to complete the 400-m walk, (b) the balance component of the SPPB, (c) meters walked in the 2.5-min walk, (d) 2.5-min gait speed, and (e) weight. In summary, when using all the variables in the BLSA, the number of seconds to complete the 400-m walk showed to be the most important variable across the random forest and tree-boosted XGBoost algorithms, and in the OPACH-predictor algorithms (i.e., in the absence of the 400-m walk), age became the most important variable. See Table 4 for the top 10 most important variables.

4. Discussion

We developed and assessed the performance of multiple ML, non-exercise-based VO_{2max} prediction algorithms that may enable large-scale epidemiologic cohorts with older, ambulatory, community-dwelling adults to accurately estimate VO_{2max}, an important biomarker of aging resiliency. The performance of all the ML algorithms evaluated in this study were reasonably good in relation to the performance of previously published RMSE values. Our RMSE values ranged from 2.8 to 4.2 mL/kg/min. For additional context, if one assumes the standard conversion of 3.5 mL/kg/min as being equivalent to 1 metabolic equivalent, the errors in VO_{2max} prediction based on the ML algorithms were about 0.8 and 1.2 metabolic equivalents. These predictive error values are lower than previously published non-exercise-based VO_{2max} prediction equations derived using ordinary least squares and lower than several RMSEs of previously published ML VO_{2max} prediction algorithms[12] Further, these non-exercise based predictive error values are comparable to those obtained when predicting VO_{2max} using exercise-based covariates such as the duration of maximal treadmill exercise tests[29] and timed walk tests.[30] These RMSE values, coupled with the strong correlations between predicted and measured VO_{2max}, further enhance our confidence in the VO_{2max} prediction algorithms described herein, even when performance-based assessment of CRF is not feasible.

For the total sample, the LASSO and tree-boosted XGBoost algorithms yielded the lowest RMSEs. When restricting to the OPACH predictors, LASSO had the lowest RMSE (3.5 mL/kg/min) followed by the 2 XGBoost algorithms and SVM at 3.6 mL/kg/min. Across all the algorithms, the RMSE values for women were lower than for men. This is likely due to the larger variation in men's VO_{2max} measurements compared to the women's. Despite the better prediction of VO_{2max} for the BLSA women than men, the associations between measured and predicted VO_{2max} and all-cause mortality were notably stronger for the men than the women, though the number of deaths in each quartile after stratifying by sex are few.

Minimal differences in RMSEs were observed when using the BLSA compared to OPACH covariate inputs, indicating that the variables that are not measured in OPACH are not critical to obtaining an accurate prediction of VO_{2max}, or at least that other variables were able to compensate for their absence using these ML approaches. For example, in the BLSA-predictor random forest algorithms, the number of seconds it took to complete the 400-m walk, an objective measure of physical performance capacity, is the most important variable in VO_{2max} prediction (RMSE = 3.6 mL/kg/min). However, since OPACH does not have a 400-m walk measure, age becomes the most important variable in the OPACH-predictor random forest algorithms; nonetheless, the effectiveness of this model for predicting VO_{2max} is nearly identical (RMSE = 3.7 mL/kg/min). Since age and physical performance capacity are inversely correlated, it could be that age serves as a proxy of physical performance in OPACH.

Few non-exercise-based VO_{2max} prediction ML models have been published to this point, and even fewer have been developed specifically for older adults. Our prior work assessing the performance of previously published linear regression models[31] showed that when these OLS models are used to predict VO_{2max} in the BLSA, the RMSE values range from 5.1 (using equations from Bradshaw et al.[32] and Sloan et al.'s[33] HR equation) to 20.4 (Jang et al.[34]) mL/kg/min. After recalibrating these formulas to measured VO_{2max} in the BLSA (i.e., obtaining new regression weights derived from the distribution of covariates in the BLSA) the RMSE values decrease to a range of 3.8 mL/kg/min (Bradshaw et al.[32]) to 4.2 mL/kg/min (Myers et al.[2]). A recent meta-analysis of 16 VO_{2max} prediction equations using ML,[12] few of which use non-exercise predictors and none of which were developed in older adults (the majority of the 16 equations were trained men and women in their mid- to late-20s; the oldest age range included in the meta-analysis was 18−65 years), found RMSEs (mL/kg/min) of 2.90 (SVM), 3.14 (multilayer perceptron neural network), 3.38 (tree boost), 4.78 (multilayer perceptron), 4.07 (artificial neural networks), 2.91 (feature selection with SVM), 3.37 (generalized regression neural networks), 4.51 (single decision tree), and 4.78 (multiple input single output with multilayer perceptron, SVM, and artificial neural networks with radial basis functions). Interestingly, in the multiple input single output model, the RMSEs were 4.07 for the women and 5.30 for the men, suggesting similar sex differences to those seen in the present study. The majority of the RMSEs in the algorithms for the present study outperform (lower RMSE values) those reported in this meta-analysis.

While several of the ML algorithms yielded reasonable predictions of $VO_{2max,}$ the utility of predicted VO_{2max} in estimating mortality risk was not as clear as measured VO_{2max}. In

Table 4
Top 10 important variables by algorithm.

Algorithm	Total BLSA, all BLSA predictors	Total BLSA, OPACH predictors	BLSA men, all BLSA predictors	BLSA men, OPACH predictors	BLSA women, all BLSA predictors	BLSA women, OPACH predictors
XGBoost, linear						
1	Non-Hispanic, other race	Non-Hispanic, other race	Non-Hispanic, other race	Male	High blood sugar despite medication	Fair/poor self-rated health
2	2.5-min gait speed	Very good/good self-rated health	≥150 min/week exercise	History: heart failure/CHF	2.5-min gait speed	Former smoker
3	History: heart attack or MI	History: heart attack or MI	30−74 min/week exercise	≥150 min/week exercise	30−74 min/week exercise	30−74 min/week exercise
4	Usual gait speed	History: breast cancer	History: heart attack or MI	Non-Hispanic, Asian/other Pacific Islander	Usual gait speed	Very good/good self-rated health
5	Former smoker	≥150 min/week exercise	Usual gait speed	Usual gait speed	History: heart attack or MI	≥150 min/week exercise
6	History: breast cancer	Former smoker	Never smoker	30−74 min/week exercise	Non-Hispanic, Asian/other Pacific Islander	Post-college education
7	Non-college graduate	Non-college graduate	Non-Hispanic, Asian/other Pacific Islander	History: diabetes	Former smoker	Non-Hispanic, Asian/other Pacific Islander
8	≥150 min/week exercise	History: diabetes	History: heart failure/CHF	Non-college graduate	History: stroke	Non-Hispanic, White
9	History: diabetes	75−149 min/week exercise	Non-Hispanic, White	History: heart attack or MI	Post-college education	BMI
10	Never smoker	Non-Hispanic, Asian/other Pacific Islander	High blood sugar despite medication	Non-Hispanic, White	Never smoker	History: heart attack or MI
XGBoost, tree						
1	400-m walk time	Age	400-m walk time	Age	400-m walk time	BMI
2	Calories from all activity	SF12: physical	Minutes of any walking/week	Diastolic BP	Calories per kg weight	Right-hand grip
3	Calories from exercise	Minutes of vigorous activity/ week	Radial pulse	Weight	Waist circumference	Height
4	Right-hand grip	Non-Hispanic, Black	Systolic BP	Right-hand grip	SF12: physical	Weight
5	Diastolic BP	Height	SF12: mental	SF12: mental	Minutes of vigorous activity/ week	Calories from all activity
6	Walking score	Minutes of any walking/week	2.5-min gait meters	Beta blocker use	Non-Hispanic, Black	Waist circumference
7	SF12: physical	Calories per kg weight	Calories per kg weight	Calories per kg weight	Minutes of any walking/week	Non-Hispanic, Black
8	Weight	Calories from all activity	Weight	Minutes of any walking/week	Weight	Radial pulse
9	Radial pulse	Beta blocker use	Calories from all activity	Calories from all activity	Radial pulse	Systolic BP
10	Waist circumference	Diastolic BP	Height	Systolic BP	Height	Diastolic BP
Random forest						
1	400-m walk time	Age	400-m walk time	Age	400-m walk time	Age
2	SPPB−balance	SPPB−chair stands	HR at end of 2.5-min walk	Left-hand grip	2.5-min gait meters	Calories from all activity
3	2.5-min gait meters	Waist circumference	Right-hand grip	SF12: physical	Calories per kg weight	Minutes of any walking/week
4	2.5-min gait speed	Calories per kg weight	Left-hand grip	Calories from all activity	SF12: physical	Left-hand grip
5	Weight	SF12: physical	Self-rated health	Right-hand grip	Calories from all activity	Height
6	Usual gait speed	Calories from all activity	Diastolic BP	Race/ethnicity	SPPB−chair stands	Right-hand grip
7	Height	Race/ethnicity	Calories per kg weight	Radial pulse	Calories from exercise	Race/ethnicity
8	Calories from exercise	Minutes of any walking/week	Rapid gait speed	Diastolic BP	HR at end of 2.5-min walk	Minutes of vigorous activity/ week
9	Waist circumference	Radial pulse	Walking score	Minutes of vigorous activity/ week	HR at end of 400-m walk	SF12: mental
10	SF12: physical	Minutes of vigorous activity/ week	Calories from all activity	Height	Minutes of vigorous activity/ week	Systolic BP

Abbreviations: BLSA = Baltimore Longitudinal Study of Aging; BMI = body mass index; BP = blood pressure; CHF = congestive heart failure; HR = heart rate; MI = myocardial infarction; OPACH = Objective Physical Activity and Cardiovascular Health in Older Women; SF12 = 12-item Short Form Survey of self-rated health; SPPB = Short Physical Performance Battery; XGBoost = extreme gradient boosted.

unadjusted models, all predicted VO_{2max} variables produced HRs comparable to measured VO_{2max}. However, after adjustment for even the limited set of covariates, these HRs were attenuated compared to measured VO_{2max}, though significant inverse trends in mortality risk remained evident in men (less so in women). The C-statistics were comparable for measured and predicted VO_{2max}. Direct measurement of VO_{2max} provides a more accurate representation of the underlying physiological construct of CRF than is possible using prediction. However, the present study indicates that ML prediction of VO_{2max} in older adults has relatively low prediction error and is associated with a clinical aging outcome (i.e., all-cause mortality) in a similar pattern and magnitude of association as measured VO_{2max} in unadjusted analysis. The attenuation of associations with mortality for predicted VO_{2max} but not measured VO_{2max} when adjusting for even a limited set of demographic covariates likely reflects the effect of controlling for factors correlated with mortality risk that were used in the prediction of VO_{2max}. Replication of the present investigation using large study samples with greater numbers of outcome events for analysis are needed to build upon our findings.

These findings should be confirmed and extended as ML algorithms continue to evolve to enable more precise estimations. The main limitation of this study, though not unique to it, would be the black box nature of these algorithms. However, and in direct response to the call for future research in the aforementioned ML meta-analysis,[12] we implemented the use of multiple ML methods to allow for meaningful comparisons of the algorithms' performances. Further, we compared these algorithms' associations with all-cause mortality for the total BLSA sample and by sex. To assess the transportability of these algorithms, we provided these metrics and associations with respect to a restricted universe of variables likely to be available in most aging studies. Another strength of our study is the prospective follow-up, enabling the evaluation of the accuracy of predicted VO_{2max} with respect to measured VO_{2max} and their associations with mortality. BLSA enrolled a large group of racially and ethnically diverse older adults, included objectively measured VO_{2max}, followed participants for mortality outcomes after VO_{2max} assessment, and collected data that enabled adjustment for confounders.

5. Conclusion

Measured VO_{2max} is a strong predictor of all-cause mortality in aging men and women enrolled in the BLSA, which further supports the recognition of VO_{2max} as a biomarker of aging resiliency. Given the infeasibility of direct measurement of VO_{2max} in large epidemiologic cohorts, simple linear regression models have been proposed to predict VO_{2max} and guide exercise prescription in older adults, but these more simplistic predicted VO_{2max} measures are not robust to adjustment in multivariable analyses. Using ML can improve the accuracy of VO_{2max} prediction as compared to simple OLS approaches but estimates of association with mortality remain sensitive to adjustments in multivariable analyses. Future studies should seek to reproduce these results to further improve the ability to predict VO_{2max} in community-dwelling older adults so that this "vital sign" can be more broadly studied as a modifiable target for promoting functional resiliency and healthy aging.

Acknowledgments

The authors acknowledge the BLSA participants and staff for their participation in this important scientific endeavor. We thank Sandy Liles for his thorough review of and contributions to this manuscript.

This research was supported in part by the Intramural Research Program of the National Institute on Aging. HPJr. was supported by the National Cancer Institute (K01 CA234317), the San Diego State University/UC San Diego Comprehensive Cancer Center Partnership (U54 CA132384 and U54 CA132379), and the Alzheimer's Disease Resource Center for Minority Aging Research at the University of California San Diego (P30 AG059299).

The content is solely the responsibility of the authors and does not necessarily represent the official views of funding agencies, which had no role in the study design, the collection, analysis, and interpretation of data, the writing of the report, or the decision to submit the manuscript for publication.

Authors' contributions

BTS led the conceptualization and methodology of the project, led the analysis, and wrote the original draft; MJL supported the conceptualization and methodology of the project, secured funding, and edited the manuscript; AZL supported the conceptualization and methodology of the project, secured funding, and edited the manuscript; EMS supported the conceptualization and methodology of the project, led the data curation, secured funding, and edited the manuscript; SPH supported the methodology of the project and edited the manuscript; HPJr. supported the methodology of the project and edited the manuscript; JB supported the methodology of the project and edited the manuscript; AK supported the conceptualization and methodology of the project, supported the analysis, and edited the manuscript. All authors have read and approved the final version of the manuscript, and agree with the order of presentation of the authors.

Competing interests

The authors declare that they have no competing interests.

Supplementary materials

Supplementary materials associated with this article can be found in the online version at doi:10.1016/j.jshs.2024.02.004.

References

1. Garber CE, Blissmer B, Deschenes MR, et al. American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: Guidance for prescribing exercise. *Med Sci Sports Exerc* 2011;**43**:1334–59.

2. Myers J, McAuley P, Lavie CJ, Despres JP, Arena R, Kokkinos P. Physical activity and cardiorespiratory fitness as major markers of cardiovascular risk: Their independent and interwoven importance to health status. *Prog Cardiovasc Dis* 2015;**57**:306–14.
3. Kodama S, Saito K, Tanaka S, et al. Cardiorespiratory fitness as a quantitative predictor of all-cause mortality and cardiovascular events in healthy men and women: A meta-analysis. *JAMA* 2009;**301**:2024–35.
4. Sui X, LaMonte MJ, Blair SN. Cardiorespiratory fitness as a predictor of nonfatal cardiovascular events in asymptomatic women and men. *Am J Epidemiol* 2007;**165**:1413–23.
5. Lee D, Artero EG, Sui X, Blair SN. Mortality trends in the general population: The importance of cardiorespiratory fitness. *J Psychopharmacol Oxf Engl* 2010;**24**(Suppl. 4):S27–35.
6. 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.
7. Phan DQ, Zheng C, Thai T, et al. Cardiorespiratory fitness and mortality in patients aged 60 to 90 years. *Am J Cardiol* 2022;**170**:132–7.
8. Shephard RJ. Maximal oxygen intake and independence in old age. *Br J Sports Med* 2009;**43**:342–6.
9. Kritchevsky SB, Forman DE, Callahan KE, et al. Pathways, contributors, and correlates of functional limitation across specialties: Workshop summary. *J Gerontol A Biol Sci Med Sci* 2019;**74**:534–43.
10. Matthews CE, Heil DP, Freedson PS, Pastides H. Classification of cardiorespiratory fitness without exercise testing. *Med Sci Sports Exerc* 1999;**31**:486–93.
11. de Souza E, Silva CG, Kaminsky LA, et al. A reference equation for maximal aerobic power for treadmill and cycle ergometer exercise testing: Analysis from the FRIEND registry. *Eur J Prev Cardiol* 2018;**25**:742–50.
12. Ashfaq A, Cronin N, Müller P. Recent advances in machine learning for maximal oxygen uptake (VO_{2max}) prediction: A review. *Inform Med Unlocked* 2022;**28**:100863. doi:10.1016/j.imu.2022.100863.
13. Vespa J, Medina L, Armstrong DM. *Demographic turning points for the United States: Population projections for 2020 to 2060 population estimates and projections current population reports*. Available at: www.census.gov/programs-surveys/popproj. [accessed 17.08.2021].
14. Stone JL, Norris AH. Activities and attitudes of participants in the Baltimore longitudinal study. *J Gerontol* 1966;**21**:575–80.
15. National Institute on Aging. *BLSA history*. Available at: http://www.nia.nih.gov/research/labs/blsa/history. [accessed 17.08.2021].
16. Ferrucci L. The Baltimore Longitudinal Study of Aging (BLSA): A 50-year-long journey and plans for the future. *J Gerontol A Biol Sci Med Sci* 2008;**63**:1416–9.
17. Simonsick E, Fan E, Fleg J. Estimating cardiorespiratory fitness in well-functioning older adults: Treadmill validation of the long distance corridor walk. *J Am Geriatr Soc* 2006;**54**:127–32.
18. Balke B, Ware RW. An experimental study of physical fitness of Air Force personnel. *U S Armed Forces Med J* 1959;**10**:675–88.
19. Cowper DC, Kubal JD, Maynard C, Hynes DM. A primer and comparative review of major U.S. mortality databases. *Ann Epidemiol* 2002;**12**:462–8.
20. Ware J, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: Construction of scales and preliminary tests of reliability and validity. *Med Care* 1996;**34**:220–33.
21. Kallman DA, Plato CC, Tobin JD. The role of muscle loss in the age-related decline of grip strength: Cross-sectional and longitudinal perspectives. *J Gerontol* 1990;**45**:M82–8.
22. Guralnik JM, Simonsick EM, Ferrucci L, et al. A Short Physical Performance Battery assessing lower extremity function: Association with self-reported disability and prediction of mortality and nursing home admission. *J Gerontol* 1994;**49**:M85–94.
23. Guralnik JM, Ferrucci L, Pieper CF, et al. Lower extremity function and subsequent disability: Consistency across studies, predictive models, and value of gait speed alone compared with the Short Physical Performance Battery. *J Gerontol Ser A* 2000;**55**:M221–31.
24. Simonsick EM, Montgomery PS, Newman AB, Bauer DC, Harris T. Measuring fitness in healthy older adults: The Health ABC Long Distance Corridor Walk. *J Am Geriatr Soc* 2001;**49**:1544–8.
25. Simonsick EM, Newman AB, Nevitt MC, et al. Measuring higher level physical function in well-functioning older adults: Expanding familiar approaches in the Health ABC study. *J Gerontol A Biol Sci Med Sci* 2001;**56**:M644–9.
26. Talbot LA, Metter EJ, Fleg JL. Leisure-time physical activities and their relationship to cardiorespiratory fitness in healthy men and women 18−95 years old. *Med Sci Sports Exerc* 2000;**32**:417–25.
27. LaCroix AZ, Rillamas-Sun E, Buchner D, et al. The Objective Physical Activity and Cardiovascular Disease Health in Older Women (OPACH) Study. *BMC Public Health* 2017;**17**:192. doi:10.1186/s12889-017-4065-6.
28. Harrell Jr FE, Lee KL, Mark DB. Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;**15**:361–87.
29. Pollock ML, Foster C, Schmidt D, Hellman C, Linnerud AC, Ward A. Comparative analysis of physiologic responses to three different maximal graded exercise test protocols in healthy women. *Am Heart J* 1982;**103**:363–73.
30. Mänttäri A, Suni J, Sievänen H, et al. Six-minute walk test: A tool for predicting maximal aerobic power (VO_{2max}) in healthy adults. *Clin Physiol Funct Imaging* 2018. doi:10.1111/cpf.12525. [Epub ahead of print].
31. Schumacher BT, Di C, Bellettiere J, et al. Validation, recalibration, and predictive accuracy of published $\dot{V}O_{2max}$ prediction equations for adults ages 50−96 yr. *Med Sci Sports Exerc* 2023;**55**:322–32.
32. Bradshaw DI, George JD, Hyde A, et al. An accurate VO_{2max} nonexercise regression model for 18−65-year-old adults. *Res Q Exerc Sport* 2005;**76**:426–32.
33. Sloan RA, Haaland BA, Leung C, Padmanabhan U, Koh HC, Zee A. Cross-validation of a non-exercise measure for cardiorespiratory fitness in Singaporean adults. *Singapore Med J* 2013;**54**:576–80.
34. Jang TW, Park SG, Kim HR, Kim JM, Hong YS, Kim BG. Estimation of maximal oxygen uptake without exercise testing in Korean healthy adult workers. *Tohoku J Exp Med* 2012;**227**:313–9.