UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Developing, Validating, and Applying Measurements of Relative Intensity Activity in Older Adults from Observational Accelerometry Studies

Permalink

https://escholarship.org/uc/item/3jh4p4vd

Author

Schumacher, Benjamin Troy

Publication Date

2022

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA SAN DIEGO

SAN DIEGO STATE UNIVERSITY

Developing, Validating, and Applying Measurements of Relative Intensity Activity in Older Adults from Observational Accelerometry Studies

A dissertation submitted in partial satisfactory of the

requirements for the degree of Doctor of Philosophy

in

Public Health (Epidemiology)

by

Benjamin Troy Schumacher

Committee in charge:

University of California San Diego

Professor Andrea LaCroix, Chair
Professor John Bellettiere
Professor Arun Kumar

San Diego State University

Professor Steven Hooker
Professor Humberto Parada

2022

The Dissertation of Benjamin Schumacher is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California San Diego

San Diego State University

2022

Table of Contents

List of Tables

Acknowledgements

I would first like to thank the chair of my committee, Dr. Andrea LaCroix, whose expertise, teaching, mentoring, and friendship was invaluable throughout my doctoral training. I would also like to thank my loved ones—family, friends, and classmates—who supported, encouraged, and loved me throughout my journey leading to my Ph.D. To my other professors, dissertation committee members, research advisors, and mentors, namely Dr. Caroline Thompson, thank you for patience and dedication to training me in and out of the classroom. To Ryan Smith, thank you for offering me an epidemiology internship that day in 2015 as I wouldn't be on the path I am today without my experience at the County. Lastly, to my lovely wife, Katie, thank you for standing by my side for better, for worse, for richer, for poorer, 'til death do us part.

Chapter 2, in part, is currently being prepared for submission for publication of the material. Schumacher, Benjamin T.; Di, Chongzhi; Bellettiere, John; LaMonte, Michael J.; Simonsick, Eleanor M.; Parada, Humberto; Hooker, Steven; LaCroix, Andrea Z. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in part, is currently being prepared for submission for publication of the material. Schumacher, Benjamin T.; LaMonte, Michael J.; LaCroix, Andrea Z.; Simonsick, Eleanor M.; Hooker, Steven P.; Parada, Humberto; Bellettiere, John; Kumar, Arun. The dissertation author was the primary investigator and author of this paper.

Chapter 4, in part, is currently being prepared for submission for publication of the material. Schumacher, Benjamin T.; LaMonte, Michael J.; Di, Chongzhi; Simonsick, Eleanor M.; Parada, Humberto; Hooker, Steven P.; Bellettiere, John; LaCroix, Andrea Z. The dissertation author was the primary investigator and author of this paper.

Vita

Education

2022 Doctor of Philosophy, University of California San Diego & San Diego State University

2018 Master of Public Health, San Diego State University

2016 Bachelor of Science, University of California San Diego

Publications

Garduno, A., LaCroix, A., LaMonte, M., Dunstan, D., Evenson, K., Wang, G., Di, C., **Schumacher, B.T.**, Bellettiere, J. (2022). Associations of Daily Steps and Step Intensity With Incident Diabetes in a Prospective Cohort Study of Older Women: The OPACH Study. Diabetes Care. https://doi.org/10.2337/dc21-1202

Evenson, K., Bellettiere, J., Cuthbertson, C., Di, C., Dushkes, R., Howard, A., Parada Jr., H., **Schumacher, B.T.**, Shiroma, E., Wang, G., Lee, I., LaCroix, A. (2021). Cohort Profile: The Women's Health Accelerometry Collaboration. BMJ Open. https://doi.org/10.1136/bmjopen-2021-052038

Barnette, B., **Schumacher, B.T.**, Armenta, R., Wynn, J., Richardson, A., Bradley, J., Lazar, S., Lawrence, S. (2021). Contribution of Concurrent Comorbidities to Sepsis-Related Mortality in Preterm Infants ≤ 32 Weeks of Gestation at an Academic Neonatal Intensive Care Network Journal of Perinatology. American Journal of Perinatology. Advance online publication. https://doi.org/10.1055/a-1675-2899

Schumacher, B.T., Bellettiere, J., LaMonte, M., Evenson, K., Di, C., Lee, I., Sleet, D., Eaton, C., Lewis, C., Margolis, K., Tinker, L., & LaCroix, A. (2021). Accelerometer-Measured Daily Steps, Physical Function, and Subsequent Fall Risk in Older Women: The Objective Physical Activity and Cardiovascular Disease in Older Women Study. Journal of Aging and Physical Activity. Advance online publication. https://doi.org/10.1123/japa.2021-0159

Bandoli, G., Jelliffe-Pawlowski, L., **Schumacher, B.T.**, Baer, R., Felder, J., Fuchs, J., Oltman, S., Steurer, M., Marienfeld, C. (2021). Cannabis-related diagnosis in pregnancy and adverse maternal and infant outcomes. Drug and Alcohol Dependence. 225, 1 August 2021, 108757. https://doi.org/10.1016/j.drugalcdep.2021.108757

Thompson, C., Jin, A., Luft, H., Lichtensztajn, D., Allen, L., Liang, S., **Schumacher, B.T.**, Gomez, S. (2020). Population-Based Registry Linkages to Improve Validity of Electronic Health Record–Based Cancer Research. Cancer Epidemiology, Biomarkers & Prevention. Volume 29, Issue 4. https://doi.org/10.1158/1055-9965.EPI-19-0882

Abstract of the Dissertation

Developing, Validating, and Applying Measurements of Relative Intensity Activity in Older Adults from Observational Accelerometry Studies

by

Benjamin Troy Schumacher

Doctor of Philosophy in Public Health (Epidemiology)

University of California San Diego, 2022

San Diego State University, 2022

Professor Andrea Z. LaCroix, Chair

Background: Regular physical activity (PA) reduces the risk of chronic diseases, slows the progression of prevalent chronic diseases, and promotes other health benefits. One's level of energy expenditure while performing an activity (*absolute* intensity) may be discordant with their level of exertion relative to their maximal possible effort (*relative* intensity). $\dot{V}O_{2max}$, the gold standard measurement of cardiorespiratory fitness, can be used to individualize *absolute*

activity relative to maximum effort. If the $\dot{V}O_{2max}$ of an individual is known and their instantaneous oxygen uptake ($\dot{V}O_2$) can be measured, (($\dot{V}O_2$ / $\dot{V}O_{2max}$)*100) gives the percent of their maximal exercise capacity (i.e., their *relative* intensity).

Methods: Aim 1 assessed the performance of published $\dot{V}O_{2max}$ prediction equations in relation to measured $\dot{V}O_{2max}$ and recalibrated the equations using the Baltimore Longitudinal Study of Aging (BLSA). Aim 2 developed new machine-learned (ML) $\dot{V}O_{2max}$ prediction algorithms in the BLSA. In Aim 3, daily hours spent in light and moderate-to-vigorous PA (MVPA) were calculated on the *absolute* scale (accelerometer-measured PA) and *relative* scale (accelerometer-estimated $\dot{V}O_2$ / predicted $\dot{V}O_{2max}$ using Aim 2's algorithms). The associations between *absolute* and *relative* intensity PA, total mortality, and incident major cardiovascular disease (CVD) were estimated in the Objective Physical Activity and Cardiovascular Health (OPACH) Study.

Results: In Aim 1, the prediction equations yielded root mean squared error (RMSE) values ranging from 4.2-20.4 $mL \cdot kg^{-1} \cdot min^{-1}$ and from 3.9-4.2 $mL \cdot kg^{-1} \cdot min^{-1}$ after recalibration. The newly developed ML algorithms in Aim 2 yielded RMSE values ranging from 2.9-4.4 $mL \cdot kg^{-1} \cdot min^{-1}$. In Aim 3, on each PA measurement scale (relative and absolute), higher levels of light PA and MVPA were associated with reduced risk of both outcomes. On the *absolute* scale, MVPA was more strongly associated with both outcomes than light intensity, but on the relative scale, light intensity was more strongly associated with both outcomes.

Conclusion: The PA intensity paradigm should keep shifting towards recommendation of more movement, regardless of intensity, and placing greater emphasis on *relative* light intensity (37%-46% of maximal capacity) as modifiable behavioral targets that are more easily achieved, reduce the risk of CVD and death, and promote healthy aging.

1. Introduction

1.1. Physical Activity and Health in Older Adults

For people of all ages, regular physical activity (PA) is known to reduce the risk of developing new chronic diseases, slow the progression of prevalent chronic diseases, and promote a myriad of other health benefits.[1] Older adults (≥ 65 years) that engage in the recommended amounts of physical activity have a reduced risk of developing/experiencing: dementia, cancers (specifically of the breast, colon, bladder, endometrial, kidney, lung, and stomach), falls, among many other deleterious health outcomes.[1] Further, physical activity is also known to improve executive function, sleep quality, and overall quality of life.[1] Despite the benefits from engaging in regular PA, the proportion of older adults that meet the PA guidelines was only 28%, according to a 2016 study.[2]

1.2. Classification and Measurement of Physical Activity

PA can be defined as "bodily movement produced by skeletal muscles that results in energy expenditure"[3] and has a wide range of intensity categories: light, moderate, vigorous, and a frequently used aggregate category of moderate-to-vigorous PA (MVPA).[1] The PA intensity category is determined by the amount of energy expended while completing a task, most commonly measured in metabolic equivalents (METs).[1] One MET is the amount of metabolic energy that is expended while sitting quietly at rest, which, for most individuals takes 3.5 milliliters of oxygen (mL) per kg of body weight (kg) per minute (min) ($mL \bullet kg^{-1} \bullet min^{-1}$). Therefore, an activity of five METs requires five times the amount of energy expended (or 17.5 $mL \bullet kg^{-1} \bullet min^{-1}$) while sitting at rest. Categories of PA intensity have been defined as ≤ 1.5, 1.6 - 2.9, 3.0 - 5.9, and ≥ 6.0 METs as sedentary behavior (SB), light, moderate, and vigorous PA, respectively.[1] Lastly, for an activity to truly be considered as SB, one's energy expenditure must

be ≤ 1.5 METs *and* they must be sitting, reclining, or lying.[1,4] The measurement of PA in epidemiologic cohort studies is generally conducted using triaxial accelerometers. These accelerometers are generally placed on the right hip and are worn for 7 days. When analyzing accelerometry data, accelerations from all three axes are aggregated into vector magnitude (VM) counts over a recording period (epoch, usually 15-seconds or one minute) to summarize activity within that epoch.[5] The VM counts per epoch can then be used to classify the intensity of activity engaged in during that epoch. Epochs can then be aggregated across the device wear time to yield an objective measurement of PA and PA intensity over a specified time interval (e.g. per hour or day).

1.3. Absolute vs. Relative Intensity Activity

According to the 2018 U.S. Physical Activity Guidelines Advisory Committee (PAGAC) Scientific Report, PA intensity can be measured on an absolute or relative scale.[1] The absolute intensity scale focuses solely on the amount of energy expenditure needed to complete an activity, while the relative intensity scale considers the absolute energy expenditure in relation to one's maximal possible effort.[1] The aforementioned MET-based categories of PA intensity (see Chapter 1.2.) are on the absolute scale. The absolute intensity scale assumes that energy expenditure for a given activity is the same for all individuals and does not account for health status, age, cardiorespiratory fitness level, or any other observed or unobserved phenotype. Therefore, published, commonly used absolute intensity categories may adequately correlate with perceived exertion for a middle-aged, generally healthy adult.[1] However, given that older adults have a lower resting metabolic rate[6] coupled with an increased energy cost of movement[7], the absolute intensity categories will likely have lower correlations with perceived exertion and underestimate amounts of moderate and vigorous PA in older populations.[1] This discordance

between the absolute energy expenditure and perceived level of effort is not exclusive to older adults, but also applies to those in poor health, those with comorbidities, and those with poor cardiorespiratory fitness. The following adapted example[8] more clearly illustrates the differences between absolute and relative intensity activity. Assume walking at a constant, slow pace requires 3 METs. For a younger adult capable of 12 MET activities, this slow walk requires minimal effort relative to their 12 MET capacity (3/12 = 25% of their maximal effort). For an older adult, capable of 5 METs, this same walk requires much more relative effort (3/5 = 60% of their maximal effort). Both adults are exerting the same amount of energy on an absolute scale (3 METs) but the exertion relative to their maximum is quite disparate.

1.4. Percent Maximal Oxygen Uptake as a Relative Intensity Activity Metric

The capacity of the circulatory and respiratory systems to deliver oxygen to skeletal muscles for use during physical activity and exercise can be quantified by one's cardiorespiratory fitness (CRF) level.[9] Maximal oxygen uptake, $\dot{V}O_{2max}$, is the gold standard measurement of cardiorespiratory fitness.[9] $\dot{V}O_{2max}$ measures the volume of oxygen (O_2) that physiologic systems can uptake and utilize during activity.[10,11] $\dot{V}O_{2max}$ is measured using a gas exchange monitoring system attached to the participant's face, covering the nose and mouth, while a graded exercise test is conducted on a stationary bicycle or treadmill. Participants will begin walking or biking at a slow speed, and the incline of the device will be increased in a graded, stepwise fashion until the participant indicates they have reached total exhaustion. Thus, $\dot{V}O_{2max}$ marks the point in which one's body cannot increase oxygen (O_2) consumption and utilization despite an increase in the requested workload. This volume of O_2, measured in milliliters (mL), is standardized per kilogram (kg) of body mass, and per minute of exercise, yielding units of $\dot{V}O_{2max}$ to be $mL \bullet kg^{-1} \bullet min^{-1}$. Traditionally, $\dot{V}O_{2max}$ can be converted to METs

by dividing by 3.5, though studies have shown that a conversion factor of 3.0 for older adults more accurately captures the decrease in resting metabolic rate.[5] If the $\dot{V}O_{2max}$ of an individual is known and their instantaneous oxygen uptake ($\dot{V}O_2$) can be measured, then [($\dot{V}O_2$ / $\dot{V}O_{2max}$)*100 or (METs / maximal MET capacity)*100] gives the percent of their maximal exercise capacity (i.e., the *relative* intensity of their effort). The American College of Sports Medicine (ACSM) proposes the following relative intensity categories based on percent of maximal effort: < 37% as very light, 37 – 45% as light, 46 – 63% as moderate, 64 – 90% as vigorous, and ≥ 91% as maximal effort.[9]

1.5. Review of Epidemiologic Studies on Relative Intensity Activity

To my knowledge, this dissertation is the first study to assess relative vs. absolute intensity of physical activity with an estimated percent of maximal exercise capacity. As direct measurements of instantaneous $\dot{V}O_2$ and $\dot{V}O_{2max}$ require specialized equipment, trained personnel, the presence of a licensed physician (for $\dot{V}O_{2max}$), and extensive economic resources that are generally not feasible for large epidemiologic cohort studies, only indirect estimations of relative intensity have been used, e.g. the talk test[12] and the Borg Rating of Perceived Exertion (RPE).[13] In short, the Borg Scale is a subjective survey completed by the participant during PA that ranges from "no exertion at all" to "maximal exertion" and the talk test states that, generally, while engaged in relative moderate intensity activity, one can talk but not sing, and while engaged in vigorous activity one can hardly talk.

In a prospective cohort study of 7,337 men in the Harvard Alumni Health Study (mean age: 66 years), participants rated their usual level of exertion when exercising on a 10-point Borg Scale, categorized as 0 to 2 ("nothing to weak"), 3 ("moderate"), 4 ("somewhat strong"), and ≥ 5 ("strong to maximal").[14] Adjusted relative risks (RR (95% CI)) of coronary heart disease (CHD)

for men reporting usual perceived exertion as "moderate," "somewhat strong," and "strong to maximal" were: 0.86 (0.66-1.13), 0.69 (0.51-0.94), and 0.72 (0.52-1.00), respectively (P_{trend} = 0.02), when compared with "nothing to weak".[15] The authors of the study note that, "[the finding] suggests that physical activity recommendations need to be tailored to the individual and that global requirements for activities of ≥ 3 METs may not be appropriate, especially for older persons."[14]

In a study from the same cohort as used in parts of this dissertation, the Baltimore Longitudinal Study of Aging (BLSA), 619 healthy men and 497 healthy women had their activity assessed on both the relative and absolute scales.[16] The proportion of participants meeting the national recommendations for moderate and high intensity PA on an *absolute* intensity scale decreased with age, but this same proportion increased when activity was assessed on a *relative* intensity scale, further exemplifying the need to measure activity of older and younger adults on different scales. Talbot et al. assert that more older adults are compliant with national PA recommendations on a *relative* intensity scale and that the *absolute* intensity scale is inappropriate to measure and motivate older adults' physical activity.[16]

1.6. Specific Aims

In this dissertation, I examined the performance of VO_{2max} prediction equations, use machine-learned (ML) methods to develop new VO_{2max} prediction algorithms, and then assess relationships between absolute and relative intensity PA and selected health outcomes. The following aims are addressed:

Aim 1: Quantify the association between $\dot{V}O_{2max}$, both measured and as predicted from numerous published prediction equations, and all-cause mortality in the

Baltimore Longitudinal Study of Aging. Several published non-exercise $\dot{V}O_{2max}$ prediction models will be used to predict $\dot{V}O_{2max}$ in the Baltimore Longitudinal Study of Aging (BLSA). After assessing performance metrics for each predicted $\dot{V}O_{2max}$ relative to laboratory-measured $\dot{V}O_{2max}$, the equations will be recalibrated to measured $\dot{V}O_{2max}$, and associations between the predicted $\dot{V}O_{2max}$, recalibrated predicted $\dot{V}O_{2max}$, and all-cause mortality will be quantified.

Aim 2: Train multiple ML algorithms to develop non-exercise based VO_{2max} prediction algorithms in the BLSA. Given the logistical challenges of measuring $\dot{V}O_{2max}$ in large, epidemiologic cohorts, there exists a need for an accurate, reliable $\dot{V}O_{2max}$ prediction models that can be broadly applied to epidemiologic cohorts. These $\dot{V}O_{2max}$ prediction models will be trained using all covariates in the Baltimore Longitudinal Study of Aging, after restricting to commonly available non-exercise covariates to increase the transportability of these algorithms to external epidemiologic cohorts, and within sex-specific strata.

Aim 3: Estimate the associations between absolute intensity PA, relative intensity PA, total mortality, and incident major CVD. Using accelerometry data from the Objective Physical Activity and Cardiovascular Health (OPACH) Study, VM counts in each epoch will be used to categorize absolute intensity of activity. Epoch-level accelerometer-estimated METs will be divided by each persons predicted VO_{2max} (using the best performing algorithms from Aim 1 and Aim 2) to yield percent maximal effort in that epoch. Percent maximal effort will be classified using the ACSM's categorization scheme. The associations between absolute intensity PA, relative intensity PA, total mortality, and incident major CVD will then be quantified and compared.

2. Validation, Recalibration, and Predictive Accuracy of Published $\dot{V}O_{2max}$ Prediction Equations for Older Adults

Benjamin T. Schumacher, Chongzhi Di, John Bellettiere, Michael J. LaMonte, Eleanor M. Simonsick, Humberto Parada Jr., Dr. Steven P. Hooker, Andrea Z. LaCroix

2.1. Abstract

Background: Maximal oxygen uptake ($\dot{V}O_{2max}$) is the criterion measure of cardiorespiratory fitness (CRF). Lower CRF is a strong predictor of poor health outcomes, including all-cause mortality. Since $\dot{V}O_{2max}$ testing is resource intensive, several non-exercise based $\dot{V}O_{2max}$ prediction equations have been published. We assess these equations' ability to predict measured $\dot{V}O_{2max}$, recalibrate these equations, and quantify the association of measured and predicted $\dot{V}O_{2max}$ with all-cause mortality.

Methods: Baltimore Longitudinal Study of Aging participants with valid $\dot{V}O_{2max}$ tests were included (n=1,080). Using published $\dot{V}O_{2max}$ prediction equations, we calculated predicted $\dot{V}O_{2max}$ and present performance metrics before and after recalibration (deriving new regression estimates by regressing measured $\dot{V}O_{2max}$ on BLSA). Cox proportional hazards models were fit to quantify associations of measured, predicted, and recalibration-predicted values of $\dot{V}O_{2max}$ with mortality.

Results: Mean age and $\dot{V}O_{2max}$ were 69.0±10.4 years and 21.6±5.9 $mL \bullet kg^{-1} \bullet min^{-1}$, respectively. The prediction equations yielded root mean squared error values ranging from 4.2-20.4 $mL \bullet kg^{-1} \bullet min^{-1}$. After recalibration, these values decreased to 3.9-4.2 $mL \bullet kg^{-1} \bullet min^{-1}$. Adjusting for all covariates, all-cause mortality risk was 66% lower for the highest quartile of measured $\dot{V}O_{2max}$

relative to the lowest. Predicted $\dot{V}O_{2max}$ variables yielded similar estimates in unadjusted models but were not robust to adjustment.

Conclusion: Measured $\dot{V}O_{2max}$ is an extremely strong predictor of all-cause mortality. Several published $\dot{V}O_{2max}$ prediction equations yielded: (1) reasonable performance metrics relative to measured $\dot{V}O_{2max}$, especially when recalibrated, (2) all-cause mortality hazard ratios similar to those of measured $\dot{V}O_{2max}$, especially when recalibrated, yet (3) were not robust to adjustment for basic demographic covariates likely because these were used in the equation for predicted $\dot{V}O_{2max}$.

2.2. Introduction

The capacity of the circulatory and respiratory systems to deliver oxygen to skeletal muscles for use during physical activity and exercise can be quantified by one's cardiorespiratory fitness (CRF) level [9]. CRF is a physiological attribute determined by several factors including age, sex, health status, and genetics; however, the principal modifiable determinant is habitual physical activity (PA) level [9]. Through increases in the frequency, duration, and intensity of PA, CRF can incrementally increase, especially among the sedentary, though CRF declines soon after the frequency, duration, and/or intensity of PA declines. Thus, CRF often is used as an objective surrogate of recent PA patterns. Decades of clinical, epidemiologic, and exercise science studies have reported that higher CRF is a strong and independent predictor of a myriad of beneficial health outcomes [17–19]. Low CRF is among the strongest predictors of cardiovascular and all-cause mortality, with associations as strong or stronger as those of smoking, obesity, and high blood pressure with the same outcomes [20,21]. Likewise, higher CRF is associated with lower: coronary heart disease/cardiovascular disease incidence and mortality [22–24], incidence of cardiometabolic risk factors [25,26], cancer incidence and cancer mortality [27–29], dementias [30] including Alzheimer's disease [31] and their progression, depression symptoms [32,33], rates of loss of independence for older adults [34], and all-cause mortality [20,21,23,35].

The gold standard measure of CRF is maximal oxygen uptake ($\dot{V}O_{2max}$)[9]. In research settings, $\dot{V}O_{2max}$ measurements are conducted using maximal graded exercise tests on a treadmill or stationary cycle ergometer and require specialized testing equipment, highly trained personnel, and direct physician supervision in most instances. Further, in vulnerable populations such as older adults, $\dot{V}O_{2max}$ testing may be contraindicated as it requires maximal, strenuous

activity to the point of absolute exhaustion. Thus, conducting direct measures of $\dot{V}O_{2max}$ in large epidemiologic cohort studies is largely infeasible [36]. As an alternative approach, several non-exercise based $\dot{V}O_{2max}$ prediction equations have been published to enable the approximation of $\dot{V}O_{2max}$ in a variety of settings, including large epidemiologic cohorts [22,37–44]. However, few equations have been developed specifically for use in older adult populations [40,42]. There is a critical need for accurate $\dot{V}O_{2max}$ prediction models in older adults, given that by the year 2060, almost a quarter of the United States (U.S.) population will be comprised of adults 65 years of age or older (i.e., older adults) [45] and $\dot{V}O_{2max}$ has been identified as a hallmark biomarker of successful aging [46]. Given the shifting demographics, the challenges older adults face with $\dot{V}O_{2max}$ testing, and the benefits of increased CRF on health, we aimed to quantify the performance of published $\dot{V}O_{2max}$ prediction models in relation to measured $\dot{V}O_{2max}$ in the Baltimore Longitudinal Study of Aging (BLSA), recalibrate the equations to the BLSA cohort, and assess their predictive accuracy in relation to all-cause mortality.

2.3. Methods

Study Population

The analytic sample for the present study was derived from the BLSA, the longest running scientific study of aging [47,48]. The BLSA was established in 1958 and is conducted by the National Institute on Aging Intramural Research Program [49]. BLSA participants have been asked to visit the BLSA testing facility every one to four years to undergo a three-day battery of health, cognitive, and functional evaluations. More than 3,000 participants have participated in the BLSA since its inception, and over 1,300 participants are still active [47]. To date, 1,080 BLSA participants have had laboratory-based $\dot{V}O_{2max}$ measurements that meet criteria for a maximal test. Extensive details about the design, recruitment, and measurements collected in the BLSA

have been published elsewhere [48]. This study was approved by the relevant Institutional Review Boards and all participants provided written informed consent.

Measures

$\dot{V}O_{2max}$ Measurement

$\dot{V}O_{2max}$ (measured in milliliters of O_2 uptake / kilogram body weight / minute; $mL \bullet kg^{-1} \bullet min^{-1}$) was assessed in the BLSA using a modified Balke treadmill testing protocol [50,51]. This protocol consists of a graded exercise test; walking on a treadmill at a constant pace at 3.0 miles per hour (mph) for women and 3.5 mph for men, with the incline of the treadmill increasing 3% every 2 minutes until the participant indicates they have reached exhaustion. During this test, expired gas volumes were measured using a Parkinson-Cowan gas meter and concentrations of oxygen and carbon dioxide were measured using a medical mass spectrometer (Perkin-Elmer MGA-1110), which was calibrated daily using standard gases. A computerized interface between the gas meter and mass spectrometer calculated average expired gas concentrations every 30 seconds throughout the test and the highest 30-second value for O_2 uptake defined the participant's $\dot{V}O_{2max}$.

Achievement of maximal effort during the treadmill test was defined as reaching a respiratory exchange ratio (RER) >1.0. Fifty-two participants had a $\dot{V}O_{2max}$ test just below this RER cutoff when the treadmill test was stopped. Of these 52 participants, 11 achieved ≥ 85% of their age-predicted maximal heart rate (beats per minute, bpm; calculated as 220 – age) and a Borg rating of perceived exertion (RPE) ≥17 on a 20-point scale, so their tests were considered to reflect maximal effort and were included in the present analysis. Of the remaining 41 participants with an RER <1.0 at the time the treadmill was stopped, 31 were excluded because

they had no other $\dot{V}O_{2max}$ test that met the aforementioned maximal effort criteria, and 10 participants were included who provided a subsequent $\dot{V}O_{2max}$ test that fit the criteria for a maximal test, resulting in a final analytic sample of 1,080. For participants with multiple $\dot{V}O_{2max}$ measurements, the first measurement satisfying these criteria was used in the present study.

Non-Exercise Test $\dot{V}O_{2max}$ Prediction Models

Google Scholar was used to query previously published studies using the terms "non-exercise based $\dot{V}O_{2max}$ prediction models" and "older adults", yielding a total of 12 $\dot{V}O_{2max}$ prediction equations from nine published studies that were assessed in the present study. Studies that developed $\dot{V}O_{2max}$ prediction equations derived solely for younger populations, were developed using any form of exercise testing or physical performance as a predictor of $\dot{V}O_{2max}$, or included variables in the prediction equation not available in the BLSA were not included in the present study. Each prediction equation included sex, age, and some measure of body mass. Some equations additionally included variables such as self-reported PA scores, smoking history, height, and resting heart rate. In the present analysis, covariates in the published $\dot{V}O_{2max}$ prediction equations were matched with their closest equivalent covariate in the BLSA.

Outcome Ascertainment

All-cause mortality status and date of death were ascertained by linking participants to the National Death Index, a centralized database of death record information compiled from state vital statistics records, and by correspondence from relatives [52]. Follow-up for mortality occurred from first $\dot{V}O_{2max}$ test date, the earliest of which was January 1st, 2007, through April 15th, 2021. Mortality ascertainment was high with 96% of participants having a classified vital status. Over a

median follow-up time of 9.6 years (range: 0.60 – 14.1 years), 141 participants died from any cause.

Covariates

Covariates for the $\dot{V}O_{2max}$ prediction equations or their closest approximations in the BLSA included participant's sex, age, body mass index (BMI), resting heart rate, self-reported PA/exercise level, self-rated general health status, and smoking history. In the BLSA, a participant's sex and age were self-reported during each health history interview. Height and weight were measured using a stadiometer and calibrated scale, respectively, and BMI was calculated as weight in kilograms divided by height in meters squared. Resting heart rate was assessed by a nurse after the participant had been sitting quietly for at least 5 minutes [53]. Participants were asked how much time they spent each week engaging in weight/circuit training, moderate-to-high intensity exercise, or brisk walking which was then categorized as: 0–29 (coded as 0), 30–74 (1), 75–149 (2), or ≥150 (3) minutes. Health-related quality-of-life was assessed using the 12-item short form health survey (SF-12) [54]. Smoking history (never, current, or former smoker) was self-reported using a standardized questionnaire [55]. The following covariates were not used in any $\dot{V}O_{2max}$ prediction models, but were employed in the description of the study sample: self-reported race (White, Black, Asian/Other Pacific Islander, Other/not classifiable), self-reported educational attainment (non-college graduate, college graduate, post-college graduate), beta blocker use (yes or no), systolic and diastolic blood pressures (mmHg; oscillometric brachial blood pressure was measured with the participant in a supine position on both arms three times and the minimum systolic and diastolic blood pressures were used).

Statistical Analysis

We compared covariates by sex-specific quartiles of measured $\dot{V}O_{2max}$ using chi-square tests for categorical variables and analysis of variance (ANOVA) tests for continuous variables.

Predicted $\dot{V}O_{2max}$ was calculated using each $\dot{V}O_{2max}$ prediction equation as originally published. The performance (ability to accurately predict measured $\dot{V}O_{2max}$) of each equation was evaluated by comparing the predicted $\dot{V}O_{2max}$ to the measured $\dot{V}O_{2max}$ using the root mean square error (RMSE), bias, mean absolute percentage error (MAPE), the Bland-Altman 95% Limits of Agreement (LOA) [56], correlation coefficients, and R^2. These analyses were conducted in the overall sample and within sex strata. In brief, RMSE quantifies the concentration of the data around the line of best fit by estimating the square of all predicted $\dot{V}O_{2max}$ minus measured $\dot{V}O_{2max}$ pairs, taking the mean of these squared differences, and obtaining the square root of the mean squared errors. Bias was computed by taking the mean of the measured $\dot{V}O_{2max}$ minus predicted $\dot{V}O_{2max}$ pairs. MAPE was computed by taking the mean of the absolute value of the percent deviation of the predicted $\dot{V}O_{2max}$ from the measured $\dot{V}O_{2max}$. The lower the RMSE, bias, and MAPE, the better the performance of the prediction model, with 0 indicating perfect prediction of the measured $\dot{V}O_{2max}$. The calculation for the Bland-Altman 95% LOA has been described elsewhere, but these limits are expected to capture 95% of the differences between measured and predicted $\dot{V}O_{2max}$; a more narrow range of limits indicates a better prediction [56]. The Bland-Altman 95% LOA were obtained using the blandr package in R [57].

Because the accuracy of each $\dot{V}O_{2max}$ prediction equation is strongly influenced by the distribution of covariates and measured $\dot{V}O_{2max}$ in the source population from which the equation was derived, the application of a prediction equation from one population to another can affect predictive accuracy. Therefore, each $\dot{V}O_{2max}$ prediction equation was recalibrated by

regressing measured $\dot{V}O_{2max}$ in the BLSA on the BLSA covariates representing those used in each prediction equation. With recalibration, the regression coefficients for each covariate in relation to measured $\dot{V}O_{2max}$ derive fully from the BLSA, as opposed to applying regression weights calculated in a different population to BLSA covariates. Recalibration has been used in other settings to evaluate accuracy of prediction equations when transported from the source to other populations [58]. Residuals vs. Fitted, Normal Q-Q, Scale-Location, and Residuals vs. Leverage plots were used to assess model diagnostics of the recalibrated $\dot{V}O_{2max}$ prediction equations [59]. After evaluation of all recalibrated equations, their predicted $\dot{V}O_{2max}$ values were output. Performance metrics for the recalibrated equations included the same metrics as described above for evaluation of the original equations, as well as the 10-fold cross-validation (CV) RMSE and R^2 values.

To further evaluate the validity of predicted $\dot{V}O_{2max}$ values, sequentially adjusted Cox proportional hazards regression models were used to estimate the associations between quartiles of $\dot{V}O_{2max}$ (measured $\dot{V}O_{2max}$, predicted $\dot{V}O_{2max}$, and the recalibration-predicted $\dot{V}O_{2max}$) and all-cause mortality. Model 1 was unadjusted, Model 2 adjusted for age and sex, and Model 3 adjusted for Model 2 covariates in addition to race and ethnicity, and education. Linear trends across quartiles (*P*-value for Trend) were tested by specifying the quartile indicator in the model as a continuous variable. Associations between a one standard deviation increase in each $\dot{V}O_{2max}$ variable and all-cause mortality were also assessed using the same modeling approach, and the *P*-value for the centered and scaled $\dot{V}O_{2max}$ variable in the model are presented. The concordance statistic (*C*-Statistic), the proportion of pairs of participants where the model correctly predicts which participant will experience a mortality event first, is also presented. Tests of the proportional hazards assumption were conducted using the cox.zph function of the

survival package [60] in R through the testing of the correlation of each covariate's (and the whole model's) scaled Schoenfeld residuals with time to ensure independence between the residuals and time; no violations were noted.

All analyses were conducted in R version 3.6.3 (R Foundation for Statistical Computing, Vienna, Austria).

2.4. Results

Sample Characteristics

The 565 women and 515 men with measured $\dot{V}O_{2max}$ included in this study had a mean age, BMI, and $\dot{V}O_{2max}$ of 69.0 ± 10.4 years, 27.0 ± 4.4 kg/m^2, and 21.6 ± 5.9 mL•kg^{-1}•min^{-1}, respectively (see Table 2.1.). Two-thirds of study participants were non-Hispanic White, one-fourth were non-Hispanic Black, 4.6% were non-Hispanic Asian, 3.2% were Hispanic, and the remaining 0.7% were from other non-Hispanic race/ethnicity groups or could not be classified. The majority of the sample (61.9%) had a post-college education. Current smoking prevalence was 1.8% and mean systolic and diastolic blood pressures were 114.1 ± 14.1 and 66.7 ± 8.8 mmHg, respectively. Age, BMI, current smoking, and systolic blood pressure were inversely related with incremental quartiles of measured $\dot{V}O_{2max}$, whereas education, self-reported exercise, self-rated health status, and diastolic blood pressure were positively related with $\dot{V}O_{2max}$ (see Table 2.1.).

$\dot{V}O_{2max}$ Prediction Equations

When each prediction equation was used to estimate $\dot{V}O_{2max}$ in the BLSA sample, the lowest and highest RMSE values (in units of mL•kg^{-1}•min^{-1}) of the $\dot{V}O_{2max}$ prediction equations were 4.2 (Bradshaw et al.'s [37] equations) and 20.4 (Jang et al. [43]), respectively (see Table 2.3.).

The absolute value of the bias (unitless) ranged from 0.1 (Matthews et al. [40]) to 19.3 (Jang et al. [43]). Bradshaw et al. [37] had the lowest MAPE value (15.4%) and Jang et al. [43] had the highest MAPE value (97.7%).

After recalibration of the equations to the BLSA data, every equation improved on all performance metrics (see Table 2.4.). The recalibrated formulas' cross-validated RMSE values ranged from 3.9 (Bradshaw et al. [37]) to 4.2 (Myers et al. [22]) and, as expected, all bias values were 0. MAPE values were similar across the recalibrated prediction equations, ranging from 14.4% (Bradshaw et al. [37]) to 15.7% (Myers et al. [22]). The R^2 for the recalibrated equations ranged from 49% (Myers et al. [22]) to 58% (Bradshaw et al. [37]), which compares favorably to an age and sex adjusted model R^2 of 36%. Additional recalibrated performance metrics including sex-stratified performance metrics are reported in Tables 2.3 and 2.4.

$\dot{V}O_{2max}$ Associations with Mortality

When assessing the associations between quartiles of measured $\dot{V}O_{2max}$ and all-cause mortality, a steep inverse gradient in mortality risk across incremental $\dot{V}O_{2max}$ quartiles was evident in both unadjusted and adjusted models. Adjusting for Model 3 covariates, the hazard ratios (HRs) and (95% CI) 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 $\dot{V}O_{2max}$, respectively, $P_{trend} < 0.001$ (see Table 2.5.). To further investigate the robustness of measured $\dot{V}O_{2max}$ to adjustments beyond the Model 3 covariates, we additionally adjusted for the following variables: BMI, smoking history, self-rated health, diagnosed diabetes, glucose intolerance, or high blood sugar, history of heart attack or myocardial infarction, history of heart failure or CHF, history of stroke mini stroke or slight stroke, and current hypertension. The HRs from this model slightly strengthened in magnitude,

remained statistically significant, and maintained their trend across quartiles (HRs for Q2-Q4 relative to Q1: 0.56 (0.36-0.88), 0.30 (0.16-0.59), and 0.31 (0.13-0.75); $P_{trend} < 0.001$).

Results from the Cox proportional hazards regression models estimating the associations between predicted $\dot{V}O_{2max}$ (each equation separately), and all-cause mortality are shown in Table 4a. For most equations, predicted $\dot{V}O_{2max}$ was associated with mortality in a pattern and strength similar to that of measured $\dot{V}O_{2max}$ in the crude model (Model 1), but adjustment for basic covariates in Models 2 and 3 attenuated the HRs, widened the confidence intervals to statistical insignificance, and eliminated all linear trends (see Table 2.5.).

After recalibration, unadjusted HRs for Q2 – Q4 relative to Q1 of predicted $\dot{V}O_{2max}$ exhibited patterns and magnitudes of association that more closely reflected those for measured $\dot{V}O_{2max}$. For example, no published equation had an HR of 0.10 (the Q4 HR of measured $\dot{V}O_{2max}$ relative to Q1) in Q4 relative to Q1, but recalibrated Q4 HRs were ≤0.10 for most equations. However, after adjustment for covariates in Models 2 and 3, the HRs were attenuated again, confidence intervals widened to statistical insignificance, and linear trends were not statistically significant (see Table 2.6.).

2.5. Discussion

In the present study, we sought to provide validation, recalibration, and predictive accuracy metrics of published $\dot{V}O_{2max}$ prediction equations with the aim of enabling large scale epidemiologic cohorts with older, ambulatory, community-dwelling adults to accurately estimate $\dot{V}O_{2max}$. Performance metrics of several of the extant equations yielded reasonable results relative to measured $\dot{V}O_{2max}$, e.g. the Bradshaw [37] equation had an RMSE value of 4.2 $mL \bullet kg^{-1} \bullet min^{-1}$. This means that, on average, this equation's errors were within ~1.2 metabolic

equivalents (METs) assuming the standard conversion of 3.5 $mL \bullet kg^{-1} \bullet min^{-1}$ to 1 MET. The Matthews [40] equation had absolute bias value of 0.1, meaning that, on average, this model's predictions were within 0.03 METs. The recalibration of these equations using the BLSA measured $\dot{V}O_{2max}$ and covariate data improved every performance metric, although such recalibration would not be possible in epidemiologic cohorts unless $\dot{V}O_{2max}$ *and* the covariates used in the derivation cohort were directly measured.

Cox proportional hazards modeling showed measured $\dot{V}O_{2max}$ is an *extremely* powerful predictor of all-cause mortality in BLSA participants in both the unadjusted and adjusted models. Compared to participants in the lowest quartile of measured $\dot{V}O_{2max}$, those in the highest quartile had a 3-fold reduction in the risk of all-cause mortality, after adjusting for age, sex, race and ethnicity, and education. These HRs are similar to, though slightly stronger than, those reported in other studies of $\dot{V}O_{2max}$ and all-cause mortality for those with the highest levels of CRF relative to those with the lowest CRF [61–63].

Among the previously published $\dot{V}O_{2max}$ prediction models, there was no discernable pattern of covariate types (i.e. demographics, body mass, self-reported PA) that contributed to the performance of the model more than others (e.g. the Bradshaw equation [37], one of the best performing models, has the same covariates as the Jurca equations [38], which did not perform as well in relation to measured $\dot{V}O_{2max}$ in the BLSA). Several of the published equations yielded HRs similar in pattern and magnitude to those of measured $\dot{V}O_{2max}$ before adjustment, but these associations were not robust to even minimal adjustments. After adjustment for only age and sex, the ability of the equations to predict mortality was substantially weakened, suggesting that much of the association observed in the unadjusted models was due to these two variables alone. In

regression models using the recalibrated equations, the patterns of association were more similar to those estimated using measured $\dot{V}O_{2max}$ in unadjusted models, (i.e. closer to the pattern of the unadjusted HRs of measured $\dot{V}O_{2max}$): Q1 – Q4: 1.00 (ref.), 0.43 (0.29-0.63), 0.16 (0.09-0.29), and 0.10 (0.05-0.20).

Despite the pattern of the recalibrated equations' HRs in unadjusted models, these associations were still not robust to adjustment. These findings strongly suggest that while the equations may be valid and useful, to varying degrees, for individual exercise prescriptions in the field, their ability to predict mortality is severely compromised after adjustment for basic demographic and anthropometric covariates, some of which are components of the prediction equations themselves. $\dot{V}O_{2max}$, and CRF in general, are complex constructs reflecting an integration of multifaceted organ systems and metabolic processes [64]. Without direct measures of the physiologic variability across individuals inherent in measured CRF, even well-performing prediction equations based on basic demographic and health characteristics do not predict mortality independent of sex and age. To a large extent, this is because demographic and behavioral characteristics do not adequately capture the integrated physiological signal reflected in measured $\dot{V}O_{2max}$.

There are some limitations to the present study. First, not all covariates from the published equations had exact counterpart covariates in the BLSA. While these discrepancies could potentially limit the performance metrics of the equations when applied in the BLSA, this limitation would be eliminated once the equations were recalibrated to the BLSA measured $\dot{V}O_{2max}$. Next, the majority of the sample (61.9%) had a post-college education, which is higher than the general population. One substantial strength of the present study is the prospective

follow-up, enabling the evaluation of the accuracy of predicted $\dot{V}O_{2max}$ with respect to measured $\dot{V}O_{2max}$ and their associations with mortality. BLSA enrolled a large group of racially and ethnically diverse older adults, included laboratory-based measurements $\dot{V}O_{2max}$, followed participants for mortality outcomes after $\dot{V}O_{2max}$ assessment, and collected data that enabled adjustment for confounders. The conclusions drawn from these data and analyses are robust across our approaches—the performance metrics and the HRs contribute to a consistent and unified narrative regarding the importance of accurately assessing $\dot{V}O_{2max}$ in older adults and the relevance of this aging biomarker [46] to clinical outcomes such as all-cause mortality.

In conclusion, measured $\dot{V}O_{2max}$ is an extremely strong predictor of all-cause mortality in aging men and women. Those in the highest sex-specific quartile of measured $\dot{V}O_{2max}$ experienced a 66% lower risk of death relative to those in the lowest quartile of $\dot{V}O_{2max}$ after adjustment for age, race, sex, and education. Several published $\dot{V}O_{2max}$ prediction models yielded: (1) reasonable performance metrics relative to measured $\dot{V}O_{2max}$, especially when recalibrated, (2) all-cause mortality hazard ratios similar to those of measured $\dot{V}O_{2max}$, especially when recalibrated, yet (3) were not robust to adjustment for basic demographic covariates. These findings make an important contribution to research on the development of an inexpensive surrogate for direct measurement of CRF that could be broadly used to guide healthy aging in the older population. Future studies should investigate whether modern analytic methods such as machine learning can improve prediction of $\dot{V}O_{2max}$ in community-dwelling older adults so that this critical "vital sign" can be more broadly studied as a modifiable target for promoting functional resiliency and healthy aging.

Table 2.1. Characteristics of BLSA participants overall and according to quartiles of measured $\dot{V}O_{2max}$ (n = 1,080)

		Measured $\dot{V}O_{2max}$				
Characteristic†	**Total (n = 1,080)**	**Quartile 1†† (n = 270)**	**Quartile 2†† (n = 277)**	**Quartile 3†† (n = 265)**	**Quartile 4†† (n = 268)**	**P-value***
Age, Mean (SD)	69.0 (10.4)	75.5 (8.8)	72.1 (9.7)	67.3 (8.9)	60.9 (8.2)	**< 0.01**
Race, n (%)						**< 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, n (%)						**< 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)	
BMI (kg/m^2), Mean (SD)	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, n (%)	152 (14.1)	78 (28.9)	39 (14.1)	22 (8.3)	13 (4.9)	**< 0.01**
Minutes of Exercise, n (%)						**< 0.01**
0-29	465 (43.1)	171 (63.3)	127 (45.8)	93 (35.1)	74 (27.6)	
30 - 74	169 (15.6)	36 (13.3)	48 (17.3)	33 (12.5)	52 (19.4)	
75 - 149	165 (15.3)	25 (9.3)	42 (15.2)	52 (19.6)	46 (17.2)	
150+	272 (25.2)	36 (13.3)	59 (21.3)	84 (31.7)	93 (34.7)	
Self Rated Health, n (%)						**< 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)	
Systolic BP (mmHg), Mean (SD)	114.1 (14.1)	117.3 (14.8)	116 (13.3)	113 (13.7)	110.2 (13.3)	**< 0.01**
Diastolic BP (mmHg), Mean (SD)	66.7 (8.8)	65 (8.4)	66.3 (9.3)	66.9 (8.6)	68.5 (8.5)	**< 0.01**

Table 2.1. (cont'd) Characteristics of BLSA participants overall and according to quartiles of measured $\dot{V}O_{2max}$ (n = 1,080)

		Measured $\dot{V}O_{2max}$				
Characteristic†	**Total (n = 1,080)**	**Quartile 1†† (n = 270)**	**Quartile 2†† (n = 277)**	**Quartile 3†† (n = 265)**	**Quartile 4†† (n = 268)**	**P-value***
Smoking Status, n (%)						**< 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)	
Maximal Exercise Test						
$\dot{V}O_{2max}$ (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, Mean (SD)	3.3 (68.1)	1.2 (0.1)	1.2 (0.1)	1.2 (0.1)	9.5 (136.8)	0.18
Borg Score, Mean (SD)	16.5 (1.7)	16.1 (1.7)	16.2 (1.7)	16.7 (1.7)	17 (1.6)	**< 0.01**
% of Max. Predicted HR, Mean (SD)**	98.8 (50.2)	89.6 (13.3)	97.5 (9.3)	100.3 (8.5)	107.8 (98.3)	**< 0.01**

†Percentages may not sum to 100% due to missing data.

*P-value for continuous variables from One-way ANOVA and Chi-Sq. goodness of fit test for categorical variables across VO_{2max} quartiles.

**Maximum predicted heart rate: 220 - age

Bold indicates significance at the $P < 0.05$ level.

††Sex-specific quartile definitions were as follows:

Q1: Men : < 19.9; n = 129 & Women: < 16.5; n = 141

Q2: Men : ≥ 19.9 & ≤ 23.7; n = 131 & Women: ≥ 16.5 & ≤ 19.9; n = 146

Q3: Men : > 23.7 & ≤ 27.4; n = 128 & Women: > 19.9 & ≤ 23.7; n = 137

Q4: Men : > 27.4; n = 127 & Women: > 23.7; n = 141

Table 2.2. Extant $\dot{V}O_{2max}$ Prediction Equations, Adaptations for the BLSA, and the Recalibrated Formulas

Study	Prediction Equation from Study	Variable Definitions	Variable Adaptations
Jurca et al; NASA	Original: 18.07 + 2.77(Sex) - 0.10(Age) - 0.17(BMI) - 0.03(Resting HR) + 0.32(SRPA1) + 1.06(SRPA2) + 1.76(SRPA3) + 3.03(SRPA4) Recalibrated: 58.64 + 4.62(Sex) - 0.31(Age) - 0.47(BMI) - 0.04(Resting HR) - 2.39(exercise0) - 1.73(exercise1) - 1.15(exercise2)	Sex coded as Men = 1 and Women = 0. SRPA0 = Little activity other than walking for pleasure (Ref.). SRPA1 = Some regular participation in modest physical activities involving sports, recreational activities. SRPA2 = Aerobic exercise such as run/walk for 20 to 60 minutes per week. SRPA3 = Aerobic exercise such as run/walk for 1 to 3 hours per week. SRPA4 = Aerobic exercise such as run/walk for > 3 hours per week.	BLSA participants self-reported the weekly minutes of weight/circuit training, moderate-to-high intensity exercise, and/or brisk walking and were categorized into 0 - 29 minutes (coded as 0), 30 – 74 minutes (1), 75 – 149 minutes (2), and 150+ minutes (3). Dummy variables were created and exercise0 = SRPA1, exercise1 = SRPA2, exercise3 = SRPA3, and exercise3 = SRPA4.
Jurca et al; ACLS	Original: 18.81 + 2.49(Sex) - 0.08(Age) - 0.17(BMI) - 0.05(Resting HR) + 0.81(SRPA1) + 1.17(SRPA2) + 2.16(SRPA3) + 3.05(SRPA4) Recalibrated: 58.64 + 4.62(Sex) - 0.31(Age) - 0.47(BMI) - 0.04(Resting HR) - 2.39(exercise0) - 1.73(exercise1) - 1.15(exercise2)	Sex coded as Men = 1 and Women = 0. SRPA0 = No activity (Ref.). SRPA1 = Participated in sporting or leisure-time physical activity other than walking, jogging, or running. SRPA2 = Walk, jog, or run up to 10 miles per week. SRPA3 = Walk, jog, or run from 10 to 20 miles per week. SRPA4 = Walk, jog, or run > 20 miles per week	BLSA participants self-reported the weekly minutes of weight/circuit training, moderate-to-high intensity exercise, and/or brisk walking and were categorized into 0 - 29 minutes (coded as 0), 30 – 74 minutes (1), 75 – 149 minutes (2), and 150+ minutes (3). Dummy variables were created and exercise0 = SRPA1, exercise1 = SRPA2, exercise3 = SRPA3, and exercise3 = SRPA4.
Jurca et al; ADNFS	Original: 21.41 + 2.78(Sex) - 0.11(Age) - 0.17(BMI) - 0.05(Resting HR) + 0.35(SRPA1) + 0.29(SRPA2) + 0.64(SRPA3) + 1.21(SRPA4) Recalibrated: 58.64 + 4.62(Sex) - 0.31(Age) - 0.47(BMI) - 0.04(Resting HR) - 2.39(exercise0) - 1.73(exercise1) - 1.15(exercise2)	Sex coded as Men = 1 and Women = 0. SRPA0 = From 0 to 4 occasions of at least moderate activity in past 4 weeks (Ref.). SRPA1 = From 5 to 11 occasions of at least moderate activity in past 4 weeks. SRPA2 = G.E. 12 Occasions of moderate activity in past 4 weeks. SRPA3 = G.E. 12 Occasions of a mix of moderate and vigorous activities in past 4 weeks. SRPA4 = G.E. 12 occasions of vigorous activity in past 4 weeks.	BLSA participants self-reported the weekly minutes of weight/circuit training, moderate-to-high intensity exercise, and/or brisk walking and were categorized into 0 - 29 minutes (coded as 0), 30 – 74 minutes (1), 75 – 149 minutes (2), and 150+ minutes (3). Dummy variables were created and exercise0 = SRPA1, exercise1 = SRPA2, exercise3 = SRPA3, and exercise3 = SRPA4.
Bradshaw et al	Original: 48.073 + 6.178(Sex) - 0.246(Age) - 0.619(BMI) + 0.712(PFA) + 0.671(PA-R) Recalibrated: 46.61 + 4.82(Sex) - 0.31(Age) - 0.46(BMI) + 1.58(sfhealth) + 0.57(exercise)	Sex coded as Men = 1 and Women = 0. PFA = two questions that ascertain how fast participants feel they can cover a 1- and 3-mile distance at a comfortable pace. The participant's sum total of both 13-point questions is counted as the PFA score (range: 2–26). PA-R = individuals rate their activity level over the previous 6-months using a 10-point scale.	Reverse coded SF-12 self-rated health score was used in lieu of PFA (5 = "Excellent", 4 = "Very Good", 3 = "Good", 2 = "Fair", 1 = "Poor". Used the aforementioned exercise variable in lieu of PA-R.

Table 2.2. (cont'd) Extant $\dot{V}O_{2max}$ Prediction Equations, Adaptations for the BLSA, and the Recalibrated Formulas

Study	Prediction Equation from Study	Variable Definitions	Variable Adaptations
Jackson et al	Original: 56.363 + 1.921(PA-R) - 0.381(Age) - 0.754(BMI) + 10.987(Sex) Recalibrated: 53.89 + 0.74(exercise) - 0.31(Age) - 0.5(BMI) + 4.7(Sex)	Sex coded as Men = 1 and Women = 0. SRPA0 = Little activity other than walking for pleasure (Ref.). PA-R from Jurca et al. NASA equation.	Used the aforementioned exercise variable in lieu of PA-R.
Matthews et al	Original: 34.142 + 11.403(Sex) + 0.133(Age) - (0.005(Age*Age)) + 1.463(PAS) + 9.170*(ht meters) - 0.254(Body mass in kg) Recalibrated: 32.58 + 5.15(Sex) - 0.32(Age) + 0.74(exercise) + 12.97(ht meters) - 0.18(wtkg)	Sex coded as Men = 1 and Women = 0. PAS = Physical Activity Status (0 - 7); This instrument has subjects rate their last month of physical activity participation on a 0-7 scale. Responses of 0 and 1 represented no regular physical activity, whereas a response of 2 or 3 represented moderate intensity activities, and responses of 4 to 7 represented regular vigorous physical activity participation of increasing exercise time.	Used the aforementioned exercise variable in lieu of PAS.
Sloan et al	Original: Male w/ HR: 52.23 − 0.20(Age) − 0.35(BMI) − 0.06(HRrest/min) + 2.05(Physical activity score) Original: Female w/ HR: 47.79 − 0.21(Age) − 0.35(BMI) − 0.06(HRrest/min) + 2.08(Physical activity score) Original: Male w/o HR: 49.9 − 0.21(Age) − 0.36(BMI) + 2.12(Physical activity score) Original: Female w/o HR: 43.27 − 0.22(Age) − 0.37(BMI) + 2.17(Physical activity score) Recalibrated w/ HR: 56.06 + 4.62(Sex) - 0.31(Age) - 0.47(BMI) - 0.04(Resting HR) + 0.78(exercise) Recalibrated w/o HR: 53.89 + 4.7(Sex) - 0.31(Age) - 0.5(BMI) + 0.74(exercise)	Physical activity score: In accordance with the procedure outlined by Jurca et al, participants were asked to select one of the five levels of self-reported physical activity that best described their usual activity pattern: (a) level 0 – inactive or little activity other than usual daily activities; (b) level 1 – regular (> 5 days/wk) participation in physical activities requiring low levels of exertion that result in slight increases in breathing and heart rate for at least 10 mins at a time; (c) level 2 – participation in aerobic exercises such as brisk walking, jogging or running, cycling, swimming or vigorous sports at a comfortable pace, or other activities requiring similar levels of exertion for 20–60 mins/wk; (d) level 3 – participation in aerobic exercises such as brisk walking, jogging or running at a comfortable pace, or other activities requiring similar levels of exertion for 1–3 hrs/wk; and (e) level 4 – participation in aerobic exercises such as brisk walking, jogging or running at a comfortable pace, or other activities requiring similar levels of exertion for > 3 hrs/wk.	Used the aforementioned exercise variable in lieu of physical activity score.

Table 2.2. (cont'd) Extant $\dot{V}O_{2max}$ Prediction Equations, Adaptations for the BLSA, and the Recalibrated Formulas

Study	Prediction Equation from Study	Variable Definitions	Variable Adaptations
de Souza e Silva et al	Original: 44.74 – 10.9(Sex) – 0.35(Age) – 0.15(Weight pounds) + 0.68(Height inches); Treadmill constant into the intercept. Recalibrated: 44.3 - 5.31(Sex) - 0.33(Age) - 0.09(Weight pounds) + 0.35(Height inches)	Sex coded as Men = 1 and Women = 2.	-
Baynard et al	Original: 77.96 − 10.35(Sex) − 0.32(Age) − 0.92(BMI) Recalibrated: 61.45 - 4.82(Sex) - 0.33(Age) - 0.54(BMI)	Sex coded as Men = 0 and Women = 1.	-
Jang et al	Original: 50.543 - 0.069(Age) + 13.525(Sex) - 0.403(BMI) - 1.530(Smoking) Recalibrated: 56.58 - 0.33(Age) + 4.85(Sex) - 0.53(BMI) - 1.49(Smoking)	Smoking: 0 = never or former, 1 = current.	-
Myers et al	Original: 79.9 - 0.39(Age) - 13.7(Sex) - 0.127(wt lbs) Recalibrated: 62.71 - 0.35(Age) - 6.85(Sex) - 0.08(wt lbs)	Sex coded as Men = 0 and Women = 1.	-

Table 2.3. Performance Metrics for Previously Published VO_{2max} Prediction Equations Compared to Measured $\dot{V}O_{2max}$ in the BLSA

Formula	RMSE	Bias	MAPE	LOA	Correlation w/ Measured $\dot{V}O_{2max}$	R^2
Jurca et al; NASA	15.8	15.1	67.8	(5.7, 24.5)	**0.68**	0.46
Male	16.8	15.8	63.5	(4.9, 26.7)	**0.60**	0.36
Female	15.0	14.5	71.2	(6.7, 22.3)	**0.67**	0.45
Jurca et al; ACLS	15.1	14.3	63.5	(4.6, 23.9)	**0.66**	0.44
Male	16.1	15.1	60.2	(4.0, 26.2)	**0.58**	0.34
Female	14.2	13.6	66.2	(5.5, 21.6)	**0.64**	0.41
Jurca et al; ADNFS	15.4	14.7	65.4	(5.1, 24.2)	**0.69**	0.48
Male	16.4	15.4	61.5	(4.4, 26.4)	**0.66**	0.44
Female	14.6	14.1	68.6	(6.0, 22.1)	**0.68**	0.46
Bradshaw et al	4.2	1.0	15.4	(-7.1, 9.0)	**0.72**	0.52
Male	4.6	0.2	15.7	(-8.8, 9.2)	**0.67**	0.45
Female	3.8	1.7	15.2	(-5.1, 8.5)	**0.74**	0.55
Jackson et al	7.2	4.7	29.4	(-6.0, 15.4)	**0.69**	0.48
Male	5.1	1.5	17.1	(-8.0, 11.1)	**0.64**	0.41
Female	8.7	7.6	40.5	(-0.7, 15.9)	**0.73**	0.53
Matthews et al	5.3	-0.1	20.7	(-10.4, 10.2)	**0.72**	0.52
Male	5.6	-2.3	20.1	(-12.2, 7.6)	**0.67**	0.45
Female	5.0	1.9	21.2	(-7.2, 10.9)	**0.72**	0.52
Sloan et al; HR	5.1	-2.3	21.0	(-11.2, 6.6)	**0.67**	0.45
Male	6.0	-2.9	23.2	(-13.1, 7.4)	**0.58**	0.34
Female	4.2	-1.8	19.2	(-9.4, 5.7)	**0.66**	0.44

Table 2.3. (cont'd) Performance Metrics for Previously Published VO_{2max} Prediction Equations Compared to Measured $\dot{V}O_{2max}$ in the BLSA

Formula	RMSE	Bias	MAPE	LOA	Correlation w/ Measured $\dot{V}O_{2max}$	R^2
Sloan et al; No HR	5.3	-2.3	21.3	(-11.6, 7.0)	**0.65**	0.42
Male	6.4	-4.0	26.0	(-13.9, 5.8)	**0.57**	0.32
Female	3.9	-0.8	17.0	(-8.3, 6.8)	**0.67**	0.45
de Souza e Silva et al	5.4	-1.6	21.4	(-11.8, 8.5)	**0.67**	0.45
Male	6.6	-4.4	26.3	(-14.1, 5.2)	**0.61**	0.37
Female	4.1	0.9	16.9	(-6.9, 8.7)	**0.71**	0.50
Baynard et al	6.3	-3.6	25.7	(-13.8, 6.6)	**0.66**	0.44
Male	8.0	-6.4	33.7	(-15.9, 3.2)	**0.61**	0.37
Female	4.2	-1.1	18.4	(-9.0, 6.9)	**0.70**	0.49
Jang et al	20.4	-19.3	97.7	(-32.5, -6.1)	**0.44**	0.19
Male	24.8	-24.2	113.3	(-35.1, -13.2)	**0.45**	0.20
Female	15.4	-14.8	83.4	(-23.0, -6.7)	**0.61**	0.37
Myers et al	5.7	-2.3	22.2	(-12.4, 7.8)	**0.66**	0.44
Male	7.0	-5.0	28.4	(-14.6, 4.5)	**0.61**	0.37
Female	4.0	0.2	16.6	(-7.7, 8.0)	**0.68**	0.46

Bold indicates significance at the $P < 0.01$ level.

Table 2.4. Performance Metrics for Recalibrated $\dot{V}O_{2max}$ Prediction Equations Compared to Measured $\dot{V}O_{2max}$ in the BLSA

Recalibrated Formula	RMSE	MAPE	LOA	Correlation w/ Measured $\dot{V}O_{2max}$	R^2*
Jurca et al	4.1	15.4	(-8.1, 8.1)	**0.73**	0.53
Male	4.8	16.2	(-9.4, 9.4)	**0.68**	0.46
Female	3.5	14.8	(-6.8, 6.8)	**0.73**	0.53
Bradshaw et al	3.9	14.4	(-7.6, 7.6)	**0.76**	0.58
Male	4.3	14.6	(-8.4, 8.4)	**0.72**	0.52
Female	3.4	14.3	(-6.7, 6.7)	**0.74**	0.55
Jackson et al	4.0	15.0	(-7.9, 7.9)	**0.73**	0.53
Male	4.5	15.3	(-8.9, 8.9)	**0.67**	0.45
Female	3.5	14.7	(-6.8, 6.8)	**0.73**	0.53
Matthews et al	4.0	15.0	(-7.9, 7.9)	**0.73**	0.53
Male	4.5	15.3	(-8.9, 8.9)	**0.67**	0.45
Female	3.5	14.7	(-6.9, 6.9)	**0.73**	0.53
Sloan et al; HR	4.1	15.4	(-8.1, 8.1)	**0.73**	0.53
Male	4.8	16.1	(-9.4, 9.4)	**0.68**	0.46
Female	3.5	14.8	(-6.8, 6.8)	**0.72**	0.52

Table 2.4. (cont'd) Performance Metrics for Recalibrated $\dot{V}O_{2max}$ Prediction Equations Compared to Measured $\dot{V}O_{2max}$ in the BLSA

Recalibrated Formula	RMSE	MAPE	LOA	Correlation w/ Measured $\dot{V}O_{2max}$	R^2*
Sloan et al; No HR	4.0	15.0	(-7.9, 7.9)	**0.73**	0.53
Male	4.5	15.3	(-8.9, 8.9)	**0.67**	0.45
Female	3.5	14.7	(-6.8, 6.8)	**0.73**	0.53
de Souza e Silva et al	4.1	15.4	(-8.1, 8.1)	**0.72**	0.52
Male	4.6	15.7	(-9.1, 9.1)	**0.66**	0.44
Female	3.6	15.1	(-7.1, 7.1)	**0.71**	0.50
Baynard et al	4.1	15.4	(-8.1, 8.1)	**0.72**	0.52
Male	4.6	15.8	(-9.1, 9.1)	**0.66**	0.44
Female	3.6	15.1	(-7.0, 7.0)	**0.71**	0.50
Jang et al	4.1	15.4	(-8.1, 8.1)	**0.72**	0.52
Male	4.6	15.7	(-9.1, 9.1)	**0.66**	0.44
Female	3.6	15.1	(-7.0, 7.0)	**0.71**	0.50
Myers et al	4.2	15.7	(-8.2, 8.2)	**0.70**	0.49
Male	4.6	15.9	(-9.1, 9.1)	**0.65**	0.42
Female	3.7	15.5	(-7.3, 7.3)	**0.68**	0.46

Bold indicates significance at the $P < 0.01$ level.

Table 2.5. Hazard Ratios (HR) of All-Cause Mortality by Measured and Predicted $\dot{V}O_{2max}$ in the Selected BLSA Sample (n = 1,080)

Author, Model	Sex- Specific Quartiles of $\dot{V}O_{2max}$ Q1	Q2	Q3	Q4	*P*-Trend	HR for 1 SD Increase	*P*-value	*C*-Statistic
Measured								
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 (0.02)
2	1.00 (ref.)	0.55 (0.37-0.81)	0.30 (0.17-0.54)	0.34 (0.16-0.75)	**< 0.01**	0.51 (0.39-0.66)	**< 0.01**	0.79 (0.02)
3	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 (0.02)
Baynard et al								
1	1.00 (ref.)	0.67 (0.45-0.99)	0.42 (0.27-0.66)	0.15 (0.07-0.29)	**< 0.01**	0.89 (0.75-1.05)	0.16	0.66 (0.02)
2	1.00 (ref.)	0.72 (0.49-1.07)	0.82 (0.51-1.33)	0.58 (0.28-1.20)	0.12	0.91 (0.66-1.24)	0.55	0.78 (0.02)
3	1.00 (ref.)	0.71 (0.48-1.06)	0.85 (0.53-1.37)	0.63 (0.30-1.32)	0.19	0.94 (0.68-1.30)	0.72	0.78 (0.02)
Bradshaw et al								
1	1.00 (ref.)	0.66 (0.44-0.98)	0.33 (0.20-0.54)	0.24 (0.14-0.42)	**< 0.01**	0.79 (0.67-0.93)	**< 0.01**	0.64 (0.02)
2	1.00 (ref.)	0.74 (0.50-1.10)	0.76 (0.45-1.28)	1.09 (0.58-2.02)	0.62	0.90 (0.68-1.19)	0.47	0.78 (0.02)
3	1.00 (ref.)	0.73 (0.49-1.09)	0.78 (0.47-1.32)	1.27 (0.67-2.41)	0.83	0.93 (0.70-1.25)	0.64	0.78 (0.02)
de Souza e Silva et al								
1	1.00 (ref.)	0.59 (0.40-0.88)	0.43 (0.28-0.66)	0.14 (0.07-0.28)	**< 0.01**	0.86 (0.73-1.01)	0.07	0.66 (0.02)
2	1.00 (ref.)	0.66 (0.44-0.98)	0.87 (0.54-1.39)	0.57 (0.28-1.18)	0.13	0.89 (0.65-1.21)	0.45	0.78 (0.02)
3	1.00 (ref.)	0.65 (0.44-0.97)	0.90 (0.56-1.45)	0.64 (0.31-1.34)	0.21	0.93 (0.67-1.28)	0.64	0.78 (0.02)
Jackson et al								
1	1.00 (ref.)	0.47 (0.31-0.72)	0.32 (0.20-0.51)	0.19 (0.11-0.34)	**< 0.01**	0.81 (0.69-0.96)	**0.01**	0.66 (0.02)
2	1.00 (ref.)	0.69 (0.46-1.05)	0.82 (0.50-1.35)	0.98 (0.51-1.89)	0.52	0.92 (0.68-1.25)	0.59	0.77 (0.02)
3	1.00 (ref.)	0.69 (0.46-1.06)	0.83 (0.51-1.37)	1.12 (0.57-2.21)	0.70	0.94 (0.69-1.29)	0.71	0.78 (0.02)
Jang et al								
1	1.00 (ref.)	1.08 (0.70-1.67)	0.88 (0.56-1.38)	0.54 (0.33-0.91)	**0.02**	1.33 (1.12-1.57)	**< 0.01**	0.57 (0.02)
2	1.00 (ref.)	1.03 (0.67-1.59)	0.74 (0.47-1.17)	1.00 (0.59-1.69)	0.49	0.80 (0.38-1.67)	0.55	0.78 (0.02)
3	1.00 (ref.)	1.05 (0.68-1.63)	0.76 (0.48-1.20)	1.05 (0.62-1.77)	0.60	0.87 (0.41-1.84)	0.71	0.78 (0.02)

Table 2.5. (cont'd) Hazard Ratios (HR) of All-Cause Mortality by Measured and Predicted $\dot{V}O_{2max}$ in the Selected BLSA Sample (n = 1,080)

Author, Model	Sex- Specific Quartiles of $\dot{V}O_{2max}$ Q1	Q2	Q3	Q4	*P*-Trend	HR for 1 SD Increase	*P*-value	*C*-Statistic
Jurca et al; ACLS								
1	1.00 (ref.)	0.55 (0.31-0.99)	0.33 (0.17-0.64)	0.28 (0.13-0.58)	**< 0.01**	0.74 (0.59-0.94)	**0.01**	0.63 (0.04)
2	1.00 (ref.)	0.65 (0.36-1.17)	0.65 (0.33-1.26)	0.89 (0.40-1.96)	0.39	0.87 (0.60-1.26)	0.47	0.80 (0.03)
3	1.00 (ref.)	0.62 (0.34-1.12)	0.60 (0.31-1.18)	0.90 (0.40-2.01)	0.34	0.87 (0.60-1.27)	0.47	0.80 (0.03)
Jurca et al; ADNFS								
1	1.00 (ref.)	0.76 (0.45-1.29)	0.24 (0.11-0.50)	0.14 (0.06-0.37)	**< 0.01**	0.67 (0.52-0.85)	**< 0.01**	0.67 (0.03)
2	1.00 (ref.)	1.18 (0.69-2.02)	0.74 (0.33-1.68)	1.31 (0.42-4.09)	0.95	0.89 (0.54-1.46)	0.64	0.80 (0.03)
3	1.00 (ref.)	1.14 (0.66-1.97)	0.72 (0.32-1.64)	1.33 (0.41-4.36)	0.89	0.88 (0.53-1.47)	0.63	0.80 (0.03)
Jurca et al; NASA								
1	1.00 (ref.)	0.44 (0.24-0.79)	0.31 (0.16-0.59)	0.22 (0.10-0.48)	**< 0.01**	0.67 (0.52-0.85)	**< 0.01**	0.64 (0.04)
2	1.00 (ref.)	0.69 (0.38-1.26)	0.69 (0.35-1.33)	1.12 (0.47-2.66)	0.58	0.84 (0.57-1.24)	0.39	0.80 (0.03)
3	1.00 (ref.)	0.67 (0.36-1.22)	0.65 (0.33-1.27)	1.15 (0.48-2.78)	0.54	0.84 (0.57-1.25)	0.40	0.81 (0.02)
Matthews et al								
1	1.00 (ref.)	0.25 (0.16-0.40)	0.19 (0.12-0.31)	0.09 (0.04-0.18)	**< 0.01**	0.60 (0.50-0.71)	**< 0.01**	0.71 (0.02)
2	1.00 (ref.)	0.47 (0.29-0.75)	0.57 (0.32-1.03)	0.60 (0.25-1.45)	**0.04**	0.82 (0.57-1.16)	0.26	0.78 (0.02)
3	1.00 (ref.)	0.47 (0.29-0.75)	0.62 (0.34-1.13)	0.69 (0.28-1.68)	0.07	0.85 (0.59-1.22)	0.37	0.79 (0.02)
Myers et al								
1	1.00 (ref.)	0.62 (0.42-0.91)	0.36 (0.23-0.56)	0.13 (0.06-0.26)	**< 0.01**	0.83 (0.70-0.98)	**0.03**	0.66 (0.02)
2	1.00 (ref.)	0.82 (0.55-1.21)	0.80 (0.49-1.30)	0.70 (0.32-1.53)	0.24	0.78 (0.56-1.07)	0.13	0.77 (0.02)
3	1.00 (ref.)	0.86 (0.58-1.27)	0.86 (0.53-1.40)	0.79 (0.35-1.76)	0.43	0.82 (0.59-1.15)	0.26	0.78 (0.02)

Table 2.5. (cont'd) Hazard Ratios (HR) of All-Cause Mortality by Measured and Predicted $\dot{V}O_{2max}$ in the Selected BLSA Sample (n = 1,080)

Author, Model	Sex- Specific Quartiles of $\dot{V}O_{2max}$				*P*-Trend	HR for 1 SD Increase	*P*-value	*C*-Statistic
	Q1	Q2	Q3	Q4				
Sloan et al; HR								
1	1.00 (ref.)	0.41 (0.22-0.75)	0.32 (0.17-0.60)	0.22 (0.10-0.47)	**< 0.01**	0.65 (0.51-0.83)	**< 0.01**	0.64 (0.04)
2	1.00 (ref.)	0.63 (0.34-1.15)	0.61 (0.32-1.17)	1.05 (0.44-2.50)	0.37	0.84 (0.59-1.20)	0.34	0.80 (0.03)
3	1.00 (ref.)	0.62 (0.34-1.16)	0.58 (0.30-1.12)	1.08 (0.44-2.61)	0.34	0.84 (0.59-1.21)	0.36	0.81 (0.02)
Sloan et al; No HR								
1	1.00 (ref.)	0.38 (0.24-0.59)	0.39 (0.25-0.60)	0.26 (0.15-0.43)	**< 0.01**	0.86 (0.73-1.02)	0.08	0.64 (0.02)
2	1.00 (ref.)	0.67 (0.42-1.05)	0.74 (0.47-1.15)	0.94 (0.53-1.66)	0.39	0.94 (0.72-1.24)	0.67	0.78 (0.02)
3	1.00 (ref.)	0.67 (0.43-1.06)	0.73 (0.46-1.14)	1.04 (0.58-1.88)	0.50	0.96 (0.72-1.26)	0.75	0.78 (0.02)

Model 1 = $\dot{V}O_{2max}$ quartiles; crude

Model 2 = Model 1 + age + sex

Model 3 = Model 2 + race and ethnicity + education

Table 2.6. Hazard Ratios (HR) of All-Cause Mortality by Measured and Recalibrated, Predicted $\dot{V}O_{2max}$ in the Selected BLSA Sample (n = 1,080)

	Sex- Specific Quartiles of $\dot{V}O_{2max}$							
Author, Model	**Q1**	**Q2**	**Q3**	**Q4**	***P*-Trend**	**HR for 1 SD Increase**	***P*-value**	***C*-Statistic**
Measured								
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 (0.02)
2	1.00 (ref.)	0.55 (0.37-0.81)	0.30 (0.17-0.54)	0.34 (0.16-0.75)	**< 0.01**	0.51 (0.39-0.66)	**< 0.01**	0.79 (0.02)
3	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 (0.02)
Baynard et al								
1	1.00 (ref.)	0.39 (0.26-0.58)	0.23 (0.14-0.37)	0.10 (0.05-0.20)	**< 0.01**	0.60 (0.50-0.71)	**< 0.01**	0.70 (0.02)
2	1.00 (ref.)	0.75 (0.48-1.16)	0.78 (0.43-1.39)	0.98 (0.39-2.46)	0.48	0.90 (0.64-1.27)	0.55	0.77 (0.02)
3	1.00 (ref.)	0.79 (0.51-1.22)	0.85 (0.47-1.55)	1.19 (0.47-3.05)	0.75	0.94 (0.66-1.33)	0.72	0.78 (0.02)
Bradshaw et al								
1	1.00 (ref.)	0.50 (0.33-0.73)	0.26 (0.16-0.42)	0.13 (0.07-0.25)	**< 0.01**	0.60 (0.50-0.71)	**< 0.01**	0.68 (0.02)
2	1.00 (ref.)	0.87 (0.57-1.32)	0.83 (0.48-1.46)	1.16 (0.50-2.70)	0.77	0.86 (0.63-1.17)	0.34	0.78 (0.02)
3	1.00 (ref.)	0.90 (0.59-1.36)	0.91 (0.51-1.62)	1.34 (0.56-3.19)	0.97	0.89 (0.65-1.22)	0.48	0.78 (0.02)
de Souza e Silva et al								
1	1.00 (ref.)	0.41 (0.27-0.60)	0.23 (0.14-0.37)	0.10 (0.05-0.20)	**< 0.01**	0.59 (0.49-0.70)	**< 0.01**	0.70 (0.02)
2	1.00 (ref.)	0.82 (0.53-1.27)	0.78 (0.44-1.39)	1.03 (0.41-2.59)	0.55	0.86 (0.61-1.21)	0.40	0.77 (0.02)
3	1.00 (ref.)	0.88 (0.57-1.36)	0.87 (0.48-1.58)	1.26 (0.50-3.23)	0.89	0.91 (0.64-1.29)	0.58	0.78 (0.02)
Jackson et al								
1	1.00 (ref.)	0.40 (0.27-0.60)	0.23 (0.14-0.38)	0.10 (0.05-0.21)	**< 0.01**	0.61 (0.51-0.72)	**< 0.01**	0.69 (0.02)
2	1.00 (ref.)	0.65 (0.43-1.00)	0.72 (0.41-1.26)	0.82 (0.35-1.96)	0.22	0.91 (0.66-1.25)	0.57	0.77 (0.02)
3	1.00 (ref.)	0.65 (0.42-0.99)	0.78 (0.44-1.39)	0.95 (0.40-2.29)	0.34	0.94 (0.68-1.30)	0.72	0.78 (0.02)
Jang et al								
1	1.00 (ref.)	0.40 (0.27-0.60)	0.23 (0.15-0.38)	0.10 (0.05-0.20)	**< 0.01**	0.59 (0.50-0.70)	**< 0.01**	0.70 (0.02)
2	1.00 (ref.)	0.79 (0.51-1.22)	0.84 (0.47-1.50)	1.07 (0.42-2.69)	0.66	0.90 (0.64-1.28)	0.56	0.78 (0.02)
3	1.00 (ref.)	0.83 (0.53-1.29)	0.91 (0.50-1.65)	1.33 (0.52-3.40)	0.96	0.94 (0.66-1.34)	0.72	0.78 (0.02)

Table 2.6. (cont'd) Hazard Ratios (HR) of All-Cause Mortality by Measured and Recalibrated, Predicted $\dot{V}O_{2max}$ in the Selected BLSA Sample (n = 1,080)

Author, Model	Sex- Specific Quartiles of $\dot{V}O_{2max}$ Q1	Q2	Q3	Q4	*P*-Trend	HR for 1 SD Increase	*P*-value	*C*-Statistic
Jurca et al								
1	1.00 (ref.)	0.39 (0.22-0.68)	0.21 (0.11-0.43)	0.09 (0.03-0.24)	**< 0.01**	0.52 (0.40-0.66)	**< 0.01**	0.71 (0.03)
2	1.00 (ref.)	0.66 (0.37-1.19)	0.74 (0.33-1.66)	0.92 (0.25-3.34)	0.42	0.82 (0.52-1.29)	0.38	0.80 (0.03)
3	1.00 (ref.)	0.64 (0.35-1.16)	0.73 (0.32-1.65)	1.03 (0.27-4.02)	0.43	0.81 (0.50-1.29)	0.37	0.81 (0.02)
Matthews et al								
1	1.00 (ref.)	0.41 (0.27-0.61)	0.24 (0.15-0.38)	0.11 (0.05-0.21)	**< 0.01**	0.60 (0.50-0.71)	**< 0.01**	0.69 (0.02)
2	1.00 (ref.)	0.70 (0.46-1.07)	0.74 (0.42-1.29)	0.79 (0.34-1.84)	0.25	0.88 (0.64-1.21)	0.43	0.77 (0.02)
3	1.00 (ref.)	0.69 (0.45-1.06)	0.79 (0.45-1.41)	0.88 (0.37-2.07)	0.37	0.92 (0.66-1.27)	0.60	0.78 (0.02)
Myers et al								
1	1.00 (ref.)	0.41 (0.27-0.60)	0.21 (0.13-0.35)	0.09 (0.04-0.19)	**< 0.01**	0.56 (0.47-0.67)	**< 0.01**	0.70 (0.02)
2	1.00 (ref.)	0.76 (0.50-1.15)	0.66 (0.37-1.19)	0.86 (0.33-2.22)	0.22	0.76 (0.54-1.08)	0.13	0.77 (0.02)
3	1.00 (ref.)	0.80 (0.52-1.21)	0.72 (0.40-1.31)	1.11 (0.42-2.95)	0.42	0.81 (0.57-1.16)	0.26	0.78 (0.02)
Sloan et al; HR								
1	1.00 (ref.)	0.39 (0.22-0.67)	0.19 (0.09-0.39)	0.08 (0.03-0.23)	**< 0.01**	0.51 (0.40-0.66)	**< 0.01**	0.71 (0.03)
2	1.00 (ref.)	0.61 (0.34-1.10)	0.61 (0.27-1.41)	0.78 (0.22-2.82)	0.21	0.81 (0.51-1.27)	0.36	0.80 (0.03)
3	1.00 (ref.)	0.60 (0.33-1.08)	0.59 (0.25-1.37)	0.86 (0.22-3.32)	0.21	0.80 (0.50-1.28)	0.35	0.81 (0.03)
Sloan et al; No HR								
1	1.00 (ref.)	0.40 (0.27-0.60)	0.23 (0.14-0.38)	0.10 (0.05-0.21)	**< 0.01**	0.61 (0.51-0.72)	**< 0.01**	0.69 (0.02)
2	1.00 (ref.)	0.65 (0.43-1.00)	0.72 (0.41-1.26)	0.82 (0.35-1.96)	0.22	0.91 (0.66-1.25)	0.57	0.77 (0.02)
3	1.00 (ref.)	0.65 (0.42-0.99)	0.78 (0.44-1.39)	0.95 (0.40-2.29)	0.34	0.94 (0.68-1.30)	0.72	0.78 (0.02)

Model 1 = $\dot{V}O_{2max}$ quartiles; crude

Model 2 = Model 1 + age + sex

Model 3 = Model 2 + race and ethnicity + education

Chapter 2, in part, is currently being prepared for submission for publication of the material. Schumacher, Benjamin T.; Di, Chongzhi; Bellettiere, John; LaMonte, Michael J.; Simonsick, Eleanor M.; Parada, Humberto; Hooker, Steven; LaCroix, Andrea Z. The dissertation author was the primary investigator and author of this paper.

3. Development, Validation, and Transportability of Several Machine-Learned, Non-Exercise Based $\dot{V}O_{2max}$ Prediction Models for Older Adults

Benjamin T. Schumacher, Michael J. LaMonte, Andrea Z. LaCroix, Eleanor M. Simonsick, Steven P. Hooker, Humberto Parada Jr., John Bellettiere, Arun Kumar

3.1. Abstract

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

Methods: Baltimore Longitudinal Study of Aging participants with valid $\dot{V}O_{2max}$ tests were included (n=1,080). LASSO, linear- and tree-boosted xgboost, random forest, and SVM algorithms were trained to predict $\dot{V}O_{2max}$. We developed these algorithms for: (1) the overall BLSA, (2) by sex, (3) using all BLSA variables, and (4) variables common in aging cohorts. Finally, we quantified the associations between measured and predicted $\dot{V}O_{2max}$ and mortality.

Results: Mean age was 69.0 (SD = 10.4) years and mean measured $\dot{V}O_{2max}$ was 21.6 (SD = 5.9) $mL \bullet kg^{-1} \bullet min^{-1}$. LASSO, linear- and tree-boosted xgboost, random forest, and SVM yielded root mean squared errors (RMSEs) of 3.4, 3.6, 3.4, 3.6, and 3.5 $mL \bullet kg^{-1} \bullet min^{-1}$, respectively. Incremental quartiles of measured $\dot{V}O_{2max}$ showed an inverse gradient in mortality risk. Predicted $\dot{V}O_{2max}$ variables yielded similar effect estimates but were not robust to adjustment.

Conclusion: Measured $\dot{V}O_{2max}$ 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 this vital sign can be more broadly studied as a modifiable target for promoting functional resiliency and healthy aging.

3.2. Introduction

An individual's cardiorespiratory fitness (CRF) refers to their circulatory and respiratory systems' capacity to provide oxygen to skeletal muscles for engaging in physical activity (PA).[9] While factors such as age, sex, health status, and genetics are strong determinants of CRF, one's level of habitual PA is the principal *modifiable* determinant of this physiological attribute.[9] CRF can be improved by increasing PA frequency, duration, and/or intensity, especially for sedentary individuals; however, CRF declines rapidly once PA declines in frequency, duration, and/or intensity, making CRF a commonly used objective surrogate marker of recent PA patterns. Scientific evidence accumulated over many years from clinical, epidemiologic, and exercise science studies has consistently shown higher CRF to have a strong, independent, beneficial association with a number of health-related factors and clinical outcomes. Higher CRF predicts lower incidence and mortality from coronary heart disease/cardiovascular disease[22–24], longer survival times[20,21,23,35,65], and lower rates of loss of independence for older adults.[34]

Maximal oxygen uptake ($\dot{V}O_{2max}$) is the gold standard measure of CRF and is recognized as a hallmark biomarker of healthy aging.[9,46] $\dot{V}O_{2max}$ 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, must include direct physician supervision to reduce the risk of adverse events. Because $\dot{V}O_{2max}$ testing involves strenuous activity to the point of absolute exhaustion, it is often contraindicated for vulnerable populations, including older adults. These features make direct measurement of $\dot{V}O_{2max}$ infeasible in large epidemiologic cohort studies.[36] In an attempt to provide more practical alternatives, researchers have published non-exercise based $\dot{V}O_{2max}$ prediction equations that can be used to approximate laboratory-measured $\dot{V}O_{2max}$ in various

contexts, including in large epidemiologic cohorts.[22,37–44] However, few of these equations were designed for use specifically in older adults.[40,42] A recent systematic review of the published $\dot{V}O_{2max}$ 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-derived models were developed in older adult populations.[66] By the year 2060, nearly one-fourth of the United States' (U.S.) population will be ≥ 65 years of age. Given the aforementioned strong, independent associations of higher CRF with a number of beneficial health outcomes, the ability to precisely estimate $\dot{V}O_{2max}$ in older adults is growing as critical need to enable continued investigation on the effects of cardiorespiratory fitness on healthy aging.[45]

Thus, in this study, we aimed to develop, validate, and compare multiple machine-learned, non-exercise based $\dot{V}O_{2max}$ prediction algorithms for older adults using laboratory-measured $\dot{V}O_{2max}$ in the Baltimore Longitudinal Study of Aging (BLSA). We aimed to develop these algorithms for the BLSA sample overall and within sex-specific strata, assess the associations between measured and predicted $\dot{V}O_{2max}$ and all-cause mortality for the total sample and within sex-specific strata, and assess the feasibility of transporting these algorithms for use in an external epidemiologic cohort of older women, the Women's Health Initiative's (WHI) Objective Physical Activity and Cardiovascular Health in Older Women (OPACH) Study where mortality follow-up is available.[67]

3.3. Methods

Study Population

The analytic sample was drawn from the BLSA, which is conducted by the National Institute on Aging Intramural Research Program.[49] Established in 1958, the BLSA is the longest on-going scientific study of aging.[47,48] The study protocol specifies visits by participants to the BLSA testing facility every one to four years for health, cognitive, and functional evaluations lasting three days. Since its inception, over 3,500 individuals have participated in the BLSA, and more than 1,300 remain active.[47] Extensive details about BLSA design, recruitment, and measurements are available elsewhere.[48] All participants provided written informed consent for the current study, which was approved by the applicable Institutional Review Boards.

Measures

$\dot{V}O_{2max}$ Measurement

Using a modified Balke treadmill testing protocol,[50,51] $\dot{V}O_{2max}$ was measured as milliliters of oxygen uptake per kilogram of body weight per minute ($mL \bullet kg^{-1} \bullet min^{-1}$). The participants walked on a treadmill at a set pace (3.0 miles per hour for women; 3.5 miles per hour for men) and the incline of the treadmill increased by 3% every two minutes until the participant indicated having reached exhaustion. Because participants included here were without known or suspected cardiopulmonary disease at the time of exercise testing, none of the data included in the present study were from tests terminated early due to medical contraindications. During this test, a Parkinson-Cowan gas meter was used to measure expired gas volumes. A medical mass spectrometer (Perkin-Elmer MGA-1110; calibrated daily using standard gases), was used to measure oxygen and carbon dioxide concentrations. Every 30 seconds during the test, average expired gas concentrations were calculated by a programmed interface between the gas meter and mass spectrometer, and $\dot{V}O_{2max}$ was defined as the highest 30-second oxygen uptake value.

Maximal effort on the treadmill test was specified as a respiratory exchange ratio (RER) greater than 1.0. Of 52 participants with an RER value just below the cutoff when the treadmill was stopped, 11 achieved ≥ 85% of their age-predicted maximal heart rate in beats per minute (bpm; computed as 220 - age in years) and had a value greater than 17 on the 20-point Borg rating of perceived exertion (RPE) scale. Their test results were considered to reflect maximal effort and were included in the present analysis. Of the remaining 41 participants with an RER less than 1.0 when the treadmill was stopped, 31 had no other $\dot{V}O_{2max}$ test meeting the aforementioned criteria and were excluded from the present analysis, and 10 provided a subsequent $\dot{V}O_{2max}$ test that satisfied these maximal test criteria and that subsequent measurement was included, resulting in a final analytic sample of 1,080 participants. For participants having more than one $\dot{V}O_{2max}$ measurement, only the first measurement meeting the maximal effort criteria was analyzed in the present study.

Outcome Ascertainment

Participant information was linked to the National Death Index[68] to ascertain vital status and, for those deceased, their date of death. Vital status surveillance using the National Death Index has been shown to provide an accurate mortality follow-up even in historical cohort datasets[69] even when limited personal identifying information is available.[70] Follow-up occurred from the participant's $\dot{V}O_{2max}$ test date (the earliest $\dot{V}O_{2max}$ test was administered on January 1, 2007), 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: 0.60 to 14.1 years).

Covariates

Demographics and Physical Attributes

Demographic variables included self-identified sex (male, 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), the participant's age at $\dot{V}O_{2max}$ testing, height (centimeters; cm) measured using a stadiometer, weight (kilograms; kg) measured using a calibrated scale, body mass index (BMI) calculated as weight divided by height (meters) squared, and waist circumference (cm) using a tape measure.

Health Status/Health History/Functional Capacity

Health status variables included the SF-12 self-rated health[54] scale and its physical and mental health composite scores, hand grip muscle strength scores (kg) in both hands, and Short Physical Performance Battery (SPPB) physical function score (0-12, higher is better) and its three component pieces[71,72]: (1) the number of seconds to complete five sit-to-stand movements from a chair, (2) whether the participant was able to balance with their feet placed side by side, semi-tandem, and in tandem for 10 seconds each, and (3) the number of seconds needed to complete a four-meter walk. Additional timed walk tests included the number of meters walked at usual pace for 2.5 minutes, 2.5-minute walk pace (m/s), the number of seconds to walk 400m at a fast pace, 400m fast walk pace (m/s), and a walking capacity summary score. In brief, the walking capacity summary score is an aggregate index score from participant's responses to several questions about their ability to walk a variety of distances, with a score of 0 representing an inability to walk ¼ mile and 9 representing the ability to easily walk one mile. Details about the derivation of the walking capacity summary score have been published elsewhere.[73] Health history variables included dichotomous indicators (yes/no) for a physician diagnosis of

myocardial infarction, congestive heart failure (CHF), stroke, diabetes glucose intolerance, high blood sugar, and breast cancer. Additionally, measurements of seated, resting systolic and diastolic blood pressure from both arms, resting heart rate, and heart rate at the end of the 2.5-minute usual pace walk.

Health Behaviors

Self-reported health behavior variables were included: smoking history (never, current, or former smoker), calories expended in all activity, calories expended in all activities per kilogram of body weight, calories expended in exercise related activity as scored by the Harvard alumni scale[74], minutes of any exercise per week (0-29, 30 -74, 75 – 149, or 150+ minutes), minutes of any walking per week, minutes of brisk walking per week, minutes of vigorous activity per week, beta blocker use (yes/no), and blood sugar medication use (yes/no).

Machine Learning Algorithms

We first trained a Least Absolute Shrinkage and Selection Operator (LASSO) model. The LASSO process yields a parsimonious model as it adds a penalty term that is equivalent to the absolute value of the magnitude of coefficients. The penalty coefficient, lambda, was tuned using a grid search and the lambda value that yielded the lowest mean squared error (MSE) in the 10-fold cross-validation (CV) process was used to train the final algorithm. This final algorithm was then used to predict $\dot{V}O_{2max}$ ($mL \bullet kg^{-1} \bullet min^{-1}$) and the root MSE (RMSE) between measured and predicted $\dot{V}O_{2max}$. LASSO was implemented using the glmnet package in R.[75]

The next algorithm, extreme gradient boosting (XGBoost), was employed using two separate boosters, linear and tree. In brief, XGBoost is an extension of the gradient boosting framework as it iteratively fits a model to the data, fits a subsequent model based on the previous

model's residuals, and fits another subsequent model using previous models to minimize the gradient descent function. Both algorithms were trained on the BLSA data using 10-fold CV and a grid search to tune the hyperparameters. The linear-boosted model hyperparameters included: the number of boosting rounds (nrounds), L1 (LASSO) regularization weight (alpha), and learning rate (eta). Fine-tuning of the hyperparameters was conducted until no hyperparameters were at their boundary value. The combination of these hyperparameters that yielded the minimum RMSE in the 10-fold CV process was selected as the final algorithm. The same process was executed for the tree-boosted algorithm, but the hyperparameters included nrounds, the maximum depth of the tree (max_depth), eta, alpha, subsample ratio of the training instances (subsample), and the subsample ratio of columns when constructing each tree (colsample_bytree). The combination of these hyperparameters that yielded the minimum predictive error (RMSE) in the 10-fold CV process was selected as the final algorithm. The minimum 10-fold CV test RMSE values are presented in Table 2. XGBoost was implemented using the xgboost package in R.[76]

Next, random forest models were trained to predict $\dot{V}O_{2max}$. Random forest is a tree-based ensemble algorithm where every tree is trained on a bootstrapped sample of the training data, tested against the sample not in the bootstrapped sample (the out-of-bag sample; OOB), and the prediction from all trees are averaged to get the predicted value. The following hyperparameters were tested using a grid search approach: the number of trees to grow (ntrees), the number of variables randomly sampled as candidates at each split (mtry), the proportion of the training data to include in the bootstrapped sample vs. remaining OOB (sampsize), and the minimum size of terminal nodes (node_size). The combination of hyperparameters that minimized the OOB RMSE was selected as the final model.

The final algorithm trained to predict $\dot{V}O_{2max}$ in the BLSA was Support Vector Regression (SVR), a specific application of Support Vector Machine (SVM). In brief, SVR tries to fit hyperplane through kernel transformation to best classify the data points. Similar to the approaches of the other ML algorithms, 10-fold CV was used to find the combination of the cost of constraint violations (cost) and radial kernel coefficient (gamma) that minimized the test RMSE.

Variable importance scores for the linear-boosted xgboost algorithm were extracted as Weight (the linear coefficient of each variable; a higher percentage indicates a more important predictive feature) and as Gain for the tree-boosted xgboost algorithm (contribution of each variable to the model calculated by taking each variable's contribution to each tree in the model; a higher percentage indicates a more important predictive feature). Variable importance scores for the random forest were extracted as the percent change in the OOB MSE with the given variable excluded from the algorithm (a higher percentage indicates a more important predictive feature).

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 if the results are transportable to an external cohort, using only the variables common between BLSA and OPACH, for a total of 24 combinations.

OPACH Covariates

We assessed the ML algorithms' performance when the universe of eligible predictors was restricted to BLSA variables that also exist in OPACH. Extensive details about OPACH

have been published elsewhere.[67] For the purposes of the present study, the OPACH dataset contained all the BLSA covariates except for measures of rapid gait speed, 2.5-minute usual pace walk, 400m fast walk, walking capacity summary score, and heart rate measures during and after the 2.5-minute walk.

Statistical Analysis

Analysis of variance (ANOVA) tests for continuous variables and chi-square tests for categorical variables were used to compare baseline covariates by sex-specific quartiles of measured $\dot{V}O_{2max}$. Correlations between measured $\dot{V}O_{2max}$, all predicted $\dot{V}O_{2max}$ variables, age, BMI, and SPPB were calculated.

Next, Cox proportional hazards regression models were used to estimate the associations between quartiles of $\dot{V}O_{2max}$ (measured and predicted $\dot{V}O_{2max}$) and all-cause mortality. 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 $\dot{V}O_{2max}$ as a continuous variable estimating adjusted hazard ratios for all-cause mortality associated with a one standard deviation increase in $\dot{V}O_{2max}$. *P*-values for mean-centered, standard deviation scaled $\dot{V}O_{2max}$ variable for models 1 and 2 are presented. The concordance statistic (*C*-Statistic), a measure of discrimination for time-to-event models which gives the proportion of participant pairs for which the model correctly predicts the participant in the pair who experiences a mortality event first, are also presented.[77] To test whether the proportional hazards assumption was violated, we used the cox.zph function in the R survival package.[60] The correlation of the scaled Schoenfeld residuals for each covariate (and for the

whole model) with time was examined to ensure independence of residuals and time. No violations in the proportional hazards assumption were found.

All analyses were conducted in R version 3.6.3 (R Foundation for Statistical Computing, Vienna, Austria).

3.4. Results

Sample Characteristics

For the 565 women and 515 men with laboratory measures of $\dot{V}O_{2max}$ that achieved the criteria for maximal effort, mean age was 69.0 (SD = 10.4) years, mean BMI was 27.0 (SD = 4.4) kg/m^2, and mean measured $\dot{V}O_{2max}$ was 21.6 (SD = 5.9) mL•kg^{-1}•min^{-1} (see Table 3.1.). The median $\dot{V}O_{2max}$ for the men was 23.7 (SD = 6.1) mL•kg^{-1}•min^{-1} (range: 9.5 – 48.9 mL•kg^{-1}•min^{-1}) and the median $\dot{V}O_{2max}$ for the women was 19.9 (SD = 5.1) mL•kg^{-1}•min^{-1} (range: 6.2 – 42.1 mL•kg^{-1}•min^{-1}). 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 or could not be classified. The majority of participants (61.9%) had post-college education. The prevalence of current smoking was 1.8%. Mean systolic and diastolic blood pressure was 114.1 (SD = 14.1) mmHg and 66.7 (SD = 8.8) mmHg, respectively. Education, diastolic blood pressure, self-rated health status, and self-reported exercise were positively associated with increasing quartiles of measured $\dot{V}O_{2max}$, while age, systolic blood pressure, BMI, and current smoking status were inversely associated with increasing $\dot{V}O_{2max}$ quartiles (see Table 3.1.).

Performance of Machine-Learned $\dot{V}O_{2max}$ Prediction Algorithms

The first algorithm, LASSO, yielded an RMSE of 3.4 mL•kg^{-1}•min^{-1} for $\dot{V}O_{2max}$ prediction in the total sample using all predictors (see Table 3.2.). For the subgroups (sex-stratified in combination with the BLSA-predictor and OPACH-predictor algorithms), predicted $\dot{V}O_{2max}$ RMSEs ranged from 2.9 to 3.9 mL•kg^{-1}•min^{-1} for the women BLSA-predictor and men OPACH-predictor, respectively. The linear xgboost yielded an RMSE of 3.6 mL•kg^{-1}•min^{-1} for $\dot{V}O_{2max}$ prediction in the total sample using all predictors and OPACH predictors. For the subgroups, RMSEs ranged from 3.2 to 4.0 mL•kg^{-1}•min^{-1} for both women's algorithms and both men's algorithms, respectively. The tree-boosted xgboost algorithm yielded an RMSE of 3.4 mL•kg^{-1}•min^{-1} for $\dot{V}O_{2max}$ prediction in the total sample using all predictors. For the subgroups, RMSEs ranged from 3.0 to 4.0 mL•kg^{-1}•min^{-1} for the women BLSA-predictor and men OPACH-predictor algorithms, respectively. The random forest algorithm yielded an RMSE of 3.6 mL•kg^{-1}•min^{-1} for the total sample using all predictors. For the subgroups, RMSEs ranged from 2.9 to 4.2 mL•kg^{-1}•min^{-1} for the women BLSA-predictor and men OPACH-predictor algorithms, respectively. The SVR algorithm yielded an RMSE of 3.5 mL•kg^{-1}•min^{-1} for the total sample using all predictors. For the subgroups, RMSEs ranged from 2.8 to 4.1 mL•kg^{-1}•min^{-1} for the women BLSA-predictor and men 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^{-1}•min^{-1}). LASSO had the best RMSE for the entire sample when using the OPACH predictors (3.5 mL•kg^{-1}•min^{-1}). Further details about the combination of subgroups can be found in Table 3.2. Finally, for all algorithms the RMSE values for the women were lower than the RMSE values for the men.

Correlations of Measured and Predicted $\dot{V}O_{2max}$ with Selected Covariates

Correlations between measured $\dot{V}O_{2max}$, all predicted $\dot{V}O_{2max}$ estimates, age, BMI, and SPPB are show in in Supplemental Table 3.5. In short, the correlations between predicted $\dot{V}O_{2max}$ and measured $\dot{V}O_{2max}$ ranged rom 0.80 (OPACH-predictor linear-boosted xgboost) to 0.93 (BLSA-predictor tree-boosted xgboost). All predicted $\dot{V}O_{2max}$ estimates were more strongly associated with age, BMI, and SPPB than measured $\dot{V}O_{2max}$.

Associations of Measured and Predicted $\dot{V}O_{2max}$ with All-Cause Mortality

When assessing the associations between quartiles of measured $\dot{V}O_{2max}$ and all-cause mortality, a steep inverse gradient in mortality risk across incremental $\dot{V}O_{2max}$ quartiles was evident in both unadjusted and adjusted models. Adjusting for Model 2 covariates, the HRs (95% CI) 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 $\dot{V}O_{2max}$, respectively, $P_{trend} < 0.01$ (see Table 3.3.). When evaluated in continuous format, every one SD increment (5.9 $mL \bullet kg^{-1} \bullet min^{-1}$) in measured $\dot{V}O_{2max}$ was, on average, 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 $\dot{V}O_{2max}$ prediction algorithm demonstrated patterns that were similar to those seen for measured $\dot{V}O_{2max}$—an inverse gradient in mortality risk across incremental predicted $\dot{V}O_{2max}$ quartiles (Q4 HRs ranged from 0.09 – 0.17). However, adjustment for the model 2 covariates attenuated the HRs for Q2 - Q4, and, while the majority of the confidence intervals 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 one SD increment in predicted $\dot{V}O_{2max}$ were similar to that seen for measured $\dot{V}O_{2max}$ (HRs

ranged from 0.48 to 0.61). The *C*-statistics for all predicted $\dot{V}O_{2max}$ models were 0.78 and 0.79 after adjustment for Model 2 covariates (see Table 3.3. for the *C*-statistics' 95% CIs).

Among the BLSA men, there were 91 deaths: 53, 27, 8, and 3 in Q1 – Q4 of measured $\dot{V}O_{2max}$, respectively. Among the BLSA women, there were 50 deaths: 28, 11, 6, and 5 in Q1 – Q4 of measured $\dot{V}O_{2max}$, respectively. Sex-specific associations for measured and predicted $\dot{V}O_{2max}$ with all-cause mortality can be found in Supplemental Tables 3.6 (men) and 3.7 (women). In the unadjusted and adjusted models, higher measured $\dot{V}O_{2max}$ 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 $\dot{V}O_{2max}$ 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.

Variable Importance Scores

The five most important variables in the tree-boosted BLSA-predictor xgboost algorithm were, in order from more-to-less important: (1) number of seconds to complete the 400m walk, (2) calories expended in self-reported exercise, (3) Harvard alumni calorie expenditure, (4) right-hand grip muscle strength, and (5) diastolic blood pressure. The five most important variables in the linear-boosted BLSA-predictor xgboost algorithm were: (1) non-Hispanic Other race, (2) usual gait speed in the 2.5-minute walk, (3) history of myocardial infarction, and (5) being a former smoker. The five most important variables in the random forest BLSA-predictor xgboost

algorithm were: (1) number of seconds to complete the 400m walk, (2) the balance component of the SPPB, (3) meters walked in the 2.5-minute walk, (4) 2.5-minyet gait speed, and (5) weight. In summary, when using all of the variables in the BLSA, the number of seconds to complete the 400m 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 400m walk), age became the most important variable. See Table 3.4. for the top 10 most important variables for the 18 combinations

3.5. Discussion

We developed and assessed the performance of multiple ML, non-exercise based $\dot{V}O_{2max}$ prediction algorithms that may enable large-scale epidemiologic cohorts with older, ambulatory, community-dwelling adults to accurately estimate $\dot{V}O_{2max}$, 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 $mL \bullet kg^{-1} \bullet min^{-1}$. For additional context, if one assumes the standard conversion of 3.5 $mL \bullet kg^{-1} \bullet min^{-1}$ as being equivalent to 1 metabolic equivalent (MET), the errors in $\dot{V}O_{2max}$ prediction based on the ML algorithms used herein were about 0.8 and 1.2 METs. These predictive error values are lower than previously published non-exercise based $\dot{V}O_{2max}$ prediction equations derived using ordinary least squares (see Chapter 2.4.) and lower than several RMSEs of previously published ML $\dot{V}O_{2max}$ prediction algorithms.[66] Further, these non-exercise based predictive error values are comparable to those obtained when predicting $\dot{V}O_{2max}$ using exercise based covariates such as the duration of maximal treadmill exercise tests[78] and timed walk tests.[79] These RMSE values, coupled with the strong correlations between predicted and measured $\dot{V}O_{2max}$, further enhances confidence in the $\dot{V}O_{2max}$ prediction algorithms

described herein, even when performance-based assessment of cardiorespiratory fitness 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 $mL \bullet kg^{-1} \bullet min^{-1}$) followed by the two xgboost algorithms and SVR at 3.6 $mL \bullet kg^{-1} \bullet min^{-1}$. Across all the algorithms, the RMSE values for the women were lower than the men. This is likely due to the larger variation in men's $\dot{V}O_{2max}$ measurements than the women's $\dot{V}O_{2max}$. Despite the better prediction of $\dot{V}O_{2max}$ for the BLSA women than men, the associations between measured and predicted $\dot{V}O_{2max}$ 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 $\dot{V}O_{2max}$, or at least 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-meter walk, an objective measure of physical performance capacity, is the most important variable in $\dot{V}O_{2max}$ prediction (RMSE = 3.6 $mL \bullet kg^{-1} \bullet min^{-1}$). However, since OPACH does not have a 400-meter walk measure, age becomes the most important variable in the OPACH-predictor random forest algorithms, but the effectiveness of this model to predict $\dot{V}O_{2max}$ is nearly identical (RMSE = 3.7 $mL \bullet kg^{-1} \bullet min^{-1}$). 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 $\dot{V}O_{2max}$ prediction ML models have been previously published, and even fewer have been developed specifically for older adults. Chapter 2 (namely Chapter 2.4.) on assessing the performance of previously published OLS models showed that when these OLS models are used to predict $\dot{V}O_{2max}$ in the BLSA, the RMSE values range from 5.1 (using equations from Bradshaw et al.[37] and Sloan et al.'s HR equation[41]) to 20.4 (Jang et al.[43]) $mL \bullet kg^{-1} \bullet min^{-1}$. After recalibrating these formulas' to measured $\dot{V}O_{2max}$ in the BLSA (obtaining new regression weights derived from the distribution of covariates in the BLSA) the RMSE values decrease to 3.8 (Bradshaw et al.[37]) to 4.2 (Myers et al.[22]) $mL \bullet kg^{-1} \bullet min^{-1}$. A recent meta-analysis of 16 $\dot{V}O_{2max}$ prediction equations that use ML[66], 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; oldest age range included in the meta-analysis was 18-65), found RMSEs ($mL \bullet kg^{-1} \bullet min^{-1}$) of 2.9 (SVM), 3.14 (MLP Neural Network), 3.38 (tree boost), 4.78 (multilayer perceptron; MLP), 4.07 (artificial neural networks; ANN), 2.91 (feature selection with SVM), 3.37 (Generalized Regression Neural Networks), 4.51 (Single Decision Tree), and 4.78 (Multiple input single output (MISO) with MLP, SVM, and ANN with RBF). Interestingly, in the MISO model, the RMSEs were 4.07 for the women and 5.30 for the men, suggesting the sex differences as also 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, perhaps due to the decreased variance in $\dot{V}O_{2max}$ in the older adults included herein.

While several of the ML algorithms yielded reasonable predictions of $\dot{V}O_{2max}$, as indicated by the relatively low RMSEs, the utility of predicted $\dot{V}O_{2max}$ in estimating mortality risk was not as clear as compared to measured $\dot{V}O_{2max}$. In unadjusted models, all predicted

$\dot{V}O_{2max}$ variables produced HRs comparable to measured $\dot{V}O_{2max}$. However, after adjustment for even the limited set of Model 2 covariates, these HRs were attenuated compared to measured $\dot{V}O_{2max}$, though significant inverse trends in mortality risk still were evident in men, less so in women. The *C*-statistics were comparable for measured and predicted $\dot{V}O_{2max}$. Measurement of $\dot{V}O_{2max}$ using indirect calorimetry surely provides a more accurate representation of the underlying physiological construct of CRF than is possible using prediction approaches, derived from host factors related to demographics, body habitus, and physical performance that are correlates of $\dot{V}O_{2max}$. However, the present study indicates that ML prediction of $\dot{V}O_{2max}$ in older adults has relatively low prediction error and is associated with a clinical aging outcome, all-cause mortality, in a similar pattern and magnitude of association as measured $\dot{V}O_{2max}$ in unadjusted analysis. The attenuation of associations with mortality for predicted $\dot{V}O_{2max}$ but not measured $\dot{V}O_{2max}$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 $\dot{V}O_{2max}$. Replication of the present investigation using large study samples with greater numbers of outcome events for analysis are needed to clarify and build upon our findings.

This study, in direct response to the call for future research in the aforementioned ML meta-analysis[66], 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 sex stratified. Additionally, we provided these metrics and associations when using a restricted universe of variables likely to be available in most aging cohort studies to assess the transportability of these algorithms. Another substantial strength of the present study is the prospective follow-up, enabling the evaluation of the accuracy of predicted $\dot{V}O_{2max}$ with respect to measured $\dot{V}O_{2max}$ and their

associations with mortality. BLSA enrolled a large group of racially and ethnically diverse older adults, included objectively measured $\dot{V}O_{2max}$, followed participants for mortality outcomes after $\dot{V}O_{2max}$ assessment, and collected data that enabled adjustment for confounders.

In conclusion, measured $\dot{V}O_{2max}$ is a strong predictor of all-cause mortality in aging men and women enrolled in the BLSA, which further supports the recognition of $\dot{V}O_{2max}$ as a biomarker of aging resiliency. Given the infeasibility of direct measurement of $\dot{V}O_{2max}$ in large epidemiologic cohorts, simple linear regression models have been proposed to predict $\dot{V}O_{2max}$ to guide exercise prescription in older adults, but these more simplistic predicted $\dot{V}O_{2max}$ measures are not robust to adjustment in multivariable analyses (see Chapter 2.4.). Using ML can improve the accuracy of $\dot{V}O_{2max}$ 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 $\dot{V}O_{2max}$ in community-dwelling older adults so that this critical "vital sign" can be more broadly studied as a modifiable target for promoting functional resiliency and healthy aging.

Table 3.1. Characteristics of BLSA participants overall and according to quartiles of measured $\dot{V}O_{2max}$ (n = 1,080)

		Measured $\dot{V}O_{2max}$				
Characteristic	**Total (n = 1,080)**	**Quartile 1† (n = 270)**	**Quartile 2† (n = 277)**	**Quartile 3† (n = 265)**	**Quartile 4† (n = 268)**	**P-value***
Deaths, n (%)	141 (13.1)	81 (30.0)	38 (13.7)	14 (5.3)	8 (3.0)	**< 0.01**
Age, Mean (SD)	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, n (%)						**< 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, n (%)						**< 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), Mean (SD)	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, n (%)	152 (14.1)	78 (28.9)	39 (14.1)	22 (8.3)	13 (4.9)	**< 0.01**
Minutes of Exercise, n (%)						**< 0.01**
0-29	465 (43.1)	171 (63.3)	127 (45.8)	93 (35.1)	74 (27.6)	
30 - 74	169 (15.6)	36 (13.3)	48 (17.3)	33 (12.5)	52 (19.4)	
75 - 149	165 (15.3)	25 (9.3)	42 (15.2)	52 (19.6)	46 (17.2)	
150+	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)	

Table 3.1. (cont'd) Characteristics of BLSA participants overall and according to quartiles of measured $\dot{V}O_{2max}$ (n = 1,080)

		Measured $\dot{V}O_{2max}$				
Characteristic	**Total (n = 1,080)**	**Quartile 1† (n = 270)**	**Quartile 2† (n = 277)**	**Quartile 3† (n = 265)**	**Quartile 4† (n = 268)**	**P-value***
Self-Rated Health, n (%)						**< 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), Mean (SD)	114.1 (14.1)	117.3 (14.8)	116 (13.3)	113 (13.7)	110.2 (13.3)	**< 0.01**
Diastolic BP (mmHg), Mean (SD)	66.7 (8.8)	65 (8.4)	66.3 (9.3)	66.9 (8.6)	68.5 (8.5)	**< 0.01**
Smoking Status, n (%)						**< 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						
$\dot{V}O_{2max}$ (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, Mean (SD)	3.3 (68.1)	1.2 (0.1)	1.2 (0.1)	1.2 (0.1)	9.5 (136.8)	0.18
Borg Score, Mean (SD)	16.5 (1.7)	16.1 (1.7)	16.2 (1.7)	16.7 (1.7)	17 (1.6)	**< 0.01**
% of Max. Predicted HR, Mean (SD)**	98.8 (50.2)	89.6 (13.3)	97.5 (9.3)	100.3 (8.5)	107.8 (98.3)	**< 0.01**

SD = standard deviation; BMI = body mass index; BP = blood pressure

*P-value for continuous variables from the One-way ANOVA and Chi-Sq. goodness of fit test for categorical variables across $\dot{V}O_{2max}$ quartiles.

**Maximum predicted heart rate: 220 - age

Bold indicates significance at the $P < 0.05$ level.

†Sex-specific quartile definitions were as follows:

Q1: Men : < 19.9; n = 129 & Women: < 16.5; n = 141

Q2: Men : ≥ 19.9 & ≤ 23.7; n = 131 & Women: ≥ 16.5 & ≤ 19.9; n = 146

Q3: Men : > 23.7 & ≤ 27.4; n = 128 & Women: > 19.9 & ≤ 23.7; n = 137

Q4: Men : > 27.4; n = 127 & Women: > 23.7; n = 141

Table 3.2. Root Mean Square Errors (RMSE) for Various ML Algorithms

Sample, Universe of Predictors	LASSO	xgboost, Linear	xgboost, Tree	Random Forest	SVR
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

RMSE units in units of $\dot{V}O_{2max}$: mL/kg/min.

BLSA = Baltimore Longitudinal Study of Aging

OPACH = Objective Physical Activity and Cardiovascular Health Study

LASSO = Least Absolute Shrinkage and Selection Operator; xgboost = Extreme Gradient Boosted

SVR = Support Vector Regression

Table 3.3. Hazard Ratios (HR) of All-Cause Mortality by Measured and Predicted $\dot{V}O_{2max}$ in the BLSA (n = 1,080)

Sample, Universe of Predictors	Model	Q1	Q2	Q3	Q4	HR for 1 SD Increase	*C*-Statistic
Measured $\dot{V}O_{2max}$	1	1.00 (ref.)	0.43 (0.29-0.63)	0.16 (0.09-0.29)	0.10 (0.05-0.20)	0.46 (0.38-0.57)	0.71 (0.67-0.75)
Measured $\dot{V}O_{2max}$	2	1.00 (ref.)	0.55 (0.37-0.82)	0.30 (0.17-0.54)	0.34 (0.15-0.75)	0.50 (0.38-0.66)	0.79 (0.75-0.83)
Total BLSA, All BLSA Predictors							
xgboost, Tree	1	1.00 (ref.)	0.36 (0.24-0.53)	0.18 (0.11-0.30)	0.09 (0.05-0.19)	0.48 (0.39-0.58)	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)	0.49 (0.37-0.66)	0.79 (0.75-0.83)
xgboost, Linear	1	1.00 (ref.)	0.39 (0.26-0.58)	0.18 (0.10-0.30)	0.14 (0.07-0.26)	0.53 (0.44-0.63)	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)	0.61 (0.46-0.80)	0.78 (0.74-0.82)
Random Forest	1	1.00 (ref.)	0.47 (0.29-0.77)	0.19 (0.10-0.38)	0.11 (0.05-0.27)	0.49 (0.38-0.62)	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)	0.52 (0.37-0.72)	0.79 (0.73-0.85)
SVM	1	1.00 (ref.)	0.38 (0.23-0.64)	0.18 (0.09-0.34)	0.14 (0.06-0.30)	0.54 (0.43-0.67)	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)	0.57 (0.40-0.80)	0.78 (0.72-0.84)
Total BLSA, OPACH Predictors							
xgboost, Tree	1	1.00 (ref.)	0.36 (0.24-0.54)	0.18 (0.11-0.31)	0.11 (0.05-0.22)	0.50 (0.41-0.60)	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)	0.51 (0.38-0.69)	0.79 (0.75-0.83)
xgboost, Linear	1	1.00 (ref.)	0.40 (0.27-0.60)	0.23 (0.14-0.38)	0.11 (0.06-0.23)	0.54 (0.45-0.64)	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)	0.60 (0.45-0.81)	0.78 (0.74-0.82)
Random Forest	1	1.00 (ref.)	0.35 (0.23-0.55)	0.16 (0.09-0.30)	0.10 (0.04-0.21)	0.46 (0.37-0.57)	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)	0.48 (0.35-0.67)	0.78 (0.74-0.82)
SVM	1	1.00 (ref.)	0.52 (0.32-0.85)	0.19 (0.10-0.38)	0.17 (0.08-0.37)	0.55 (0.44-0.69)	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)	0.59 (0.41-0.86)	0.79 (0.73-0.85)

Model 1 = $\dot{V}O_{2max}$ quartiles; crude

Model 2 = Model 1 + age + race and ethnicity + education

All *P*-values were statistically significant at the 0.05 level except for Total BLSA, OPACH Predictors SVM Model 2.

Table 3.4. 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
Random Forest						
1	400m walk time	Age	400m walk time	Age	400m walk time	Age
2	SPPB - balance	SPPB - chair stands	HR at end of 2.5min walk	Left-hand grip	2.5min gait meters	Calories
3	2.5min gait meters	Waist circumference	Right-hand grip	SF12: physical	Calories per kg weight	Mins. of any walking/wk
4	2.5min gait speed	Calories per kg weight	Left-hand grip	Calories	SF12: physical	Left-hand grip
5	Weight	SF12: physical	Self-rated health	Right-hand grip	Calories	Height
6	Usual gait speed	Calories	Diastolic BP	Race/Ethnicity	SPPB - chair stands	Right-hand grip
7	Height	Race/Ethnicity	Calories per kg weight	Radial pulse	Calories - Harvard scale	Race/Ethnicity
8	Calories - Harvard scale	Mins. of any walking/wk	Rapid gait speed	Diastolic BP	HR at end of 2.5min walk	Mins. of vigorous activity/wk
9	Waist circumference	Radial pulse	Walking score	Mins. of vigorous activity/wk	HR at end of 400m walk	SF12: mental
10	SF12: physical	Mins. of vigorous activity/wk	Calories	Height	Mins. of vigorous activity/wk	Systolic BP

Table 3.4. (cont'd) 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	2.5min gait speed	Very good/good self-rated health	150+ mins/wk exercise	History: heart failure/CHF	2.5min gait speed	Former smoker
3	History: heart attack or MI	History: heart attack or MI	30-74 mins/wk exercise	150+ mins/wk exercise	30-74 mins/wk exercise	30-74 mins/wk 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+ mins/wk exercise	Usual gait speed	Usual gait speed	History: heart attack or MI	150+ mins/wk exercise
6	History: breast cancer	Former smoker	Never smoker	30-74 mins/wk 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+ mins/wk exercise	History: diabetes	History: heart failure/CHF	Non-college graduate	History: sroke	non-Hispanic, White
9	History: diabetes	75-149 mins/wk 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

Table 3.4. (cont'd) 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, Tree						
1	400m walk time	Age	400m walk time	Age	400m walk time	BMI
2	Calories	SF12: physical	Mins. of any walking/wk	Diastolic BP	Calories per kg weight	Right-hand grip
3	Calories - Harvard scale	Mins. of vigorous activity/wk	Radial pulse	Weight	Waist circumference	Height
4	Right-hand grip	non-Hispanic, Black race	Systolic BP	Right-hand grip	SF12: physical	Weight
5	Diastolic BP	Height	SF12: mental	SF12: mental	Mins. of vigorous activity/wk	Calories
6	Walking score	Mins. of any walking/wk	2.5min gait meters	Beta blocker use	non-Hispanic, Black race	Waist circumference
7	SF12: physical	Calories per kg weight	Calories per kg weight	Calories per kg weight	Mins. of any walking/wk	non-Hispanic, Black race
8	Weight	Calories	Weight	Mins. of any walking/wk	Weight	Radial pulse
9	Radial pulse	Beta blocker use	Calories	Calories	Radial pulse	Systolic BP
10	Waist circumference	Diastolic BP	Height	Systolic BP	Height	Diastolic BP

Supplemental Table 3.5. Correlations of Measured and Predicted $\dot{V}O_{2max}$ with Selected Covariates in BLSA

	Measured $\dot{V}O_{2max}$	BLSA-Predictor Linear xgboost	OPACH-Predictor Linear xgboost	BLSA-Predictor Tree xgboost	OPACH-Predictor Tree xgboost	BLSA-Predictor SVR	OPACH-Predictor SVR	Age	BMI	SPPB
Measured $\dot{V}O_{2max}$	1	0.81	0.80	0.93	0.89	0.86	0.83	-0.44	-0.26	0.33
BLSA-Predictor Linear xgboost		1	0.99	0.92	0.93	0.96	0.97	-0.56	-0.33	0.37
OPACH-Predictor Linear xgboost			1	0.91	0.94	0.95	0.98	-0.57	-0.34	0.35
BLSA-Predictor Tree xgboost				1	0.96	0.96	0.93	-0.51	-0.31	0.37
OPACH-Predictor Tree xgboost					1	0.95	0.96	-0.52	-0.33	0.38
BLSA-Predictor SVR						1	0.97	-0.55	-0.30	0.41
OPACH-Predictor SVR							1	-0.56	-0.32	0.40
Age								1	-0.18	-0.30
BMI									1	-0.03
SPPB										1

BLSA = Baltimore Longitudinal Study of Aging

OPACH = Objective Physical Activity and Cardiovascular Health Study

SVR = Support Vector Regression

SPPB = Short Physical Performance Battery

Supplemental Table 3.6. Hazard Ratios (HR) of All-Cause Mortality by Measured and Predicted $\dot{V}O_{2max}$ in the BLSA Men (n = 515)

Sample, Universe of Predictors	Model	Q1	Q2	Q3	Q4	HR for 1 SD Increase	*C*-Statistic
Measured $\dot{V}O_{2max}$	1	1.00 (ref.)	0.44 (0.28-0.70)	0.13 (0.06-0.26)	0.05 (0.02-0.17)	0.31 (0.24-0.41)	0.73 (0.69-0.77)
Measured $\dot{V}O_{2max}$	2	1.00 (ref.)	0.58 (0.36-0.92)	0.23 (0.11-0.50)	0.20 (0.06-0.70)	0.43 (0.31-0.61)	0.80 (0.74-0.86)
Total BLSA, All BLSA Predictors							
xgboost, Tree	1	1.00 (ref.)	0.42 (0.26-0.68)	0.15 (0.08-0.30)	0.03 (0.01-0.14)	0.29 (0.22-0.39)	0.74 (0.70-0.78)
xgboost, Tree	2	1.00 (ref.)	0.60 (0.37-0.98)	0.30 (0.14-0.61)	0.16 (0.03-0.73)	0.39 (0.27-0.57)	0.80 (0.74-0.86)
xgboost, Linear	1	1.00 (ref.)	0.38 (0.23-0.62)	0.16 (0.08-0.31)	0.13 (0.06-0.29)	0.36 (0.28-0.45)	0.70 (0.64-0.76)
xgboost, Linear*	2	1.00 (ref.)	0.64 (0.38-1.09)	0.50 (0.23-1.08)	0.84 (0.32-2.24)	0.59 (0.42-0.83)	0.78 (0.72-0.84)
Random Forest	1	1.00 (ref.)	0.35 (0.19-0.64)	0.16 (0.07-0.36)	0.02 (0.00-0.17)	0.28 (0.20-0.40)	0.75 (0.69-0.81)
Random Forest	2	1.00 (ref.)	0.52 (0.27-0.98)	0.35 (0.15-0.85)	0.12 (0.01-0.96)	0.41 (0.26-0.64)	0.81 (0.75-0.87)
SVM	1	1.00 (ref.)	0.33 (0.18-0.62)	0.22 (0.10-0.45)	0.05 (0.01-0.21)	0.34 (0.25-0.46)	0.73 (0.67-0.79)
SVM	2	1.00 (ref.)	0.52 (0.27-0.99)	0.52 (0.23-1.17)	0.33 (0.07-1.63)	0.49 (0.32-0.75)	0.80 (0.74-0.86)
Total BLSA, OPACH Predictors							
xgboost, Tree	1	1.00 (ref.)	0.34 (0.21-0.56)	0.20 (0.11-0.36)	0.03 (0.01-0.14)	0.29 (0.22-0.38)	0.73 (0.69-0.77)
xgboost, Tree	2	1.00 (ref.)	0.49 (0.29-0.83)	0.40 (0.20-0.77)	0.19 (0.04-0.88)	0.42 (0.29-0.61)	0.80 (0.74-0.86)
xgboost, Linear	1	1.00 (ref.)	0.32 (0.19-0.53)	0.16 (0.09-0.31)	0.10 (0.04-0.24)	0.34 (0.26-0.44)	0.72 (0.66-0.78)
xgboost, Linear*	2	1.00 (ref.)	0.57 (0.33-0.98)	0.46 (0.22-0.97)	0.67 (0.23-1.94)	0.56 (0.39-0.79)	0.78 (0.72-0.84)
Random Forest	1	1.00 (ref.)	0.27 (0.15-0.48)	0.19 (0.09-0.37)	0.04 (0.01-0.16)	0.26 (0.18-0.36)	0.74 (0.68-0.80)
Random Forest	2	1.00 (ref.)	0.39 (0.21-0.73)	0.39 (0.18-0.82)	0.17 (0.04-0.78)	0.39 (0.26-0.60)	0.79 (0.73-0.85)
SVM	1	1.00 (ref.)	0.27 (0.14-0.53)	0.17 (0.08-0.37)	0.07 (0.02-0.24)	0.30 (0.21-0.41)	0.73 (0.67-0.79)
SVM*	2	1.00 (ref.)	0.52 (0.26-1.06)	0.46 (0.20-1.10)	0.60 (0.14-2.56)	0.46 (0.29-0.72)	0.80 (0.74-0.86)

Model 1 = $\dot{V}O_{2max}$ quartiles; crude

Model 2 = Model 1 + age + race and ethnicity + education

**P*-trend across quartiles not statistically significant at the 0.05 level. All *P*-values for 1 SD increase were statistically significant at the 0.05 level.

Supplemental Table 3.7. Hazard Ratios (HR) of All-Cause Mortality by Measured and Predicted $\dot{V}O_{2max}$ in the BLSA Women (n = 565)

Sample, Universe of Predictors	Model	Q1	Q2	Q3	Q4	HR for 1 SD Increase	*C*-Statistic
Measured $\dot{V}O_{2max}$	1	1.00 (ref.)	0.37 (0.18-0.74)	0.22 (0.09-0.54)	0.18 (0.07-0.47)	0.46 (0.33-0.64)	0.66 (0.58-0.74)
Measured $\dot{V}O_{2max}$*	2	1.00 (ref.)	0.51 (0.25-1.05)	0.51 (0.20-1.31)	0.63 (0.21-1.90)	0.69 (0.46-1.04)	0.77 (0.71-0.83)
Total BLSA, All BLSA Predictors							
xgboost, Tree	1	1.00 (ref.)	0.32 (0.16-0.66)	0.10 (0.03-0.32)	0.24 (0.11-0.56)	0.45 (0.32-0.63)	0.70 (0.62-0.78)
xgboost, Tree*	2	1.00 (ref.)	0.43 (0.21-0.90)	0.21 (0.06-0.71)	0.92 (0.34-2.53)	0.71 (0.47-1.08)	0.79 (0.73-0.85)
xgboost, Linear	1	1.00 (ref.)	0.49 (0.25-0.96)	0.27 (0.12-0.62)	0.20 (0.08-0.53)	0.49 (0.35-0.67)	0.65 (0.57-0.73)
xgboost, Linear*	2	1.00 (ref.)	0.64 (0.32-1.26)	0.66 (0.26-1.71)	1.16 (0.33-4.05)	0.78 (0.50-1.21)	0.78 (0.72-0.84)
Random Forest	1	1.00 (ref.)	0.57 (0.24-1.34)	0.22 (0.06-0.76)	0.36 (0.13-1.00)	0.59 (0.40-0.86)	0.63 (0.53-0.73)
Random Forest*	2	1.00 (ref.)	0.63 (0.26-1.51)	0.39 (0.10-1.45)	0.97 (0.28-3.34)	0.83 (0.52-1.34)	0.77 (0.67-0.87)
SVM	1	1.00 (ref.)	0.56 (0.24-1.30)	0.20 (0.06-0.68)	0.29 (0.10-0.86)	0.60 (0.41-0.88)	0.65 (0.55-0.75)
SVM*	2	1.00 (ref.)	0.62 (0.26-1.48)	0.28 (0.08-1.06)	0.65 (0.17-2.48)	0.87 (0.53-1.42)	0.77 (0.67-0.87)
Total BLSA, OPACH Predictors							
xgboost, Tree	1	1.00 (ref.)	0.32 (0.15-0.67)	0.25 (0.11-0.58)	0.22 (0.09-0.54)	0.47 (0.34-0.66)	0.65 (0.57-0.73)
xgboost, Tree*	2	1.00 (ref.)	0.41 (0.19-0.88)	0.52 (0.21-1.28)	1.11 (0.37-3.35)	0.78 (0.51-1.18)	0.79 (0.73-0.85)
xgboost, Linear	1	1.00 (ref.)	0.42 (0.21-0.86)	0.23 (0.10-0.56)	0.29 (0.13-0.67)	0.50 (0.37-0.69)	0.65 (0.57-0.73)
xgboost, Linear*	2	1.00 (ref.)	0.50 (0.25-1.03)	0.62 (0.23-1.70)	2.16 (0.66-7.07)	0.84 (0.53-1.34)	0.79 (0.73-0.85)
Random Forest	1	1.00 (ref.)	0.35 (0.16-0.77)	0.22 (0.09-0.58)	0.21 (0.08-0.55)	0.47 (0.33-0.67)	0.66 (0.56-0.76)
Random Forest*	2	1.00 (ref.)	0.42 (0.18-0.94)	0.48 (0.17-1.34)	0.82 (0.25-2.63)	0.77 (0.50-1.20)	0.77 (0.69-0.85)
SVM	1	1.00 (ref.)	1.05 (0.46-2.37)	0.36 (0.12-1.11)	0.39 (0.13-1.22)	0.65 (0.45-0.95)	0.60 (0.50-0.70)
SVM*	2	1.00 (ref.)	1.29 (0.55-3.03)	0.67 (0.19-2.40)	1.37 (0.32-5.86)	1.07 (0.62-1.84)	0.77 (0.69-0.85)

Model 1 = $\dot{V}O_{2max}$ quartiles; crude

Model 2 = Model 1 + age + race and ethnicity + education

**P*-trend across quartiles not statistically significant at the 0.05 level. All Model 1 *P*-values for 1 SD increase were statistically significant at the 0.05 level and none of the Model 2 *P*-values for 1 SD increase were statistically significant at the 0.05 level.

Chapter 3, in part, is currently being prepared for submission for publication of the material. Schumacher, Benjamin T.; LaMonte, Michael J.; LaCroix, Andrea Z.; Simonsick, Eleanor M.; Hooker, Steven P.; Parada, Humberto; Bellettiere, John; Kumar, Arun. The dissertation author was the primary investigator and author of this paper.

4. Associations of Relative Intensity of Physical Activity with Incident Cardiovascular Outcomes and Total Mortality

Benjamin T. Schumacher, Michael J. LaMonte, Chongzhi Di, Eleanor M. Simonsick, Humberto Parada, Steven P. Hooker, John Bellettiere, Andrea Z. LaCroix

4.1. Abstract

Background: Quantify and compare the associations of *relative* and *absolute* intensity of physical activity (PA) with total mortality and incident major cardiovascular disease (CVD).

Methods: Accelerometer-measured PA in the Objective Physical Activity and Cardiovascular Health (OPACH) Study was used to estimate daily hours spent in *absolute* light intensity and moderate-to-vigorous PA (MVPA). Accelerometer-estimated metabolic equivalents (METs) in each epoch were divided by two maximal MET capacity estimates. These percent maximal effort metrics were categorized and aggregated into daily hours of *relative* light intensity PA and MVPA. Cox proportional hazards models estimated the associations of a one-hour daily increase in *absolute* and *relative* PA with the two outcomes.

Results: Mean age was 78.5±6.7 years. On each PA measurement scale, an increase in either intensity category reduced the risk of both outcomes. A one-hour increase in *absolute* light intensity PA reduced the risks of both outcomes by 12%, and a one-hour increase in *absolute* MPVA reduced the risk of death and CVD by 45% and 27%, respectively. On the *relative* scale, light intensity PA was more strongly associated with both outcomes than MVPA. Increasing *absolute* MVPA was more strongly associated with the outcomes than increasing *relative* MVPA.

Conclusion: The PA intensity paradigm should keep shifting towards recommendation of more movement, regardless of intensity, and placing greater emphasis on *relative* light intensity

activities (between 37% and 46% of maximal effort) as modifiable behavioral targets that are more easily achieved, reduce risks of death and incident major CVD, and promote healthy aging.

4.2. Introduction

Across the life course, regular physical activity (PA) for all population subgroups is known to reduce the risk of developing new chronic diseases, slow the progression of prevalent chronic diseases, and promote a myriad of other health benefits.[1] A 2019 meta-analysis found an association between higher levels of accelerometer-measured total PA (regardless of intensity) and a substantially reduced risk for premature mortality in older adults (ages ≥ 65).[80] Similarly, higher levels of accelerometer-measured total PA and moderate-to-vigorous PA (MVPA) have been associated with a lower risk of developing cardiovascular disease (CVD).[81] Despite the benefits from engaging in regular PA, the proportion of older adults that meet the PA guidelines was only 28%, according to a 2016 study.[2]

Historically, PA has been prescribed to individuals and recommended to populations at a given intensity. For example, in 2018, the U.S. Department of Health and Human Services recommended that adults engage in 150 - 300 minutes a week of moderate intensity, or 75 - 150 minutes a week of vigorous intensity PA.[82] These recommendations are calibrated for a generally healthy, middle-aged adult, and do not consider one's cardiorespiratory fitness, physical function, general health status, or any other observed or unobserved phenotypes. Thus, especially for an older individual, the level of energy they expend while performing an activity (*absolute* intensity) may be discordant with their level of exertion relative to their maximal possible effort which expresses the very same amount of *absolute* effort as a percent of the individual's maximum possible energy expenditure (*relative* intensity).[1,83] This concept may be better understood through the following adapted example: for a younger adult capable of 12 metabolic equivalent (MET) activities, a slow walk (assume 3 METs) requires minimal effort relative to their maximum energy expenditure capacity (3/12 = 25% of their maximal MET capacity), but

for an older adult capable of 5 MET activities, this same walk (3 METs) requires a much greater relative effort (3/5 = 60%), compelling them to operate closer to their maximum capacity.[8] Though both adults are exerting the same amount of energy on an absolute scale, their exertions relative to their maximum capabilities are quite disparate.

The recognition that what constitutes "vigorous intensity activity" is not necessarily the same from one person to the next poses a challenge to the formulation of population-level physical activity guidelines, which have always been made based on absolute intensity. This led the Physical Activity Guidelines Advisory Committee (PAGAC) to call for further research that includes objective measures of relative intensity of PA.[19] As noted in the 2018 U.S. PAGAC Scientific Report, relative intensity activity can be quantified as the percentage of maximal oxygen uptake.[1] Maximal oxygen uptake, $\dot{V}O_{2max}$, is the gold standard measurement of cardiorespiratory fitness.[9] If the $\dot{V}O_{2max}$ of an individual is known and their instantaneous oxygen uptake ($\dot{V}O_2$) can be measured, (($\dot{V}O_2$ / $\dot{V}O_{2max}$)*100) gives the percent of their maximal exercise capacity (i.e., the *relative* intensity of their effort). As direct measurements of instantaneous $\dot{V}O_2$ and $\dot{V}O_{2max}$ require specialized equipment, trained personnel, the presence of a licensed physician (for $\dot{V}O_{2max}$), and extensive economic resources that are generally not feasible for large epidemiologic cohort studies, indirect estimations of relative intensity have been used, e.g. the talk test[12] and the Borg Rating of Perceived Exertion (RPE).[13] To the best of our knowledge, the present study is the first to outline an approach that leverages a previously published accelerometer-derived algorithm[84] for assessing instantaneous $\dot{V}O_2$ and two $\dot{V}O_{2max}$ prediction algorithms (one from Chapter 2.4 and one from Chapter 3.4) to estimate percent maximal effort. We sought to determine how estimated relative intensity of physical activity differs from absolute intensity in relation to risks of total mortality and incident cardiovascular

outcomes among older women followed for up to 9 years in the Objective Physical Activity and Cardiovascular Health (OPACH) Study.

4.3. Methods

Study Population

The analytic sample for the present study was derived from OPACH, an ancillary study of the Women's Health Initiative's (WHI) Long Life Study (LLS). Details about the design, recruitment, and measures have been previously published for OPACH[67] and WHI[85,86], and the description of the inclusion criteria for the present study closely emulates those of Schumacher et al.[87] In brief, all ambulatory women from the LLS (2012–2013; n = 7,875) were invited to concurrently enroll in OPACH.[67] Women were included in the present study if they returned their accelerometer (n after exclusion = 6,721), if their accelerometer had usable data (n = 6,489), if they wore their accelerometer for more than four or more adherent days (where a day with ≥10 hours of device wear while awake was considered adherent[67]; n = 6,126), and did not have prevalent cardiovascular disease at OPACH baseline, leaving 5,633 women in the analytic sample. All women provided informed consent either in writing or by telephone. The institutional review board at the Fred Hutchinson Cancer Research Center approved the study protocols for the LLS and OPACH, and the University of California, San Diego's institutional review board has approved subsequent OPACH data analysis.

Measures

Accelerometer-Measured Absolute Intensity

OPACH participants wore a GT3X+ triaxial accelerometer (ActiGraph LLC, Pensacola, FL) on their right hip, above the iliac crest, using a belt for 7 days, with the first day of

accelerometer wear serving as that participant's OPACH baseline.[67] The triaxial accelerometers measured acceleration at 30 Hz. Data were converted to 15-second epochs using the normal filter supplied with ActiLife (version 6; ActiGraph LLC). The Choi algorithm[88] was used to remove periods of accelerometer non-wear using vector magnitude acceleration counts with a 90-minute window, 30-minute stream frame, and 2-min tolerance.[88] Estimates of accelerometer-measured absolute intensity were categorized using cutpoints from the OPACH Calibration Study[89] based on each epoch's vector magnitude (VM) count. *Absolute* light intensity activity was defined as any epoch with VM counts of > 18 & ≤ 518 and *absolute* MVPA was defined as any epoch with VM > 518.

$\dot{V}O_{2max}$ Estimation

$\dot{V}O_{2max}$ was estimated using two prediction equations: one previously published ordinary least squares (OLS) model[44] that was recalibrated to measured $\dot{V}O_{2max}$ in the Baltimore Longitudinal Study of Aging (BLSA) (see Chapter 2.4.) and one machine-learned (ML) $\dot{V}O_{2max}$ prediction algorithm also derived in the BLSA (see Chapter 3.4.). A third algorithm, a Least Absolute Shrinkage and Selection Operator (LASSO) model, also developed in the BLSA, was selected based on its prediction performance in the BLSA, but when exported to OPACH yielded 38% of participants' predicted $\dot{V}O_{2max}$ values as missing due to missing covariate data. Results are not shown for the LASSO model in this study. Extensive details about the BLSA[48,49,90] and these prediction algorithms are published elsewhere (see Chapters 2 and 3). Briefly, the OLS model published by Baynard et al.[44] was recalibrated to measured $\dot{V}O_{2max}$ in the BLSA women and yielded a final regression equation of predicted $\dot{V}O_{2max}$ = 56.63 - 0.33*age - 0.54*body mass index (BMI) and then used to predict $\dot{V}O_{2max}$ in OPACH (root mean squared error (RMSE)

between predicted $\dot{V}O_{2max}$ and measured $\dot{V}O_{2max}$ in the BLSA = 3.6 mL•kg^{-1}•min^{-1}; Chapter 2.4.). An extreme gradient boosted (xgboost) algorithm using a tree-booster was used to develop a $\dot{V}O_{2max}$ prediction algorithm in the BLSA and was applied to OPACH (RMSE from BLSA 3.1 mL•kg^{-1}•min^{-1}; Chapter 3.4.).

The two $\dot{V}O_{2max}$ estimations in OPACH were converted to maximal MET capacity (predicted $\dot{V}O_{2max}$ / 3.0 mL•kg^{-1}•min^{-1}). After applying the Baynard equation to OPACH, 172 OPACH participants had implausible predicted maximal MET capacity levels (< 3 METs), largely due to very high body mass index levels which were not represented in the BLSA, and these women were excluded from analyses relying on the Baynard equation.

Outcome Ascertainment

Two outcomes were assessed in the present study: total mortality and physician adjudicated cases of incident stroke, myocardial infarction, or cardiovascular disease (CVD) related death (hereafter referred to as incident major CVD). Deaths and incident major CVD events were ascertained through March 31st, 2021 via annual mailed outcomes questionnaires, telephone follow-up, augmented by systematic reviews of the National Death Index[52], obituary notices, notification from the family of the decedent, and proxy queries.[67,91] Vital status classification was obtained for all OPACH participants in the analytic sample. Incident major CVD events were physician adjudicated following a review of participants' medical records.[67] There were 1,312 participant deaths from any cause and 748 incident major CVD cases during a median follow-up time of 7.4 years (range: 0.1 to 8.9 years).

Statistical Analysis

Analysis of variance (ANOVA) tests (continuous variables) and chi-square tests (categorical variables) were used to compare baseline covariates by quartiles of xgboost-predicted maximal MET capacity.

Correlations between all maximal MET estimations and selected covariates were assessed. To evaluate the transportability of these prediction algorithms to OPACH from BLSA, the same correlations from the BLSA cohort are also presented.

Cox proportional hazards models were fit to assess the associations between quartiles of the two estimates of predicted maximal MET capacity and the two outcomes. Model 1 was unadjusted, Model 2 adjusted for age, race and ethnicity, and education, and Model 3 adjusted for all Model 2 covariates and the RAND-36 physical function composite score.[92] Tests for linear trends across quartiles were conducted using an indicator variable for the quartile in the same model. Additionally, we assessed how a one standard deviation increase in each predicted maximal MET capacity (continuous) was associated with total mortality. Hazard ratios (HRs) and *P*-values of the mean-centered and scaled (i.e. a one standard deviation increase) maximal MET capacity variable for all models were obtained. Concordance statistics (*C*-Statistics), which give the proportion of participant pairs for which the model correctly predicts the participant in the pair who will experience an outcome event first, are also reported for each model.[77]

Next, we estimated *relative* intensity of activity. Using a previously published equation[84], VM counts in each 15-second epoch were used to estimate that epoch's energy expenditure expressed in METs (estimated METs = $1 + (0.09088*\sqrt{VM})$). The estimated MET value in each epoch was divided by both of the predicted maximal MET capacity estimates and each resulting measure of percent of maximal effort was categorized into relative intensity categories per the

American College of Sports Medicine[9]: ≥ 37% and < 46% of maximal MET capacity as *relative* light intensity activity and ≥ 46% of maximal MET capacity as *relative* MVPA. We then computed the total hours spent in each absolute and relative intensity category and divided each by the number of adherent accelerometer wear days to yield average daily hours in each category of absolute (based on VM counts alone) and relative intensity (based on the percent of maximal MET capacity). Correlations between time spent in absolute and relative intensity categories and selected covariates were assessed. Cox proportional hazards models were fit to estimate the HRs of a one-hour daily increase in each intensity category for both relative and absolute activity. Multiplicative effect measure modification tests were conducted to assess differences between all exposures and tertiles of xgboost-predicted maximal MET capacity and the Short Physical Performance Battery (SPPB) to assess whether the association between estimated relative intensity of PA and the two outcomes varied by the degree of the predicted maximal MET capacity or by lower extremity physical function. To test whether the proportional hazards assumption was violated, we used the cox.zph function in the R survival package.[60] The correlation of the scaled Schoenfeld residuals for each covariate (and for the whole model) with time were examined to ensure independence of residuals and time. No violations in the proportional hazards assumption were noted.

4.4. Results

Distribution of Maximal MET Capacity and Daily Time Spent in Each PA Category

In OPACH women, the median predicted maximal MET capacity for the xgboost and Baynard algorithms were 5.7 and 5.3 METs, respectively. These medians were slightly lower than the measured maximal MET capacity (6.2 METs) from the BLSA women in the same age range as those in OPACH (63 - 97). The Baynard-predicted maximal MET capacity had a wider

distribution of values than xgboost, ranging from 3.0 – 8.9 METs, though xgboost yielded a higher predicted maximal MET capacity than Baynard at 9.2 METs. Further details on the summary of both maximal MET capacity estimates can be found in Supplemental Table 4.5.

On the *absolute* intensity scale, OPACH women engaged in 4.8 and 0.8 hours per day (hrs/d; median) of accelerometer-measured light intensity and MVPA, respectively. On the *relative* intensity scale as defined by xgboost-predicted maximal MET capacity, these amounts were 1.3 and 1.7 hrs/d, respectively. Finally, on the Baynard-*relative* scale, these amounts were 1.3 and 2.2 hrs/d for *relative* light and *relative* MVPA, respectively.

Sample Characteristics

The 5,633 OPACH women had a mean age and BMI of 78.5 (SD = 6.7) years and 28.1 (SD = 5.7) kg/m^2, respectively. The participants were 49.2% were non-Hispanic White, 33.3% non-Hispanic Black, and 17.5% were Hispanic. Half of the participants (51.9%) reported "excellent or very good" health status, had moderate Short Physical Performance Battery scores (mean: 8.3 (SD = 2.5), and 2.4% were current smokers. Older age, non-Hispanic White race/ethnicity, higher BMI, lower SPPB, and lower quality of life rating were associated with having a lower xgboost-predicted maximal MET capacity at baseline (see Table 4.1.).

Associations Between Predicted Maximal MET Capacity and the Selected Outcomes

xgboost-Predicted Maximal MET Capacity

In unadjusted Cox proportional hazards models, xgboost-predicted maximal MET capacity was inversely associated with total mortality in both quartile and continuous forms ($P_{trend} < 0.01$; see Table 4.2.). After adjustment for Model 2 covariates, all three quartiles' HRs (in reference to the lowest quartile) remained below 1.0, statistically significant, and maintained

an observable trend across quartiles. After further adjustment for physical function (Model 3), all the HRs were above 1.0, all of the confidence intervals included 1.0, and no statistically significant trend across quartiles remained.

In unadjusted Cox proportional hazards models, xgboost-predicted maximal MET capacity was inversely associated with incident major CVD in both quartile and continuous forms ($P_{trend} < 0.01$; see Table 4.3.). After adjustment for Model 2 covariates, all three quartiles' HRs were attenuated, though the trend across quartiles remained ($P_{trend} < 0.01$). After further adjustment for physical function (Model 3), all the HRs were further attenuated with all three confidence intervals now including the null value and no trend across quartiles was evident ($P_{trend} = 0.80$).

Baynard-Predicted Maximal MET Capacity

In unadjusted Cox proportional hazards models, Baynard-predicted maximal MET capacity was inversely associated with total mortality in both quartile and continuous forms ($P_{trend} < 0.01$; see Table 4.2.). After adjustment for Model 2 covariates, all three quartiles' HRs (in reference to the lowest quartile) were above 1.00, and, after further adjustment for physical function (Model 3), all HRs again increased in strength and were positively associated with total mortality in both quartile and continuous forms ($P_{trend} < 0.01$; see Table 4.2.).

In unadjusted Cox proportional hazards models, Baynard-predicted maximal MET capacity was inversely associated with incident major CVD in both quartile and continuous forms ($P_{trend} < 0.01$; see Table 4.3.). After adjustment for Model 2 covariates, all three quartiles' HRs were attenuated and the trend across quartiles remained ($P_{trend} < 0.01$). After further adjustment for physical function (Model 3), all the HRs were further attenuated with all three

confidence intervals now including the null value and no trend across quartiles remained (P_{trend} = 0.80).

Correlations Among Predicted Maximal MET Capacity Estimates, Accelerometer-Measured Maximal MET Capacity, Daily Hours in Intensity Categories, and Selected Covariates

Both xgboost and Baynard-predicted maximal MET capacity estimates were strongly correlated with measured maximal MET capacity in the BLSA women in the same age range of the OPACH women (r = 0.87 and 0.63, respectively; see Table 4.6.). The correlations between xgboost and Baynard-predicted maximal MET capacity and age, BMI, and SPPB were of similar magnitude and direction in BLSA and OPACH.

Daily hours spent in light intensity PA on the xgboost-relative and Baynard-relative scales were quite strongly correlated with absolute light intensity PA (r = 0.88 and 0.87, respectively). Daily hours spent in MPVA on the xgboost-relative and Baynard-relative scales were moderately correlated with absolute MVPA (r = 0.47 and 0.35, respectively).

Associations of Absolute and Relative Intensity Activity with the Selected Outcomes

Accelerometer-Measured Absolute Intensity

When categorizing physical activity absolute intensity into light (VM counts of > 18 & ≤ 518 in each 15-second epoch) and MVPA (VM counts of > 518 in each 15-second epoch), we observed a 12% reduction in the risk of total mortality for every one-hour increase in daily light intensity activity, after adjusting for model 3 covaries (HR (95% CI): 0.88 (0.84-0.92); Table 4.4. Further, for every one-hour increase in daily MVPA, we observed a 45% reduction in risk of total mortality after adjusting for the same covariates (HR (95% CI): 0.55 (0.48-0.64). We observed a 12% reduction in the risk of incident major CVD for every one-hour increase in daily

light intensity activity, after adjusting for model 3 covaries (HR (95% CI): 0.88 (0.83-0.93). Further, for every one-hour increase in daily MVPA, we observed a 27% reduction in risk of incident major CVD after adjusting for the same covariates (HR (95% CI): 0.73 (0.61-0.87). In summary, *absolute* light intensity and *absolute* MVPA were significantly associated with reduced risk of total mortality and major CVD and higher levels of *absolute* MVPA carried a stronger reduction in risk than higher levels of *absolute* light intensity activity.

xgboost-Predicted Relative Intensity

When categorizing physical activity *relative* intensity into time spent in light (≥ 37% and < 46% of xgboost-predicted maximal MET capacity) and MVPA (≥ 46% of xgboost-predicted maximal MET capacity), we observed a 22% reduction in the risk of total mortality for every one-hour increase in daily *relative* light intensity activity, after adjusting for model 3 covaries (HR (95% CI): 0.78 (0.68-0.89); Table 4.4.). This was the strongest observed association between light intensity PA (across both absolute and relative) and total mortality. For every one-hour increase in daily *relative* MVPA, we observed an 18% reduction in risk of total mortality after adjusting for the same covariates (HR (95% CI): 0.82 (0.77-0.87)). The association between a one-hour increase in daily *relative* light intensity PA was somewhat more strongly associated with incident major CVD than total mortality (HR (95% CI): 0.70 (0.59-0.84)). The association between a one-hour increase in daily *relative* MVPA and incident major CVD was slightly weaker than the daily *relative* MVPA and total mortality association but remained in the direction of reduced risk and statistically significant (HR (95% CI): 0.89 (0.83-0.96); Table 4.4.). In summary, higher levels of time spent in *xgboost-relative* light intensity and *xgboost-relative* MVPA reduced risk of both total mortality and major CVD, and higher levels of *relative* light

intensity were associated with stronger reductions in risk than higher levels of *relative* MVPA for both outcomes.

Baynard-Predicted Relative Intensity

When categorizing physical activity *relative* intensity into light (≥ 37% and < 46% of Baynard-predicted maximal MET capacity) and MVPA (≥ 46% of Baynard-predicted maximal MET capacity), we observed a 21% reduction in the risk of total mortality for every one-hour increase in daily *relative* light intensity activity, after adjusting for model 3 covaries (HR (95% CI): 0.79 (0.68-0.92); Table 4.4.). Further, for every one-hour increase in daily *relative* MVPA, we observed a 22% reduction in risk of total mortality after adjusting for the same covariates (HR (95% CI): 0.78 (0.73-0.82)). Unlike the xgboost findings that showed stronger associations of *relative* light compared to *relative* MVPA associations with total mortality, the Baynard-predicted *relative* intensity daily light and MVPA associations had nearly equivalent magnitudes of association with total mortality. However, and similar to xgboost, the association between a one-hour increase in daily *relative* light intensity PA was more strongly associated with incident major CVD than total mortality (HR (95% CI): 0.77 (0.63-0.93)). The association between a one-hour increase in daily *relative* MVPA and incident major CVD was weaker than the daily *relative* MVPA and total mortality association but remained statistically significant (HR (95% CI): 0.88 (0.82-0.95); Table 4.4.). In summary, both Baynard-*relative* light intensity and Baynard-*relative* MVPA were associated with reduced risks of both total mortality and major CVD, the magnitudes of the association between Baynard-*relative* light intensity and Baynard-*relative* MVPA with total mortality were nearly equivalent, and the association between higher levels of *relative* light intensity was stronger than for *relative* MVPA with respect to incident major CVD.

Effect Modification

Accelerometer-Measured Absolute Intensity

We observed no statistically significant effect modification at the $P = 0.10$ level of *absolute* intensity on both outcomes by tertiles of xgboost-predicted maximal MET capacity (Supplemental Table 4.8.) or by SPPB categories (Supplemental Table 4.9.). In all strata, the associations between *absolute* PA and the outcomes were the same as in the total sample, except for the highest SPPB category, where *absolute* MVPA was not statistically associated with incident major CVD.

xgboost-Predicted Relative Intensity

We found statistically significant effect modification ($P < 0.10$) by tertile of xgboost-predicted maximal MET capacity in the association between xgboost-predicted *relative* MVPA and total mortality (Supplemental Table 4.8.) and no statistically significant effect modification by category of SPPB scores (Supplemental Table 4.9.). Specifically, in tertiles 2 and 3 of xgboost-predicted maximal MET capacity, the associations between xgboost-predicted *relative* MVPA and total mortality were stronger than those of xgboost-predicted *relative* light intensity, unlike the associations observed in tertile 1 and in the total sample, where xgboost-predicted *relative* light intensity was more strongly associated with total mortality than xgboost-predicted *relative* MVPA. In each tertile of xgboost-predicted maximal MET capacity, xgboost-predicted *relative* light intensity was more strongly associated with incident major CVD than xgboost-predicted *relative* MVPA. When stratifying by category of SPPB scores, xgboost-predicted *relative* MVPA was more strongly associated with total mortality than xgboost-predicted *relative*

light intensity, but xgboost-predicted *relative* light intensity was more strongly associated with incident major CVD than xgboost-predicted *relative* MVPA.

Baynard-Predicted Relative Intensity

We observed no statistically significant effect modification at the $P = 0.10$ level of Baynard-predicted *relative* MVPA on both outcomes by tertiles of xgboost-predicted maximal MET capacity (Supplemental Table 4.8.) or by SPPB categories (Supplemental Table 4.9.). Though not statistically different, in tertiles 2 and 3 of xgboost-predicted maximal MET capacity, Baynard-predicted *relative* MVPA was more strongly associated with total mortality than Baynard-predicted *relative* light intensity. In tertiles 1 and 2, Baynard-predicted *relative* light intensity was more strongly associated with incident major CVD than Baynard-predicted *relative* MVPA. When stratified by categories of SPPB scores, patterns in these HRs become obfuscated by statistically insignificant HRs.

4.5. Discussion

In the present study, we sought to quantify the association of relative intensity of physical activity—as scored by percent maximal effort—with incident major CVD and total mortality in a prospective study of ambulatory, community-dwelling older women. We found that: (1) on the *absolute* scale, increases in daily MVPA were more strongly associated with total mortality and major CVD than increases in light intensity PA in adjusted models, (2) on the *relative* scale, light intensity PA was more strongly associated with total mortality and major CVD than MVPA in adjusted models (with the exception of Baynard-predicted *relative* intensity where the HRs were nearly equivalent), (3) a one-hour increase in *relative* light intensity PA was more strongly associated with both mortality and incident major CVD than a one-hour increase in *absolute* light

intensity PA, and (4) a one-hour increase in *absolute* MVPA was more strongly associated with mortality and CVD than a one-hour increase in *relative* MVPA.

We do not have the ability to assess how well the xgboost and Baynard algorithms performed in their predictions of maximal MET capacity in OPACH. However, given that the strength and direction of the correlations between these predicted maximal MET capacity variables and the selected covariates were similar between OPACH and BLSA, coupled with the reasonable distributions of both variables (the excluded < 3 MET OPACH participants notwithstanding), and their ability to predict total mortality and incident major CVD, we have confidence in their estimations.

Few studies of relative vs. absolute intensity of physical activity exist with which to compare these results. In a prospective cohort study of 7,337 men in the Harvard Alumni Health Study (mean age: 66 years), participants rated their usual level of exertion when exercising on a 10-point Borg Scale, categorized as 0 to 2 ("nothing to weak"), 3 ("moderate"), 4 ("somewhat strong"), and ≥ 5 ("strong to maximal"). Adjusted relative risks (RR (95% CI)) of coronary heart disease (CHD) for men reporting usual perceived exertion as "moderate," "somewhat strong," and "strong to maximal" were: 0.86 (0.66-1.13), 0.69 (0.51-0.94), and 0.72 (0.52-1.00), respectively ($P_{trend} = 0.02$), when compared with "nothing to weak".[15] Despite the differences between Lee et al.[15] and the present study in measurement of relative intensity, outcome(s) being assessed, and gender of the study populations, the highest category of relative intensity activity did not provide the strongest reduction in risk in either study.

Findings from the present study *do not* support the assertion that one must achieve 6 METs on the *absolute* intensity scale to gain protection against total mortality and incident

CVD.[93] Our results suggest that reductions in risk of total mortality and major CVD are stronger for every additional hour per day engaged in activities that require between 37% and 46% maximal effort (*relative* light intensity PA) than engaging in activities that require ≥ 46% maximal effort (*relative* MVPA). These findings also align with a recent study from the OPACH cohort, the same cohort from which women in the present study were drawn, that found that higher amounts of activities in daily life (specifically, "daily life movement", e.g. performing housework or gardening) were independently associated with a lower risk of cardiovascular disease.[94]

Given the strong, graded associations of predicted maximal MET capacity (the denominator in the percent maximal effort computation) with death and CVD and the strong, graded associations of accelerometer-measured *absolute* intensity of activity (the numerator in the computation) with the two outcomes, it appears peculiar that the associations of *relative* light intensity PA were stronger than the associations of *relative* MVPA. A likely explanation is that the participants on the lower end of the predicted maximal MET capacity spectrum, e.g. participants with a predicted maximal MET capacity of 3 METs can reach MVPA on the *relative* scale by engaging in minimal activity (i.e. getting out of bed). However, those same participants have increased risk of mortality and CVD associated with their low estimated fitness levels as shown in Tables 2a and 2b. Thus, on the relative scale, the classification of time spent in MVPA includes participants with a much wider range of fitness levels, than when MVPA is classified on the absolute scale. Our stratified analyses provide some basis for this hypothesis as we did observe that xgboost-predicted *relative* MVPA was more strongly associated with mortality than xgboost-predicted *relative* light intensity in higher tertiles (2 and 3) of predicted maximal METs (Supplemental Table 4.8.).

Despite the benefits from engaging in regular PA, the proportion of older adults in the U.S. that, according to self-report, met the PA guidelines was only 28% according to a 2016 study.[2] Talbot et al[16] asserts that a higher proportion of older adults meet national PA recommendations on a *relative* intensity activity scale than on an *absolute* intensity activity scale. Because the energy costs of movement increase with age and aerobic fitness levels decline with age, these authors assert that absolute intensity is inappropriate to measure and motivate older adults' PA. Further, they note that the proportion of older adults meeting the national recommendations for moderate and high intensity PA on an *absolute* intensity scale decreases with age, but the proportion meeting guidelines when activity is assessed on a *relative* intensity scale increases with age. Our findings support this, as we observed strong correlations between *absolute* light PA and the two *relative* light PA estimations (both r values = 0.88), but weak correlations between *absolute* MVPA and the two *relative* MVPA estimations (xgboost-MVPA r = 0.47, and Baynard-MVPA r = 0.35). It is evident that measuring PA on the *relative* scale in older adults credits them with more MVPA than when measuring PA on the *absolute* scale.

There are some limitations to the present study. Harmonization of variables between BLSA and OPACH and missingness in both cohorts could have introduced error and decreased the precision with which maximal MET capacity was estimated. Also, the accuracy of the prediction equations for maximal MET capacity could be affected by differences between the BLSA cohort and OPACH or other study populations. In the present study, the Baynard maximal MET capacity equation that was calibrated in the BLSA cohort systematically underestimated predicted maximal MET capacity in OPACH compared to the xgboost estimates. This likely occurred because OPACH has women with higher BMI values than their BLSA counterparts and the Baynard estimate relied heavily on BMI. However, both estimates had distributions in

OPACH similar to the BLSA, similar magnitudes of correlations with age, physical performance, and BMI, and predictive validity for mortality and CVD until adjusted for physical function. These findings all support the external validity of the BLSA-derived predicted maximal capacity measures when applied to an external cohort, in this case OPACH. Importantly, the methods and results of the present study provide an innovative and rigorous approach for examining relative intensity of PA in large population-based studies where direct measurement of maximal MET capacity and accelerometer-measured absolute intensity in the same cohort is quite rare.

There are many strengths to this study. First, we directly respond to PAGAC's call to estimate percent maximal effort and its association with health outcomes. Next, the WHI OPACH study population is a well-characterized, diverse cohort with high-quality prospective follow-up, accelerometer-measured *absolute* PA in 5,633 older women, and sufficient data to support two estimations of maximal MET capacity. Finally, the results of this study provide a generally consistent narrative across both estimates of relative intensity. Future studies should seek to replicate these findings and will be especially informative if accelerometer-measured PA and maximal exercise capacity testing is present in the same cohort.

In conclusion, these findings show that: (1) on the *absolute* scale, increases in daily MVPA were more strongly associated with mortality and major CVD than increases in light intensity activity; however, (2) on the *relative* scale, light intensity activity was more strongly associated with mortality and major CVD than MVPA, (3) a one-hour increase in *relative* light intensity was more strongly associated with mortality and CVD than a one-hour increase in *absolute* light intensity, and (4) a one-hour increase in *absolute* MVPA was more strongly associated with both outcomes than a one-hour increase in *relative* MVPA. These findings extend the paradigm shift towards recommendation of more movement, regardless of intensity,

and placing greater emphasis on *relative* light intensity activities (between 37% and 46% of one's maximal effort) as modifiable behavioral targets that are more easily achieved in older adults to reduce risks of death and CVD and improve prospects for healthy aging.

Table 4.1. Participant Characteristics by Quartiles of xgboost-Predicted Maximal MET Capacity in OPACH (n = 5,633)

Characteristic†	Total Range (3.5 - 9.2) (n = 5,959)	Quartile 1 < 5.1 METs (n = 1,408)	Quartile 2 ≥ 5.1 & ≤ 5.7 METs (n = 1,409)	Quartile 3 > 5.7 & ≤ 6.3 METs (n = 1,408)	Quartile 4 > 6.3 METs (n = 1,408)	*P*-value*
Age, Mean (SD)	78.5 (6.7)	80.8 (6.0)	79.1 (6.5)	78.1 (6.6)	76.0 (6.7)	**<0.01**
Race/Ethnicity, n (%)						**<0.01**
White, N.H.	2772 (49.2)	891 (63.3)	738 (52.4)	634 (45.0)	509 (36.2)	
Black, N.H.	1876 (33.3)	369 (26.2)	457 (32.4)	494 (35.1)	556 (39.5)	
Hispanic	985 (17.5)	148 (10.5)	214 (15.2)	280 (19.9)	343 (24.4)	
BMI, Mean (SD)	28.1 (5.7)	32.5 (5.8)	28.8 (5.2)	26.2 (4.1)	23.8 (3.1)	**<0.01**
General Health Status, n (%)						**<0.01**
Excellent or Very Good	2924 (51.9)	461 (32.7)	682 (48.4)	850 (60.4)	931 (66.1)	
Good	2190 (38.9)	712 (50.6)	574 (40.7)	483 (34.3)	421 (29.9)	
Fair/Poor	499 (8.9)	227 (16.1)	146 (10.4)	73 (5.2)	53 (3.8)	
SPPB, Mean (SD)	8.3 (2.5)	6.9 (2.4)	8.1 (2.3)	8.9 (2.2)	9.7 (2.2)	**<0.01**
RAND36 Physical Functioning, Mean (SD)	70.0 (25.5)	48.7 (23.8)	68.9 (23.3)	79.0 (19.2)	83.3 (20.1)	**<0.01**
Quality of Life Rating, Mean (SD)	7.9 (1.5)	7.3 (1.6)	7.8 (1.5)	8.2 (1.4)	8.3 (1.4)	**<0.01**
Body Pain in Past Month, n (%)						**<0.01**
Mild/Mod/Severe	2905 (51.6)	928 (65.9)	764 (54.2)	642 (45.6)	571 (40.6)	
None/Very mild	2235 (39.7)	336 (23.9)	523 (37.1)	659 (46.8)	717 (50.9)	

Table 4.1. (cont'd) Participant Characteristics by Quartiles of xgboost-Predicted Maximal MET Capacity in OPACH (n = 5,633)

Characteristic†	Total Range (3.5 - 9.2) (n = 5,959)	Quartile 1 < 5.1 METs (n = 1,408)	Quartile 2 ≥ 5.1 & ≤ 5.7 METs (n = 1,409)	Quartile 3 > 5.7 & ≤ 6.3 METs (n = 1,408)	Quartile 4 > 6.3 METs (n = 1,408)	*P*-value*
MET-mins/week from all exercise activities, Mean (SD)	1794.9 (1617.3)	1184.0 (1138.5)	1611.5 (1465.1)	1927.7 (1524.3)	2443.2 (1957.3)	**<0.01**
Alcohol Frequency in Past 3mos, n (%)						**<0.01**
Never	1880 (33.4)	554 (39.3)	478 (33.9)	465 (33.0)	383 (27.2)	
Less than 1 per week	1782 (31.6)	451 (32.0)	457 (32.4)	443 (31.5)	431 (30.6)	
1 or more drinks per week	1487 (26.4)	256 (18.2)	353 (25.1)	400 (28.4)	478 (33.9)	
Current Smoker, n (%)	137 (2.4)	28 (2.0)	33 (2.3)	40 (2.8)	36 (2.6)	0.60

*P-value for continuous variables from the One-way ANOVA and Chi-Sq. goodness of fit test for categorical variables across tertiles. Bold indicates significance at the P < 0.05 level.

Table 4.2. Hazard Ratios (HR) of Total Mortality by Predicted Maximal MET Capacity in OPACH (n = 5,633)

Max. MET Capacity	Q1	Q2	Q3	Q4	HR for 1 SD Increase	*C*-Statistic (Std. Error)
Baynard et al	< 4.58	≥ 4.58 & ≤ 5.28	> 5.28 & ≤ 5.97	> 5.97		
n	1278	1277	1277	1278		
Deaths	399	357	281	162		
Person Years	8676.0	8864.8	9032.4	9344.6		
Rate per 1,000	46.0	40.3	31.1	17.3		
Model 1	**1.00 (ref.)**	**0.87 (0.75-1.00)**	**0.66 (0.57-0.77)**	**0.36 (0.30-0.43)**	**0.69 (0.65-0.73)**	0.59 (0.01)
Model 2	1.00 (ref.)	1.03 (0.89-1.19)	1.06 (0.91-1.25)	1.12 (0.91-1.37)	1.01 (0.94-1.09)	0.71 (0.01)
Model 3	**1.00 (ref.)**	**1.23 (1.06-1.43)**	**1.41 (1.19-1.66)**	**1.56 (1.26-1.94)**	**1.17 (1.08-1.26)**	0.73 (0.01)
xgboost	< 5.08	≥ 5.08 & ≤ 5.66	> 5.66 & ≤ 6.27	> 6.27		
n	1408	1409	1408	1408		
Deaths	475	339	295	203		
Person Years	9461.0	9837.7	10056.8	9881.3		
Rate per 1,000	50.2	34.5	29.3	20.5		
Model 1	**1.00 (ref.)**	**0.68 (0.59-0.78)**	**0.56 (0.49-0.65)**	**0.41 (0.34-0.48)**	**0.67 (0.63-0.71)**	0.59 (0.01)
Model 2	**1.00 (ref.)**	**0.79 (0.69-0.91)**	**0.77 (0.66-0.89)**	**0.70 (0.59-0.83)**	**0.84 (0.79-0.90)**	0.72 (0.01)
Model 3	1.00 (ref.)	1.02 (0.88-1.19)	1.15 (0.98-1.36)	1.11 (0.92-1.33)	1.03 (0.96-1.10)	0.74 (0.01)

Model 1 = Quartiles of Predicted Max. METs; crude

Model 2 = Model 1 + age + race and ethnicity + education

Model 3 = Model 2 + RAND-36 Physical Function Score

Bold represents statistical significance at the 0.05 level.

Table 4.3. Hazard Ratios (HR) of Incident Major CVD by Predicted Maximal MET Capacity in OPACH (n = 5,633)

VO_{2max}	Q1	Q2	Q3	Q4	HR for 1 SD Increase	*C*-Statistic (Std. Error)
Baynard et al	< 4.58	≥ 4.58 & ≤ 5.28	> 5.28 & ≤ 5.97	> 5.97		
n	1278	1277	1277	1278		
CVD Events	251	182	145	94		
Person Years	8298.5	8537.1	8796.3	9154.2		
Rate per 1,000	30.2	21.3	16.5	10.3		
Model 1	**1.00 (ref.)**	**0.70 (0.58-0.85)**	**0.54 (0.44-0.66)**	**0.34 (0.27-0.43)**	**0.68 (0.63-0.73)**	0.61 (0.01)
Model 2	1.00 (ref.)	0.81 (0.67-0.98)	0.79 (0.64-0.98)	0.85 (0.65-1.11)	**0.91 (0.83-1.00)**	0.70 (0.01)
Model 3	1.00 (ref.)	0.93 (0.76-1.13)	0.96 (0.77-1.20)	1.06 (0.80-1.41)	1.01 (0.92-1.12)	0.71 (0.01)
xgboost	< 5.08	≥ 5.08 & ≤ 5.66	> 5.66 & ≤ 6.27	> 6.27		
n	1408	1409	1408	1408		
CVD Events	273	176	166	133		
Person Years	9039.8	9517.5	9773.8	9684.6		
Rate per 1,000	30.2	18.5	17.0	13.7		
Model 1	**1.00 (ref.)**	**0.61 (0.50-0.74)**	**0.56 (0.46-0.68)**	**0.45 (0.37-0.55)**	0.71 (0.65-0.76)	0.58 (0.01)
Model 2	**1.00 (ref.)**	**0.69 (0.57-0.84)**	**0.73 (0.60-0.89)**	**0.73 (0.59-0.90)**	0.86 (0.79-0.94)	0.70 (0.01)
Model 3	1.00 (ref.)	0.83 (0.68-1.01)	0.98 (0.79-1.22)	1.00 (0.79-1.27)	0.99 (0.91-1.09)	0.71 (0.01)

Model 1 = Quartiles of Predicted Max. METs; crude

Model 2 = Model 1 + age + race and ethnicity + education

Model 3 = Model 2 + RAND-36 Physical Function Score

Bold represents statistical significance at the 0.05 level.

Table 4.4. Hazard Ratios for Absolute and Relative Intensity Activity Categories and the Selected Outcomes in OPACH (n = 5,633)

	Total Mortality					
	Absolute		Relative: xgboost		Relative: Baynard et al	
Model	Light	MVPA	Light	MVPA	Light	MVPA
1	0.73 (0.70-0.77)	0.27 (0.23-0.31)	0.51 (0.45-0.58)	0.86 (0.81-0.91)	0.49 (0.43-0.57)	0.93 (0.88-0.97)
2	0.82 (0.78-0.85)	0.45 (0.39-0.51)	0.64 (0.56-0.72)	0.85 (0.80-0.90)	0.67 (0.58-0.77)	0.76 (0.72-0.80)
3	0.88 (0.84-0.92)	0.55 (0.48-0.64)	0.78 (0.68-0.89)	0.82 (0.77-0.87)	0.79 (0.68-0.92)	0.78 (0.73-0.82)
	Incident Major CVD					
	Absolute		Relative: xgboost		Relative: Baynard et al	
Model	Light	MVPA	Light	MVPA	Light	MVPA
1	0.75 (0.71-0.80)	0.40 (0.34-0.46)	0.49 (0.41-0.58)	0.92 (0.86-0.99)	0.51 (0.42-0.61)	1.01 (0.95-1.08)
2	0.82 (0.78-0.87)	0.61 (0.52-0.72)	0.60 (0.50-0.71)	0.91 (0.84-0.98)	0.66 (0.54-0.80)	0.86 (0.81-0.93)
3	0.88 (0.83-0.93)	0.73 (0.61-0.87)	0.70 (0.59-0.84)	0.89 (0.83-0.96)	0.77 (0.63-0.93)	0.88 (0.82-0.95)

Model 1 = crude

Model 2 = Model 1 + age + race/ethnicity + education

Model 3 = Model 2 + RAND-36 Physical Function Score

Supplemental Table 4.5. Summary of Predicted Maximum METs in OPACH (n = 5,633)

Source	Minimum	Quartile 1	Median	Quartile 3	Maximum	Missing Due to Covariate Missingness	Missingness Due to Implausible Prediction (< 3 METs)
Measured, BLSA*	2.1	5.1	6.2	7.2	10.4	-	-
xgboost	3.5	5.1	5.7	6.3	9.2	0	0
Baynard	3.0	4.6	5.3	6.0	8.9	351	172

*Measured maximal MET capacity (VO_{2max} / 3.0) in BLSA women in the same age range as OPACH women (ages 63 - 97)

Supplemental Table 4.6. Correlation Matrix of Maximal MET Capacity Estimates and Selected Covariates in OPACH (BLSA*)

	Measured Max. METs	Max. METs xgboost	Max. METs Baynard	Age	BMI	SPPB	RAND-36 PF†
Measured Max. METs	- (1.00)	- (.87)	- (.63)	- (-.35)	- (-.41)	- (.34)	- (-)
Max. METs xgboost		1.00	.71 (.77)	-.31 (-.42)	-.51 (-.52)	.44 (.38)	.51 (-)
Max. METs Baynard			1.00	-.49 (-.54)	-.68 (-.67)	.17 (.27)	.30 (-)
Age				1.00	-.30 (-.25)	-.17 (-.27)	-.26 (-)
BMI					1.00	-.04 (-.07)	-.11 (-)
SPPB						1.00	.28 (-)
RAND-36 PF†							1.00

*BLSA women in the same age range as OPACH women (ages 63 - 97)

Correlations are for OPACH with the corresponding BLSA correlations in parentheses.

†RAND-36 physical function composite score (0 - 100)

Supplemental Table 4.7. Correlation Matrix of Daily Hours in PA Intensity Categories and Selected Covariates in OPACH

	Absolute : Light	xgboost: Light	Baynard : Light	Absolute : MVPA	xgboost: MVPA	Baynard : MVPA	Age	BMI	SPPB	RAND-36 PF†
Absolute: Light	1.00	0.88	0.87	0.34	0.25	0.20	-0.10	-0.23	0.12	0.18
xgboost: Light		1.00	0.91	0.35	0.14	0.13	-0.08	-0.29	0.15	0.21
Baynard: Light			1.00	0.35	0.17	0.08	-0.12	-0.25	0.12	0.18
Absolute: MVPA				1.00	0.47	0.35	-0.32	-0.09	0.25	0.31
xgboost: MVPA					1.00	0.71	0.05	0.32	-0.18	-0.17
Baynard: MVPA						1.00	0.27	0.52	0.02	-0.05
Age							1.00	-0.30	-0.17	-0.27
BMI								1.00	-0.04	-0.11
SPPB									1.00	0.28
RAND-36 PF†										1.00

†RAND-36 physical function composite score (0 - 100)

Supplemental Table 4.8. Hazard Ratios for Relative Intensity Activity Categories and the Selected Outcomes in OPACH Stratified by Tertiles of xgboost-Predicted Maximal MET Capacity (n = 5,633)

		Total Mortality					
		Absolute		xgboost		Baynard et al	
Tertile	Model	Light	MVPA	Light	MVPA	Light	MVPA
1	1	0.77 (0.72-0.83)	0.35 (0.28-0.44)	0.57 (0.45-0.71)	0.73 (0.66-0.80)	0.52 (0.41-0.67)	0.76 (0.71-0.82)
	2	0.82 (0.76-0.88)	0.50 (0.40-0.63)	0.66 (0.52-0.83)	0.77 (0.70-0.85)	0.63 (0.49-0.80)	0.72 (0.67-0.78)
	3	0.87 (0.81-0.93)	0.62 (0.49-0.79)	0.74 (0.58-0.94)	0.81 (0.74-0.89)*	0.70 (0.54-0.89)	0.76 (0.70-0.82)
2	1	0.79 (0.73-0.86)	0.26 (0.21-0.34)	0.64 (0.50-0.81)	0.48 (0.42-0.55)	0.62 (0.48-0.79)	0.78 (0.71-0.85)
	2	0.85 (0.78-0.92)	0.41 (0.32-0.53)	0.70 (0.55-0.89)	0.60 (0.53-0.69)	0.76 (0.59-0.96)	0.67 (0.61-0.74)
	3	0.88 (0.81-0.95)	0.46 (0.36-0.59)	0.79 (0.62-1.00)	0.64 (0.56-0.73)*	0.82 (0.64-1.05)	0.71 (0.64-0.78)
3	1	0.75 (0.68-0.83)	0.28 (0.22-0.37)	0.56 (0.44-0.70)	0.55 (0.45-0.66)	0.65 (0.48-0.87)	1.05 (0.93-1.19)
	2	0.82 (0.75-0.90)	0.50 (0.39-0.64)	0.66 (0.53-0.82)	0.65 (0.54-0.79)	0.78 (0.58-1.05)	0.72 (0.63-0.84)
	3	0.88 (0.80-0.97)	0.62 (0.48-0.81)	0.82 (0.65-1.03)	0.77 (0.64-0.94)*	0.86 (0.64-1.16)	0.79 (0.69-0.92)

Supplemental Table 4.8. (cont'd) Hazard Ratios for Relative Intensity Activity Categories and the Selected Outcomes in OPACH Stratified by Tertiles of xgboost-Predicted Maximal MET Capacity (n = 5,633)

		Incident Major CVD					
		Absolute		xgboost		Baynard et al	
Tertile	Model	Light	MVPA	Light	MVPA	Light	MVPA
1	1	0.77 (0.71-0.85)	0.49 (0.37-0.65)	0.52 (0.38-0.71)	0.81 (0.72-0.92)	0.50 (0.36-0.68)	0.91 (0.82-1.00)
	2	0.82 (0.75-0.90)	0.68 (0.51-0.90)	0.60 (0.44-0.82)	0.85 (0.75-0.95)	0.61 (0.44-0.85)	0.88 (0.79-0.97)
	3	0.88 (0.80-0.96)	0.84 (0.63-1.12)	0.70 (0.51-0.96)	0.90 (0.80-1.02)	0.70 (0.50-0.97)	0.93 (0.84-1.02)
2	1	0.77 (0.69-0.86)	0.42 (0.31-0.57)	0.51 (0.36-0.71)	0.59 (0.49-0.70)	0.53 (0.38-0.75)	0.82 (0.73-0.93)
	2	0.82 (0.73-0.91)	0.61 (0.45-0.83)	0.55 (0.40-0.77)	0.71 (0.59-0.85)	0.62 (0.44-0.87)	0.74 (0.65-0.84)
	3	0.85 (0.76-0.95)	0.67 (0.49-0.92)	0.62 (0.44-0.87)	0.75 (0.63-0.90)	0.67 (0.48-0.95)	0.77 (0.68-0.88)
3	1	0.81 (0.72-0.91)	0.40 (0.30-0.53)	0.60 (0.45-0.79)	0.67 (0.54-0.83)	0.77 (0.54-1.09)	1.15 (0.99-1.34)
	2	0.87 (0.78-0.98)	0.60 (0.45-0.80)	0.69 (0.53-0.91)	0.77 (0.62-0.96)	0.89 (0.62-1.27)	0.89 (0.75-1.05)
	3	0.91 (0.80-1.02)	0.66 (0.48-0.89)	0.76 (0.57-1.01)	0.83 (0.66-1.04)	0.93 (0.65-1.33)	0.92 (0.77-1.09)

Asterisk indicates a significant interaction at the $P < 0.10$ level.

Model 1 = crude

Model 2 = Model 1 + age + race and ethnicity + education

Model 3 = Model 2 + RAND-36 Physical Function Score

Supplemental Table 4.9. Hazard Ratios for Relative Intensity Activity Categories and the Selected Outcomes in OPACH Stratified by SPPB Category (n = 5,633)

		Total Mortality					
		Absolute		xgboost		Baynard et al	
Category	Model	Light	MVPA	Light	MVPA	Light	MVPA
Low (0 - 6)	1	0.79 (0.73-0.85)	0.30 (0.23-0.40)	0.56 (0.44-0.72)	0.75 (0.68-0.83)	0.52 (0.40-0.68)	0.83 (0.76-0.91)
	2	0.85 (0.78-0.92)	0.45 (0.34-0.60)	0.70 (0.54-0.90)	0.78 (0.70-0.86)	0.64 (0.49-0.83)	0.75 (0.68-0.83)
	3	0.90 (0.83-0.98)	0.54 (0.40-0.73)	0.81 (0.63-1.06)	0.79 (0.71-0.88)	0.71 (0.54-0.93)	0.77 (0.69-0.85)
Medium (7 - 9)	1	0.81 (0.75-0.87)	0.32 (0.26-0.41)	0.65 (0.52-0.82)	0.82 (0.74-0.91)	0.64 (0.51-0.81)	0.97 (0.89-1.05)
	2	0.87 (0.80-0.94)	0.50 (0.40-0.63)	0.75 (0.60-0.94)	0.82 (0.74-0.91)	0.80 (0.64-1.01)	0.80 (0.73-0.87)
	3	0.89 (0.83-0.97)	0.55 (0.44-0.70)	0.83 (0.66-1.05)	0.79 (0.71-0.87)	0.88 (0.69-1.12)	0.80 (0.73-0.87)
High (10 -12)	1	0.82 (0.74-0.91)	0.43 (0.33-0.55)	0.67 (0.50-0.90)	0.90 (0.78-1.04)	0.65 (0.48-0.88)	1.03 (0.92-1.14)
	2	0.90 (0.81-0.99)	0.64 (0.49-0.83)	0.81 (0.61-1.08)	0.86 (0.74-1.01)	0.88 (0.65-1.19)	0.80 (0.70-0.90)
	3	0.93 (0.84-1.03)	0.70 (0.53-0.92)	0.93 (0.70-1.24)	0.80 (0.68-0.93)	0.99 (0.73-1.34)	0.78 (0.69-0.89)

Supplemental Table 4.9. (cont'd) Hazard Ratios for Relative Intensity Activity Categories and the Selected Outcomes in OPACH Stratified by SPPB Category (n = 5,633)

		Incident Major CVD					
		Absolute		xgboost		Baynard et al	
Category	Model	Light	MVPA	Light	MVPA	Light	MVPA
Low (0 - 6)	1	0.80 (0.72-0.90)	0.46 (0.32-0.66)	0.59 (0.42-0.84)	0.81 (0.71-0.94)	0.57 (0.40-0.81)	0.91 (0.81-1.03)
	2	0.85 (0.76-0.95)	0.63 (0.44-0.91)	0.68 (0.48-0.97)	0.84 (0.73-0.97)	0.65 (0.45-0.93)	0.86 (0.75-0.98)
	3	0.90 (0.80-1.02)	0.75 (0.51-1.09)	0.81 (0.56-1.15)	0.86 (0.74-0.99)	0.74 (0.51-1.07)	0.88 (0.77-1.00)
Medium (7 - 9)	1	0.82 (0.74-0.90)	0.46 (0.35-0.60)	0.59 (0.44-0.80)	0.91 (0.80-1.03)	0.55 (0.41-0.75)	1.06 (0.97-1.17)
	2	0.87 (0.79-0.95)	0.66 (0.51-0.87)	0.66 (0.49-0.88)	0.91 (0.80-1.03)	0.66 (0.48-0.89)	0.93 (0.83-1.03)
	3	0.89 (0.80-0.98)	0.72 (0.54-0.94)	0.71 (0.53-0.95)	0.89 (0.79-1.02)	0.70 (0.51-0.95)	0.93 (0.84-1.04)
High (10 -12)	1	0.82 (0.72-0.94)	0.68 (0.51-0.92)	0.65 (0.45-0.94)	0.99 (0.82-1.18)	0.65 (0.44-0.96)	1.10 (0.96-1.25)
	2	0.90 (0.79-1.03)	1.00 (0.74-1.37)	0.80 (0.55-1.16)	0.92 (0.76-1.11)	0.88 (0.59-1.31)	0.89 (0.76-1.04)
	3	0.92 (0.80-1.06)	1.06 (0.77-1.46)	0.86 (0.59-1.26)	0.88 (0.72-1.08)	0.94 (0.63-1.40)	0.88 (0.75-1.03)

No statistically significant interaction terms.

SPPB = Short Physical Performance Battery

Model 1 = crude

Model 2 = Model 1 + age + race/ethnicity + education

Model 3 = Model 2 + RAND-36 Physical Function Score

Chapter 4, in part, is currently being prepared for submission for publication of the material. Schumacher, Benjamin T.; LaMonte, Michael J.; Di, Chongzhi; Simonsick, Eleanor M.; Parada, Humberto; Hooker, Steven P.; Bellettiere, John; LaCroix, Andrea Z. The dissertation author was the primary investigator and author of this paper.

5. Discussion

5.1. Summary

Aim 1 provided validation, recalibration, and predictive accuracy metrics of published $\dot{V}O_{2max}$ prediction equations with the aim of enabling large scale epidemiologic cohorts with older, ambulatory, community-dwelling adults to accurately estimate $\dot{V}O_{2max}$. Performance metrics of several of the previously published equations yielded reasonable results relative to measured $\dot{V}O_{2max}$. The recalibration of these equations using measured $\dot{V}O_{2max}$ in the BLSA improved every performance metric, although such recalibration would not be possible in epidemiologic cohorts unless they had directly measured $\dot{V}O_{2max}$ and the covariates used in the prediction equation. Among the previously published $\dot{V}O_{2max}$ prediction models, there was no discernable pattern of covariate types (i.e. demographics, body mass, self-reported PA) that contributed to the performance of the model more than others (e.g. the Bradshaw equation[37], one of the best performing models, has the same covariates as the Jurca equations[38], which did not perform as well in relation to measured $\dot{V}O_{2max}$ in the BLSA), likely because there is no group of covariate types (other than direct exercise testing) that adequately capture the integrated physiological signal reflected in measured $\dot{V}O_{2max}$.

Cox proportional hazards modeling showed measured $\dot{V}O_{2max}$ is a powerful predictor of all-cause mortality in both the unadjusted and adjusted models. Compared to participants in the lowest quartile of measured $\dot{V}O_{2max}$, those in the highest quartile had a 3-fold reduction in the risk of all-cause mortality, after adjusting for age, sex, race and ethnicity, and education. Several of the published equations yielded HRs similar in pattern and magnitude to those of measured $\dot{V}O_{2max}$ before adjustment, but these associations were not robust to even minimal adjustments. After adjustment for only age and sex, the ability of the equations to predict mortality was

substantially weakened, suggesting that much of the association observed in the unadjusted models was due to these two variables alone. In regression models using the recalibrated equations, the patterns of association were more similar to those estimated using measured $\dot{V}O_{2max}$ in unadjusted models. Despite the pattern of the recalibrated equations' HRs in unadjusted models, these associations were still not robust to adjustment.

These findings strongly suggest that while the $\dot{V}O_{2max}$ prediction equations may be valid and useful, to varying degrees, for individual exercise prescriptions in the field, their ability to predict mortality is severely compromised after adjustment for basic demographic and anthropometric covariates, some of which are components of the prediction equations themselves. $\dot{V}O_{2max}$ is a complex construct reflecting an integration of multifaceted organ systems and metabolic processes.[64] Without direct measures of the physiologic variability across individuals inherent in measured $\dot{V}O_{2max}$, even well-performing prediction equations based on basic demographic and health characteristics do not predict mortality independent of sex and age. To a large extent, this is because demographic and behavioral characteristics do not adequately capture the integrated physiological signal reflected in measured $\dot{V}O_{2max}$.

Aim 2, in direct response to the call for future research in an ML meta-analysis[66], developed and assessed the performance of multiple ML, non-exercise based $\dot{V}O_{2max}$ prediction algorithms that may enable large-scale epidemiologic cohorts with older, ambulatory, community-dwelling adults to accurately estimate $\dot{V}O_{2max}$, 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 2.9 to 4.4 $mL \bullet kg^{-1} \bullet min^{-1}$. For additional context, if one assumes the standard

conversion of 3.5 $mL \bullet kg^{-1} \bullet min^{-1}$ as being equivalent to 1 metabolic equivalent (MET), the errors in $\dot{V}O_{2max}$ prediction based on the ML algorithms used herein were about 0.8 and 1.3 METs. These predictive error values are lower than the RMSEs observed for the Aim 1 prediction equations and are lower than several RMSEs of previously published ML $\dot{V}O_{2max}$ prediction algorithms.[66]

Across all the algorithms, the RMSE values for the women were lower than the men. This is likely due to the larger variation in men's $\dot{V}O_{2max}$ measurements than the women's $\dot{V}O_{2max}$. Despite the better prediction of $\dot{V}O_{2max}$ for the BLSA women than men, the associations between measured and predicted $\dot{V}O_{2max}$ and all-cause mortality were notably stronger for the men than the women (men's Model 2 measured $\dot{V}O_{2max}$ Q4 vs. Q1 HR: 0.20 (0.06-0.70), $P_{trend} <$ 0.01; women's 0.63 (0.21-1.90), $P_{trend} = 0.14$).

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 $\dot{V}O_{2max}$, or at least other variables were able to compensate for their absence using these ML approaches. Specifically, When using all of the variables in the BLSA, the number of seconds to complete the 400m 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 400m walk), age became the most important variable.

Few non-exercise based $\dot{V}O_{2max}$ prediction ML models have been previously published, and even fewer have been developed specifically for older adults. Findings from Aim 1 on the assessment of the performance of previously published OLS models showed that when these

models are used to predict $\dot{V}O_{2max}$ in the BLSA, the RMSE values range from 5.1 - 20.4 mL•kg^{-1}•min^{-1} (see Chapter 2.4). After recalibrating these formulas' to measured $\dot{V}O_{2max}$ in the BLSA (obtaining new regression weights derived from the distribution of covariates in the BLSA) the RMSE values decrease to 3.8 - 4.2 mL•kg^{-1}•min^{-1} (see Chapter 2.4). A recent meta-analysis of 16 $\dot{V}O_{2max}$ prediction equations that use ML[66], few of which use non-exercise predictors and none of which were developed in older adults (the majority of the 16 equations were trained among men and women in their mid-to-late 20s; oldest age range included in the meta-analysis was 18-65), found RMSEs (mL•kg^{-1}•min^{-1}) of 2.9 (SVM), 3.14 (MLP Neural Network), 3.38 (tree boost), 4.78 (multilayer perceptron; MLP), 4.07 (artificial neural networks; ANN), 2.91 (feature selection with SVM), 3.37 (Generalized Regression Neural Networks), 4.51 (Single Decision Tree), and 4.78 (Multiple input single output (MISO) with MLP, SVM, and ANN with RBF). Interestingly, in the MISO model, the RMSEs were 4.07 for the women and 5.30 for the men, suggesting the sex differences as also 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, perhaps due to the decreased variance in $\dot{V}O_{2max}$ in the older adults included herein, differences in sample size, and/or ML training approaches.

The present study indicates that ML prediction of $\dot{V}O_{2max}$ in older adults has relatively low prediction error and is associated with a clinical aging outcome, all-cause mortality, in a similar pattern and magnitude of association as measured $\dot{V}O_{2max}$ in unadjusted analysis, however the utility of predicted $\dot{V}O_{2max}$ in estimating mortality risk in adjusted models was not as strong or robust as compared to measured $\dot{V}O_{2max}$ (see Chapter 3.4.). The attenuation of associations with mortality for predicted $\dot{V}O_{2max}$ but not measured $\dot{V}O_{2max}$ 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 $\dot{V}O_{2max}$.

Aim 3 was the first study to estimate participants' percent maximal effort (relative intensity of PA), to quantify its association with health outcomes, and compare these associations to those of absolute intensity of PA. As such, there exist few studies with which to compare the results herein. In a prospective cohort study of 7,337 men in the Harvard Alumni Health Study (mean age: 66 years), participants rated their usual level of exertion when exercising on a 10-point Borg Scale, categorized as 0 to 2 ("nothing to weak"), 3 ("moderate"), 4 ("somewhat strong"), and ≥ 5 ("strong to maximal"). Adjusted relative risks (RR (95% CI)) of coronary heart disease (CHD) for men reporting usual perceived exertion as "moderate," "somewhat strong," and "strong to maximal" were: 0.86 (0.66-1.13), 0.69 (0.51-0.94), and 0.72 (0.52-1.00), respectively ($P_{trend} = 0.02$), when compared with "nothing to weak".[15] Despite the differences between Lee et al.'s[15] and this dissertation's measurement of relative intensity, the outcome(s) that were assessed (incident CHD in Lee et al. and total mortality and incident major CVD in this dissertation), and the gender of the study populations, both bodies of work found that the highest category of relative intensity activity did not provide the strongest reduction in risk of outcome. In this dissertation, increases in activity requiring between 37% and 46% of maximal effort (relative light intensity PA) had the strongest reduction in risk of total mortality and incident major CVD, and Lee et al.[15] found that those engaging in PA at a perceived exertion of 4 on a 10-point scale had the lowest risk of incident CHD.

Findings from this dissertation do not support the assertion that one must achieve 6 METs on the absolute intensity scale to gain protection against total mortality and incident CVD.[93] Our results suggest that reductions in risk of total mortality and major CVD are stronger for every additional hour per day when engaged in activities that require between 37% and 46% maximal effort (relative light intensity PA) than when engaged in activities that require ≥ 46% maximal effort (relative MVPA). These findings also align with a recent study from the OPACH cohort, the same cohort from which women in the present study were drawn, that found that higher amounts of activities in daily life (specifically, "daily life movement", e.g. performing housework or gardening which was comprised of 69% light and 16% MVPA on the absolute scale) were independently associated with a lower risk of cardiovascular disease.[94]

These findings show that: (1) on the *absolute* scale, increases in daily MVPA were more strongly associated with mortality and major CVD than increases in light intensity activity, (2) on the *relative* scale, light intensity activity was more strongly associated with mortality and major CVD than MVPA, (3) a one-hour increase in *relative* light intensity activity was more strongly associated with mortality and CVD than a one-hour increase in *absolute* light intensity activity, and (4) a one-hour increase in *absolute* MVPA was more strongly associated with both outcomes than a one-hour increase in *relative* MVPA. Further, the proportion of older adults meeting the national recommendations for moderate and high intensity PA on an *absolute* intensity scale decreases with age, but the proportion meeting guidelines when activity is assessed on a *relative* intensity scale increases with age.[95] The findings in this dissertation support this, as strong correlations were observed between *absolute* light PA and the two *relative* light PA estimations (both r values = 0.88), but weak correlations between *absolute* MVPA and the two *relative* MVPA estimations (xgboost-MVPA r = 0.47, and Baynard-MVPA r = 0.35). It

is evident that measuring PA on the *relative* scale in older adults credits them with more MVPA than when measuring PA on the *absolute* scale. These findings extend the paradigm shift towards recommendation of more movement, regardless of intensity, and placing greater emphasis on *relative* light intensity activities (between 37% and 46% of one's maximal effort) as modifiable behavioral targets that are more easily achieved in older adults to reduce risks of death and CVD and improve prospects for healthy aging.

Aims 1 and 2 of this dissertation showed $\dot{V}O_{2max}$ to be a strong predictor of health outcomes, and these associations were stronger in older men than women. While previously published $\dot{V}O_{2max}$ prediction equations yielded reasonable estimates of $\dot{V}O_{2max}$, especially when recalibrated, these predictions are not robust to minimal adjustments in multivariable models (Aim 1). Using ML approaches (Aim 2), the accuracy in which $\dot{V}O_{2max}$ can be predicted improves, though the resulting predictions are still sensitive to adjustment. $\dot{V}O_{2max}$ predictions were more accurate relative to $\dot{V}O_{2max}$ in older women than men, but $\dot{V}O_{2max}$ (measured and predicted) was more strongly associated with health outcomes in men than women, perhaps due to the wider range of $\dot{V}O_{2max}$ values in men than women. When the universe of predictors for the ML algorithms was restricted to non-exercise variables common in many epidemiologic cohorts of aging (namely, OPACH), the predication errors were hardly affected, indicating that: (1) the performance-based variables important in the BLSA are not critical to accurate $\dot{V}O_{2max}$ prediction, and (2) that other variables become important in their absence. Though the transportation of one OLS equation from Aim 1 (Baynard et al.[44]) and the tree-boosted xgboost algorithm from Aim 2 to OPACH could not be directly validated, the distributions of the predicted $\dot{V}O_{2max}$ variables were reasonable and their correlations with selected covariates

closely emulated the same correlations in BLSA, bolstering confidence in the quality of the prediction.

This dissertation greatly advances the field of PA epidemiology and other exercise-science related fields through: (1) validating published $\dot{V}O_{2max}$ equations, (2) leveraging modern ML algorithms to develop new $\dot{V}O_{2max}$ prediction algorithms for older adults, (3) further establishing the advantage of measuring $\dot{V}O_{2max}$, a hallmark biomarker of healthy aging[9,46], or estimating $\dot{V}O_{2max}$ when laboratory measurements of $\dot{V}O_{2max}$ are infeasible, and, most importantly, (4) providing the first estimates of the relationship between percent maximal effort and health outcomes.

5.2. The importance of understanding the relationships between absolute intensity PA, relative intensity PA, and health outcomes

The current paradigm of absolute intensity—instead of relative intensity—is the prevailing paradigm in which epidemiologic studies of PA are conducted. Though the results from these studies have shown that increasing the frequency, duration, and intensity of PA on an absolute scale is beneficial, we do not yet know measuring PA on the relative scale contribute to our understanding of these associations. Several lab experimental studies have measured individual cardiorespiratory fitness, prescribed exercise based on these measurements, and have found that moderate relative intensity activity has beneficial associations with: blood pressure, lipids, insulin sensitivity, coagulation, and hemostasis.[14,96–99] Results from this dissertation indicate that relative light intensity (37 - 46% of maximal effort) is more important in reducing the risk of total mortality and incident major CVD than previously known to be. Gaining a deeper understanding of the associations between percent maximal effort as an estimate of

relative intensity of activity and these health outcomes would enable more attainable population-level PA recommendations and improve prospects for healthy aging.

5.3. Recommendations for future work on the intensity of PA and health outcomes

There is a critical need for the development and validation of more accurate and robust $\dot{V}O_{2max}$ prediction models in older adults. Given that by the year 2060, almost a quarter of the United States (U.S.) population will be comprised of adults 65 years of age or older[45], coupled with the association between beneficial health outcomes and higher $\dot{V}O_{2max}$, the development of a few accurate, robust prediction equations will enable $\dot{V}O_{2max}$ to be more broadly studied as a modifiable target for promoting functional resiliency and healthy aging. Further, future studies should seek to reproduce this dissertation's findings on the associations between absolute intensity PA, relative intensity PA, total mortality, and incident major CVD, especially if maximal exercise capacity testing has been conducted in the same cohort where accelerometry exists.

5.4. Concluding remarks

In conclusion, higher $\dot{V}O_{2max}$ is strongly associated with beneficial health outcomes. In large epidemiologic cohort studies that do not have the resources to directly measure $\dot{V}O_{2max}$, previously published OLS prediction equations can be used to yield reasonable estimates of $\dot{V}O_{2max}$, and ML prediction algorithms may yield more precise results. Findings from this dissertation extend the paradigm shift towards recommendation of more movement, regardless of intensity, and placing greater emphasis on *relative* light intensity activities (between 37% and 46% of one's maximal effort) as modifiable behavioral targets that are more easily achieved in older adults to reduce risks of death and CVD and improve prospects for healthy aging.

References

1. Department of Health & Human Services. 2018 Physical Activity Guidelines Advisory Committee Scientific Report. Published online 2018. doi:10.1111/j.1753-4887.2008.00136.x

2. Keadle SK, McKinnon R, Graubard BI, Troiano RP. Prevalence and trends in physical activity among older adults in the United States: A comparison across three national surveys. *Prev Med*. 2016;89:37-43. doi:10.1016/j.ypmed.2016.05.009

3. Caspersen C, Powell K CG. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. Public Health Rep. 1985;Mar-Apr100(2):126-31. *Public Health Rep*. 1985;Mar-Apr 1(2):126-131.

4. Tremblay M. Letter to the editor: Standardized use of the terms "sedentary" and "sedentary behaviours." *Appl Physiol Nutr Metab*. 2012;37(3):540-542. doi:10.1139/H2012-024

5. Evenson KR, Wen F, Herring AH, et al. Calibrating physical activity intensity for hip-worn accelerometry in women age 60 to 91years: The Women's Health Initiative OPACH Calibration Study. *Prev Med Rep*. 2015;2:750-756. doi:10.1016/j.pmedr.2015.08.021

6. Tzankoff SP, Norris AH. Effect of muscle mass decrease on age-related BMR changes. *J Appl Physiol*. 1977;43(6):1001-1006. doi:10.1152/jappl.1977.43.6.1001

7. Malatesta D, Simar D, Dauvilliers Y, et al. Energy cost of walking and gait instability in healthy 65- and 80-yr-olds. *J Appl Physiol*. 2003;95(6):2248-2256. doi:10.1152/japplphysiol.01106.2002

8. Howley ET. Type of activity: Resistance, aerobic and leisure versus occupational physical activity. In: *Medicine and Science in Sports and Exercise*. Vol 33. ; 2001. doi:10.1097/00005768-200106001-00005

9. 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(7):1334-1359. doi:10.1249/MSS.0b013e318213fefb

10. Smirmaul BPC, Bertucci DR, Teixeira IP. Is the VO2max that we measure really maximal? *Front Physiol*. 2013;4:203. doi:10.3389/fphys.2013.00203

11. Poole DC, Wilkerson DP, Jones AM. Validity of criteria for establishing maximal O2 uptake during ramp exercise tests. *Eur J Appl Physiol*. 2008;102(4):403-410. doi:10.1007/s00421-007-0596-3

12. Centers for Disease Control. Measuring Physical Activity Intensity | Physical Activity. Published February 28, 2022. Accessed March 23, 2022. https://www.cdc.gov/physicalactivity/basics/measuring/index.html

13. Borg GAV. Psychophysical bases of perceived exertion. *Med Sci Sports Exerc*. 1982;14(5):377-381. doi:10.1249/00005768-198205000-00012

14. Lee IM, Sesso HD, Oguma Y, Paffenbarger RS. Relative intensity of physical activity and risk of coronary heart disease. *Circulation*. 2003;107(8):1110-1116. doi:10.1161/01.CIR.0000052626.63602.58

15. Lee IM, Sesso HD, Oguma Y, Paffenbarger RS. Relative intensity of physical activity and risk of coronary heart disease. *Circulation*. 2003;107(8):1110-1116. doi:10.1161/01.CIR.0000052626.63602.58

16. A. Talbot E. Jeffrey Metter, Laura JLF. ABSOLUTE VERSUS RELATIVE INTENSITY CLASSIFICATION OF PHYSICAL ACTIVITY: IMPLICATIONS FOR PUBLIC HEALTH POLICY. *Educ Gerontol*. 2001;27(3-4):307-321. doi:10.1080/036012701750195012

17. Surgeon General's Report on Physical Activity and Health. *JAMA*. 1996;276(7):522. doi:10.1001/jama.1996.03540070018010

18. 2018 Physical Activity Guidelines Advisory Committee. 2018 Physical Activity Guidelines Advisory Committee Scientific Report. Published online 2018. doi:10.1111/j.1753-4887.2008.00136.x

19. Physical Activity Guidelines Advisory Committee Report, 2008 to the Secretary of Health and Human Services: (525442010-001). Published online 2008. doi:10.1037/e525442010-001

20. Lee D chul, Artero EG, Sui X, Blair SN. Mortality trends in the general population: the importance of cardiorespiratory fitness. *J Psychopharmacol Oxf Engl*. 2010;24(4 Suppl):27-35. doi:10.1177/1359786810382057

21. 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(16):1547-1553. doi:10.1001/jama.282.16.1547

22. 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(4):306-314. doi:10.1016/j.pcad.2014.09.011

23. 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(19):2024-2035. doi:10.1001/jama.2009.681

24. 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(12):1413-1423. doi:10.1093/aje/kwm031

25. LaMonte MJ, Barlow CE, Jurca R, Kampert JB, Church TS, Blair SN. Cardiorespiratory Fitness Is Inversely Associated With the Incidence of Metabolic Syndrome. *Circulation*. 2005;112(4):505-512. doi:10.1161/CIRCULATIONAHA.104.503805

26. Barlow CE, LaMonte MJ, Fitzgerald SJ, Kampert JB, Perrin JL, Blair SN. Cardiorespiratory fitness is an independent predictor of hypertension incidence among initially normotensive healthy women. *Am J Epidemiol*. 2006;163(2):142-150. doi:10.1093/aje/kwj019

27. Lakoski SG, Willis BL, Barlow CE, et al. Midlife Cardiorespiratory Fitness, Incident Cancer, and Survival After Cancer in Men: The Cooper Center Longitudinal Study. *JAMA Oncol*. 2015;1(2):231-237. doi:10.1001/jamaoncol.2015.0226

28. Peel JB, Sui X, Adams SA, Hébert JR, Hardin JW, Blair SN. A prospective study of cardiorespiratory fitness and breast cancer mortality. *Med Sci Sports Exerc*. 2009;41(4):742-748. doi:10.1249/MSS.0b013e31818edac7

29. Sui X, Lee DC, Matthews CE, et al. Influence of cardiorespiratory fitness on lung cancer mortality. *Med Sci Sports Exerc*. 2010;42(5):872-878. doi:10.1249/MSS.0b013e3181c47b65

30. Liu R, Sui X, Laditka JN, et al. Cardiorespiratory fitness as a predictor of dementia mortality in men and women. *Med Sci Sports Exerc*. 2012;44(2):253-259. doi:10.1249/MSS.0b013e31822cf717

31. Vidoni ED, Honea RA, Billinger SA, Swerdlow RH, Burns JM. Cardiorespiratory fitness is associated with atrophy in Alzheimer's and aging over 2 years. *Neurobiol Aging*. 2012;33(8):1624-1632. doi:10.1016/j.neurobiolaging.2011.03.016

32. Milani RV, Lavie CJ. Impact of cardiac rehabilitation on depression and its associated mortality. *Am J Med*. 2007;120(9):799-806. doi:10.1016/j.amjmed.2007.03.026

33. Sui X, Laditka JN, Church TS, et al. Prospective study of cardiorespiratory fitness and depressive symptoms in women and men. *J Psychiatr Res*. 2009;43(5):546-552. doi:10.1016/j.jpsychires.2008.08.002

34. Shephard RJ. Maximal oxygen intake and independence in old age. *Br J Sports Med*. 2009;43(5):342-346. doi:10.1136/bjsm.2007.044800

35. Sui X, LaMonte MJ, Laditka JN, et al. Cardiorespiratory fitness and adiposity as mortality predictors in older adults. *JAMA*. 2007;298(21):2507-2516. doi:10.1001/jama.298.21.2507

36. Fletcher GF, Ades PA, Kligfield P, et al. Exercise standards for testing and training: a scientific statement from the American Heart Association. *Circulation*. 2013;128(8):873-934. doi:10.1161/CIR.0b013e31829b5b44

37. Bradshaw DI, George JD, Hyde A, et al. An accurate VO2max nonexercise regression model for 18-65-year-old adults. *Res Q Exerc Sport*. 2005;76(4):426-432. doi:10.1080/02701367.2005.10599315

38. Jurca R, Jackson AS, LaMonte MJ, et al. Assessing cardiorespiratory fitness without performing exercise testing. *Am J Prev Med*. 2005;29(3):185-193. doi:10.1016/j.amepre.2005.06.004

39. Jackson AS, Blair SN, Mahar MT, Wier LT, Ross RM, Stuteville JE. Prediction of functional aerobic capacity without exercise testing. *Med Sci Sports Exerc*. 1990;22(6):863-870. doi:10.1249/00005768-199012000-00021

40. Matthews CE, Heil DP, Freedson PS, Pastides H. Classification of cardiorespiratory fitness without exercise testing. *Med Sci Sports Exerc*. 1999;31(3):486-493.

41. 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(10):576-580. doi:10.11622/smedj.2013186

42. de Souza E Silva CG, Kaminsky LA, Arena R, 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(7):742-750. doi:10.1177/2047487318763958

43. 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(4):313-319. doi:10.1620/tjem.227.313

44. Baynard T, Arena RA, Myers J, Kaminsky LA. The Role of Body Habitus in Predicting Cardiorespiratory Fitness: The FRIEND Registry. *Int J Sports Med*. 2016;37(11):863-869. doi:10.1055/s-0042-110572

45. 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*. www.census.gov/programs-surveys/popproj

46. 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(4):534-543. doi:10.1093/gerona/gly093

47. BLSA History. National Institute on Aging. Accessed August 17, 2021. http://www.nia.nih.gov/research/labs/blsa/history

48. 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(12):1416-1419.

49. Stone JL, Norris AH. Activities and attitudes of participants in the Baltimore longitudinal study. *J Gerontol*. 1966;21(4):575-580. doi:10.1093/geronj/21.4.575

50. BALKE B, WARE RW. An experimental study of physical fitness of Air Force personnel. *U S Armed Forces Med J*. 1959;10(6):675-688.

51. 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*. Published online 2006. doi:10.1111/j.1532-5415.2005.00530.x

52. Data Access - National Death Index - About. Published February 4, 2021. Accessed March 1, 2022. https://www.cdc.gov/nchs/ndi/about.htm

53. Schrack JA, Leroux A, Fleg JL, et al. Using Heart Rate and Accelerometry to Define Quantity and Intensity of Physical Activity in Older Adults. *J Gerontol A Biol Sci Med Sci*. 2018;73(5):668-675. doi:10.1093/gerona/gly029

54. 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(3):220-233. doi:10.1097/00005650-199603000-00003

55. Wanigatunga AA, Gresham GK, Kuo PL, et al. Contrasting characteristics of daily physical activity in older adults by cancer history. *Cancer*. 2018;124(24):4692-4699. doi:10.1002/cncr.31745

56. Martin Bland J, Altman DouglasG. STATISTICAL METHODS FOR ASSESSING AGREEMENT BETWEEN TWO METHODS OF CLINICAL MEASUREMENT. *The Lancet*. 1986;327(8476):307-310. doi:10.1016/S0140-6736(86)90837-8

57. Datta D. *Deepankardatta/Blandr: Version 0.4.2*. Zenodo; 2017. doi:10.5281/zenodo.824524

58. D'Agostino RB, Grundy S, Sullivan LM, Wilson P, CHD Risk Prediction Group. Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. *JAMA*. 2001;286(2):180-187. doi:10.1001/jama.286.2.180

59. Altman N, Krzywinski M. Regression diagnostics. *Nat Methods*. 2016;13(5):385-386. doi:10.1038/nmeth.3854

60. Therneau, Terry M. *A Package for Survival Analysis in R.* https://cran.r-project.org/package=survival

61. Farrell SW, Braun L, Barlow CE, Cheng YJ, Blair SN. The relation of body mass index, cardiorespiratory fitness, and all-cause mortality in women. *Obes Res*. 2002;10(6):417-423. doi:10.1038/oby.2002.58

62. Park MS, Chung SY, Chang Y, Kim K. Physical Activity and Physical Fitness as Predictors of All-Cause Mortality in Korean Men. *J Korean Med Sci*. 2009;24(1):13-19. doi:10.3346/jkms.2009.24.1.13

63. Salier Eriksson J, Ekblom B, Andersson G, Wallin P, Ekblom-Bak E. Scaling VO_2 max to body size differences to evaluate associations to CVD incidence and all-cause mortality risk. *BMJ Open Sport Exerc Med*. 2021;7(1):e000854. doi:10.1136/bmjsem-2020-000854

64. Mitchell JH, Blomqvist G. Maximal oxygen uptake. *N Engl J Med*. 1971;284(18):1018-1022. doi:10.1056/NEJM197105062841809

65. Phan DQ, Zheng C, Thai T, et al. Cardiorespiratory Fitness and Mortality in Patients Aged 60 to 90 Years. *Am J Cardiol*. Published online March 3, 2022:S0002-9149(22)00092-3. doi:10.1016/j.amjcard.2022.01.035

66. Ashfaq A, Cronin N, Müller P. Recent advances in machine learning for maximal oxygen uptake (VO2 max) prediction: A review. *Inform Med Unlocked*. 2022;28:100863. doi:10.1016/j.imu.2022.100863

67. 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(1):192. doi:10.1186/s12889-017-4065-6

68. Cowper DC, Kubal JD, Maynard C, Hynes DM. A Primer and Comparative Review of Major U.S. Mortality Databases. *Ann Epidemiol*. 2002;12(7):462-468. doi:10.1016/S1047-2797(01)00285-X

69. Pollack AZ, Hinkle SN, Liu D, et al. Vital Status Ascertainment for a Historic Diverse Cohort of U.S. Women. *Epidemiol Camb Mass*. 2020;31(2):310-316. doi:10.1097/EDE.0000000000001134

70. Lash TL, Silliman RA. A Comparison of the National Death Index and Social Security Administration Databases to Ascertain Vital Status. *Epidemiology*. 2001;12(2):259-261.

71. 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(2):M85-M94. doi:10.1093/geronj/49.2.M85

72. 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(4):M221-M231. doi:10.1093/gerona/55.4.M221

73. 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(10):M644-649. doi:10.1093/gerona/56.10.m644

74. Brach JS, Simonsick EM, Kritchevsky S, Yaffe K, Newman AB, Health, Aging and Body Composition Study Research Group. The association between physical function and lifestyle activity and exercise in the health, aging and body composition study. *J Am Geriatr Soc*. 2004;52(4):502-509. doi:10.1111/j.1532-5415.2004.52154.x

75. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw*. 2010;33(1).

76. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM; 2016:785-794. doi:10.1145/2939672.2939785

77. 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(4):361-387. doi:10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4

78. 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(3):363-373. doi:10.1016/0002-8703(82)90275-7

79. Mänttäri A, Suni J, Sievänen H, et al. Six-minute walk test: a tool for predicting maximal aerobic power (VO2 max) in healthy adults. *Clin Physiol Funct Imaging*. Published online May 31, 2018. doi:10.1111/cpf.12525

80. Ekelund U, Tarp J, Steene-Johannessen J, et al. Dose-response associations between accelerometry measured physical activity and sedentary time and all cause mortality: systematic review and harmonised meta-analysis. *BMJ*. 2019;366:l4570-l4570. doi:10.1136/bmj.l4570

81. Dempsey PC, Strain T, Khaw KT, Wareham NJ, Brage S, Wijndaele K. Prospective Associations of Accelerometer-Measured Physical Activity and Sedentary Time With Incident Cardiovascular Disease, Cancer, and All-Cause Mortality. *Circulation*. 2020;141(13):1113-1115. doi:10.1161/CIRCULATIONAHA.119.043030

82. Physical activity guidelines for Americans. *Okla Nurse*. 2008;53(4):25. doi:10.1249/fit.0000000000000472

83. American College of Sports Medicine (ACSM). *ACSMs Guidelines for Exercise Testing and Prescription*. 11th ed. Wolters Kluwer

84. Wu S. *A Novel Analysis of Accelerometry Data: Calibration and Flexible Association Modeling in an Epidemiological Study of Older Women*. Thesis. 2020. Accessed April 12, 2022. https://digital.lib.washington.edu:443/researchworks/handle/1773/46385

85. Anderson GL, Manson J, Wallace R, et al. Implementation of the women's health initiative study design. *Ann Epidemiol*. 2003;13(9, Supplement):S5-S17. doi:10.1016/S1047-2797(03)00043-7

86. The Women's Health Initiative Study Group. Design of the Women's Health Initiative Clinical Trial and Observational Study. *Control Clin Trials*. 1998;19(1):61-109. doi:10.1016/S0197-2456(97)00078-0

87. Schumacher BT, Bellettiere J, LaMonte MJ, et al. Accelerometer-Measured Daily Steps, Physical Function, and Subsequent Fall Risk in Older Women: The Objective Physical Activity and Cardiovascular Disease in Older Women Study. *J Aging Phys Act*. 2021;1(aop):1-11. doi:10.1123/japa.2021-0159

88. Choi L, Liu Z, Matthews CE, Buchowski MS. Validation of Accelerometer Wear and Nonwear Time Classification Algorithm. *Med Sci Sports Exerc*. 2011;43(2):357-364. doi:10.1249/MSS.0b013e3181ed61a3

89. Evenson KR, Wen F, Herring AH, et al. Calibrating physical activity intensity for hip-worn accelerometry in women age 60 to 91 years: The Women's Health Initiative OPACH Calibration Study. *Prev Med Rep*. 2015;2:750-756. doi:10.1016/j.pmedr.2015.08.021

90. BLSA History | National Institute on Aging. https://www.nia.nih.gov/research/labs/blsa/history

91. Curb JD, McTiernan A, Heckbert SR, et al. Outcomes ascertainment and adjudication methods in the Women's Health Initiative. *Ann Epidemiol*. 2003;13(9 Suppl):S122-8. doi:10.1016/s1047-2797(03)00048-6

92. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care*. 1992;30(6):473-483.

93. Shephard RJ. Absolute versus relative intensity of physical activity in a dose-response context. *Med Sci Sports Exerc*. 2001;33(6):S400.

94. Nguyen S, Bellettiere J, Wang G, et al. Accelerometer-Derived Daily Life Movement Classified by Machine Learning and Incidence of Cardiovascular Disease in Older Women: The OPACH Study. *J Am Heart Assoc*. 2022;11(5):e023433. doi:10.1161/JAHA.121.023433

95. A. Talbot E. Jeffrey Metter L Jerome L Fleg. ABSOLUTE VERSUS RELATIVE INTENSITY CLASSIFICATION OF PHYSICAL ACTIVITY: IMPLICATIONS FOR PUBLIC HEALTH POLICY. *Educ Gerontol*. 2001;27(3-4):307-321. doi:10.1080/036012701750195012

96. Hambrecht R, Wolf A, Gielen S, et al. Effect of exercise on coronary endothelial function in patients with coronary artery disease. *N Engl J Med*. 2000;342(7):454-460. doi:10.1056/NEJM200002173420702

97. Kelley DE, Goodpaster BH. Effects of exercise on glucose homeostasis in Type 2 diabetes mellitus. *Med Sci Sports Exerc*. 2001;33(6 Suppl):S495-501; discussion S528-9. doi:10.1097/00005768-200106001-00020

98. Kraus WE, Houmard JA, Duscha BD, et al. Effects of the amount and intensity of exercise on plasma lipoproteins. *N Engl J Med*. 2002;347(19):1483-1492. doi:10.1056/NEJMoa020194

99. Marceau M, Kouamé N, Lacourcière Y, Cléroux J. Effects of different training intensities on 24-hour blood pressure in hypertensive subjects. *Circulation*. 1993;88(6):2803-2811. doi:10.1161/01.cir.88.6.2803