**Reference values for and cross-validation of time to exhaustion on a modified Balke protocol in Norwegian men and women**

Eivind Aadland,[1] Ane Kristiansen Solbraa[1], Geir Kåre Resaland[1], Jostein Steene-Johannessen[1,2], Elisabeth Edvardsen[3,4], Bjørge Hermann Hansen[3], and Sigmund Alfred Anderssen[3]

[1] Sogn og Fjordane University College, Faculty of Teacher Education and Sports, Sogndal, Norway
[2] Kristiania University College – Department of Health Studies, Oslo, Norway
[3] Department of Sports Medicine, Norwegian School of Sport Sciences, Oslo, Norway
[4] Department of Pulmonary Medicine, Oslo University Hospital, Ullevål, Oslo, Norway

Short title: Reference values for modified Balke protocol

**Corresponding author**

Eivind Aadland

Faculty of Teacher Education and Sport, Sogn og Fjordane University College, Box 133, 6851 Sogndal, Norway. Phone: +47 5767 6086; Email: eivind.aadland@hisf.no

**Abstract**

**Introduction**: The aims of the present study were to provide reference values for time to exhaustion (TTE) on a modified Balke treadmill protocol, and to perform a cross-validation of TTE as a measure of maximal oxygen consumption ($VO_{2max}$), in Norwegian men and women 20 – 85 years of age. **Methods**: Reference values for TTE were derived from a national sample of 765 subjects. An additional sample of 119 subjects was included in the cross-validation (total n = 884), where prediction equations for $VO_{2max}$ was established. **Results**: A decline in TTE was seen with increased age. Prediction of $VO_{2max}$ in an independent dataset (n = 319) resulted in a $R^2$ = 0.78 and standard error of the estimate = 4.55 ml $kg^{-1}$ $min^{-1}$. The observed – predicted bias was small (mean difference < 1.24 ml $kg^{-1}$ $min^{-1}$), whereas random error was considerable (95% limits of agreement ± 7.11 – 9.70 ml $kg^{-1}$ $min^{-1}$) across age in both men and women. **Conclusions**: Despite limitations concerning the prediction of $VO_{2max}$ on an individual level, TTE from the Balke protocol is a good measure of aerobic fitness in adults across a range of settings, and could be evaluated according to the suggested reference values.

**Keywords:** Maximal graded treadmill protocol; Aerobic capacity; $VO_{2max}$; Validity;

**Introduction**

Aerobic fitness is strongly associated with health and longevity (Kodama et al., 2009; Barry et al., 2014), and is often an important outcome in exercise interventions. Thus, accurate determination of aerobic fitness is important in a range of clinical- and study settings. Maximal oxygen consumption ($VO_{2max}$) measured directly during a maximal treadmill or cycle protocol is generally considered the gold standard measure of aerobic fitness. However, due to the need for sophisticated and expensive equipment along with skilled personnel to measure $VO_{2max}$ directly, a wide array of test protocols that estimate aerobic fitness are available across populations and settings (Jorgensen et al., 2009; Solway et al., 2001; Fletcher et al., 2013), because of these tests simpler administration and superior feasibility.

Despite $VO_{2max}$ being regarded as the gold-standard for measurement of aerobic fitness, most evidence linking aerobic fitness to health are based on time to exhaustion (TTE) (or other performance measures, e.g., final stage or maximal load) on graded treadmill exercise protocols. For instance, the largest and possibly most influential study cohort in terms of establishing aerobic fitness as a strong predictor of longevity, the Cooper Center Longitudinal Study (CCLS) (previously the Aerobics Center Longitudinal Study (ACLS)) (Blair et al., 1996; Blair et al., 1989; Wei et al., 1999), rely on TTE determined from the Balke treadmill protocol (Balke and Ware, 1959). Time to exhaustion on this protocol compares well with directly measured $VO_{2max}$ ($r = 0.72 – 0.94$ and standard error of the estimate (SEE) = 2.2 – 4.3 ml $kg^{-1}$ $min^{-1}$) (Froelich and Lancaste, 1974; Froelicher et al., 1975; Pollock et al., 1976; Pollock et al., 1982). Nonetheless, due to the unambiguous relationship with health (Blair et al., 1996; Blair et al., 1989; Wei et al., 1999), TTE as determined from the Balke protocol is arguable a valid measure of aerobic fitness. Moreover, when compared to each other, directly measured $VO_{2max}$ and performance measures has been shown to be of similar prognostic value regarding mortality (Kavanagh et al., 2002; Goel et al., 2011) and of similar predictive value regarding sport performance (Hawley and Noakes, 1992; Noakes et al., 1990). Thus, performance on a graded maximal exercise protocol are a feasible and valid measure of aerobic fitness across a range of settings. Importantly, the Balke protocol are well suited in clinical settings due to its initial stages being of low intensity and as it does not requires running.

Recently, Edvardsen et al. (2013) published reference values for $VO_{2max}$, but not TTE, in a representative sample of Norwegian men and women 20 – 85 years of age based on a modified Balke treadmill protocol. To the best of our knowledge, reference values for TTE as obtained from a Balke protocol has not been published in women or subjects older than 55 years. Therefore, to make results obtained without direct measurement of $O_2$-consumption informative, the main aim of the

present study was to provide reference values for TTE from the modified Balke protocol in Norwegian adults. Moreover, we are not aware of any previous studies that have performed a cross-validation of TTE obtained from a Balke protocol for prediction of $VO_{2max}$. A secondary aim was therefore to perform a cross-validation of TTE against $VO_{2max}$ in this large sample.

## Methods

### Subjects

The present study is based on two datasets: a national sample (n = 904) and an additional sample (n = 159) recruited at one of the nine test centers that provided data for the national sample. The *national sample* is based on a large multicenter cross-sectional study to determine physical activity, aerobic fitness and determinants of physical activity in a representative sample of 20 – 85 year old Norwegians. Please see Hansen et al. (2012) and Edvardsen et al. (2013) for detailed information regarding the sample underlying the present data. In short, 1930 out of the 3485 subjects who initially provided data on objectively measured physical activity were invited to perform direct measurements of $VO_{2max}$, of which 904 undertook the examination.

The *additional sample* comprised 159 (out of 262 randomly invited) subjects, aged 40 – 43 and 53 – 54 years, who were invited for a direct measurement of $VO_{2max}$ and assessment of risk factors for cardiovascular diseases at one of the test centers (Solbraa et al., 2011).

Both studies were approved by the Regional Medical Ethics Committee and all subject provided written informed consent prior to participation.

### Test protocol

Before undertaking the maximal treadmill test, subjects were screened, and those having two or more risk factors for cardiovascular disease combined with an age > 50 years of age or a blood pressure > 180/110 mmHg were excluded from performing the maximal test.

Prior to the maximal treadmill protocol, subjects performed 2 – 7 minutes of self-paced walking on a level treadmill. After the familiarization, the treadmill was set to a start inclination of 4% and a start speed of 3.8 km per hour (km $h^{-1}$) (≥ 55 years old) or 4.8 km $h^{-1}$ (< 55 years old), and the subjects walked for 4 minutes (i.e., warm-up was included in the protocol). Thereafter, the elevation increased by 2% every minute until a maximal inclination of 20% was reached. The most fit subject

(those exceeding a test-duration of 12 minutes) thereafter received an increased speed of 0.5 km $h^{-1}$ until exhaustion.

Oxygen consumption was measured using three types of gas analyzers across the nine test centers: Oxycon Pro (Erich Jaeger GmbH, Hoechberg, Germany; n = 2), Vmax Sensormedics (CareFusion Corporation, San Diego, USA; n = 6) and Moxus Modular $VO_2$ system (AEI Technologies Inc, Pittsburgh, USA; n = 1). Volume and gas calibrations were performed daily according to the manufacturers' recommendations. In addition, all analyzers were checked for measurement precision and accuracy at two different time points with a standardized motorized mechanical lung (Motorized Syringe with Metabolic Calibration Kit; VacuMed, Ventura, USA) and/or a human calibrator, and a correction factor was applied for each gas analyzer to ensure comparable data between the test laboratories (range 0.925 – 1.059). In addition, all lab technicians involved in the data collection were experienced technicians and rigorously trained in the test procedures, and the nine treadmills were pre-programmed to ensure that identical exercise protocols were used across all labs. $VO_{2max}$ was reported as the mean oxygen uptake over a 30 second interval, whereas TTE was reported to the nearest 30 seconds. A test was considered valid (maximal) if individuals had a respiratory exchange ratio (RER) ≥1.10 or a rate of perceived exertion (RPE) ≥ 17 ($Borg_{6\text{-}20\ scale}$) (Borg, 1970). These end criteria was applied as they have been shown to be valid in this specific sample (Edvardsen et al, 2014), to end up with a reasonable sample size, and for the present results to align with the previously published reference values for $VO_{2max}$ (Edvardsen et al., 2013).

Physical activity was measured using the Actigraph GT1M accelerometer (ActiGraph, LLC, Pensacola, Florida, USA), as previously described (Hansen et al., 2012). Body mass and height was measured at the test centers. Overweight and obesity were defined as exhibiting a body mass index (BMI) ≥ 25 and 30 kg $m^{-2}$, respectively. Education level, ethnicity, smoking habits and prevalence of diabetes type 2 was self-reported.

Statistical analyses

The subject characteristics are presented as the mean values and standard deviation (SD).

Reference values were based on the *national sample* only, to retain the representativeness of the sample. Mean values and the 10 – 90 percentile range (mean ± 1.282 · SD of the residuals) for 5-year age groups (20 – 24 years, 25 – 29 years, etc.] were based on a linear mixed model determining the association between age and TTE in men and women for the low-and high intensity protocols

separately, including test center as a random effect. The quadratic effect of age was tested in all models, and retained in the model(s) if statistically significant.

The cross-validation of TTE as a measure of $VO_{2max}$ were based on *both datasets*, and assessed in several steps. First, we split our sample to construct one training dataset (~2/3 of the subjects, n = 567) and one testing dataset (~1/3 of the subjects, n = 317). Second, we tested the TTE · sex interaction in each of the two age-groups (including TTE, sex, age, body mass, and TTE · sex as independent variables) in the training dataset. Third, equations to predict $VO_{2max}$ from TTE in the training dataset was made separately in each sex- and age-group using the following model: $VO_{2max} = b_0 + b_1 \cdot TTE + b_2 \cdot body\ mass + b_3 \cdot age$. Fourth, the predicted and measured $VO_{2max}$ were compared in the testing dataset using linear regression, Bland-Altman plots and 95% limits of agreement (LoA [= SD of the differences · 1.96)]) (Bland and Altman, 1986). Finally, we calculated new sex- and age-specific equations to predict $VO_{2max}$ from TTE, body mass and age based on the whole sample. All equations for estimating $VO_{2max}$ were constructed using a linear mixed model, to take into account the clustering of observations within each test center. However, prediction models were evaluated using ordinary least squares regression to avoid overfitting of the model, as variance among test centers will be unknown in future prediction models.

The impact on TTE and $VO_{2max}$ of using RPE 17 vs. 18 – 20 and 17 vs. 19 – 20 to accept a maximal effort was analyzed using a linear mixed model including test center as a random effect and protocol, age and sex as fixed effects.

All analyses were performed using SPSS v. 23 (IBM SPSS Inc., Armonk, New York, USA). A p-value < .05 indicated statistically significant findings.

**Results**

Of the 1063 subjects that met for the examination of aerobic fitness, 22 subjects were excluded for having a high cardiovascular disease risk. Additionally, $VO_{2max}$ measurements were unavailable for 16 subjects due to technical problems. Finally, 141 subjects were excluded for not achieving the criteria for maximal effort. Reasons for ending the test prematurely were various combinations of poor motivation, medical problems or discomfort during testing (mainly musculoskeletal pain, dizziness or problems with maintaining balance), or discomfort wearing the face mask required for the measurement of $VO_{2max}$. Thus, a total of 884 subjects (448 men and 436 women) were included in the present analyses (83% of those meeting for the examination). Of these, 765 subjects (402 men and

363 women) were from the national representative sample, and served as subjects for establishing the reference values for TTE (table 1). The additional sample (n = 119) exhibited a higher physical activity level (20 – 32%, p < .001) and improved aerobic fitness (6 – 16%, p < .003) compared to the national sample.

Reference values for TTE

Table 2 shows reference values for TTE on the two protocols (i.e., start speed 3.8 and 4.8 km $h^{-1}$) for men and women. Men exhibited a longer test-duration than women in both age-groups. The 25 – 34 year old men and women exhibited the best performance, whereas a decline in performance was seen with increased age. The relationship with age was quadratic in the youngest age-group (p = .018 and .013 for $age^2$ in men and women, respectively) and linear in the oldest age-group (p = .859 and .988 for $age^2$ in men and women, respectively). Thus, reference values and the 80% reference range across age-groups (table 2; figure 1) was based on the following models: Men 20 – 54 years: TTE = 11.99 (6.33 – 17.65) + 0.27 (-0.02 – 0.57) · age - 0.00448 (-0.00819 - -0.00078) · $age^2$ (SEE = 2.23, n = 242); Women 20 – 54 years: TTE = 10.37 (5.14 – 15.60) + 0.26 (-0.02 – 0.53) · age - 0.00446 (-0.00798 - -0.00094) · $age^2$ (SEE = 2.21, n = 223); Men 55 – 85 years: TTE = 21.38 (17.70 – 25.05) -0.11 (-0.16 - -0.05) · age (SEE = 2.13, n = 160); Women 55 – 85 years: TTE = 23.06 (19.41 – 26.71) -0.17 (-0.22 - -0.11) · age (SEE = 2.50, n = 140).

Cross-validation of TTE as a measure of $VO_{2max}$

*Training dataset equations*: Splitting of the dataset left 567 subjects (286 men and 281 women; 363 20 – 54 year olds and 204 55 – 85 year olds) in the training dataset and 317 subjects (162 men and 155 women, 220 20 – 54 year olds and 97 55 – 85 year olds) in the testing dataset. In the training dataset, we found a significant sex · TTE interaction for prediction of $VO_{2max}$ in the youngest age-group (β (95% CI) -0.55 (-0.95 – -0.15, p = .007), whereas the interaction was weaker and non-significant (-0.22 (-0.64 – 0.19), p = .284) in the oldest age-group. Despite the weaker sex-specific effect in the oldest age-group, all equations were determined separately for men and women (Supporting information table 1).

*Cross-validation performance*: In the total sample of men and women of all ages (n = 319) the model fit indicated a small amount of shrinkage when predicted values was used to model the observed values of $VO_{2max}$, leaving $R^2$ = 0.78 and SEE = 4.55 ml $kg^{-1}$ $min^{-1}$ (p < .001). Corresponding values were

$R^2$ = 0.73 and SEE = 4.97 ml $kg^{-1}$ $min^{-1}$ in the 20 – 54 year old men, $R^2$ = 0.59 and SEE = 4.66 ml $kg^{-1}$ $min^{-1}$ in the 20 – 54 year old women, $R^2$ = 0.65 and SEE = 3.67 ml $kg^{-1}$ $min^{-1}$ in the 55 – 85 year old men, and $R^2$ = 0.57 and SEE = 3.97 ml $kg^{-1}$ $min^{-1}$ in the 55 – 85 year old women ($p < .001$ for all models). Prediction of $VO_{2max}$ from TTE resulted in a bias (95% confidence interval) ± 95% limits of agreement of 0.10 (-0.84 – 1.03) ± 9.70 ml $kg^{-1}$ $min^{-1}$ in the 20 – 54 year old men, -0.16 (-1.06 – 0.74) ± 9.29 ml $kg^{-1}$ $min^{-1}$ in the 20 – 54 year old women, 1.24 (0.22 – 2.27) ± 7.11 ml $kg^{-1}$ $min^{-1}$ in the 55 – 85 year old men, and 0.76 (-0.43 – 1.96) ± 7.90 ml $kg^{-1}$ $min^{-1}$ in the 55 – 85 year old women (figure 1). The bias varied from -4.00 (-5.05 – -2.93) to 3.55 (2.48 – 4.63) ml $kg^{-1}$ $min^{-1}$ across the test centers, when all age- and gender groups were merged. As the derived equations performed sufficiently well in the cross-validation, we established final equations based on the whole sample (table 3).

An examination of RPE in relation to performance showed that subjects that reported maximal values of 17 (n = 297) was not significantly different from those reporting values of 18-20 (mean [95% CI] TTE: -0.18 [-0.53 – 0.16] min, p = .309; $VO_{2max}$: -0.33 [-1.40 – 0.74] ml $kg^{-1}$ $min^{-1}$, p = .543, n = 442) or 19-20 (TTE: -0.19 [-0.61 – 0.23] min, p = .377; $VO_{2max}$: -0.04 [-1.32 – 1.25] ml $kg^{-1}$ $min^{-1}$, p = .952, n = 216).

**Discussion**

The present study provides reference values for performance on a modified Balke treadmill protocol in a large sample of Norwegian men and women aged 20 – 85 years. The 25 – 34 year old men and women exhibited the best performance, whereas a decline in performance was observed with increased age. Results from the cross-validation showed that TTE can be used as an accurate measure of $VO_{2max}$ on a group-level across age and gender. Yet, one must be aware of a substantial amount of noise when predicting $VO_{2max}$ on an individual level. Nevertheless, prediction of $VO_{2max}$ from performance on the modified Balke protocol would not be necessary in most settings, as TTE can be used directly and interpreted against the reference values presented herein.

The original Balke protocol (start speed 5.3 km $h^{-1}$ at a level treadmill; 1% increased inclination per minute) has shown to result in exhaustion at (median (10 – 90 percentile)) 17.5 (14 – 21), 16.5 (13 – 20) and 15.5 (12 – 19) minutes in 25 – 34, 35 – 44 and 45 – 54 year old men, respectively (Wolthuis et al., 1977). The protocol applied in the present study makes use of an increase in inclination of 2% every minute, but had a lower speed of walking than that of the original protocol (Balke and Ware, 1959). As a result, our findings in the 20 – 54 year old men compares well with TTE using the original Balke protocol (Balke and Ware, 1959; Wolthuis et al., 1977). To the best of our knowledge,

reference values have not been published in women or subjects older than 55 years. The lower treadmill speed used in subjects aged 55 – 85 years (3.8 km $h^{-1}$), led to test durations being quite similar to that of their younger counterparts (using a treadmill speed of 4.8 km $h^{-1}$). The long test duration might however constitute a problem with respects to the determination of maximal aerobic capacity in fit subjects, whose TTE greatly exceeds the recommended duration of 6 – 12 minutes of a test to determine aerobic fitness using a graded maximal exercise protocol (Fletcher et al., 2013; Buchfuhrer et al., 1983). Thus, for example the Bruce protocol has been shown to produce higher $VO_{2max}$ values than the Balke protocol in both men and women (Pollock et al., 1976; Pollock et al., 1982; Froelicher et al., 1975). Interestingly, Froelicher et al. (1975) did only find this difference in fit subjects. Nevertheless, underestimation of aerobic fitness in fit subjects are arguably not a big concern, as fit subjects are normally not those to target by exercise interventions aimed at improving health outcomes.

Previous studies using graded treadmill protocols have shown correlations of r = 0.72 – 0.92 and SEE = 3.95 – 4.26 ml $kg^{-1}$ $min^{-1}$ compared with directly measured $VO_{2max}$ in men (Froelich and Lancaste, 1974; Froelicher et al., 1975; Pollock et al., 1976), and r = 0.94 and SEE = 2.20 ml $kg^{-1}$ $min^{-1}$ in women (Pollock et al., 1982). The SEEs found in the training dataset in the present study were in the range of 3.36 ml $kg^{-1}$ $min^{-1}$ for the oldest women to 4.71 ml $kg^{-1}$ $min^{-1}$ for the youngest men. However, the SEE in the youngest men was reduced to 4.14 ml $kg^{-1}$ $min^{-1}$ when we excluded three subjects for having residuals for predicting $VO_{2max}$ < -15 ml $kg^{-1}$ $min^{-1}$ (results not shown). These subjects' oxygen uptake might have been underestimated due to unnoticed technical errors. Nevertheless, findings from the cross-validation were identical when these subjects were excluded, thus all subjects were kept in the training model as well as in the final model.

As expected, the cross-validation revealed a moderate degree of shrinkage of the model fit in most groups. As far as we are aware, no previous study has performed a cross-validation of a Balke protocol for prediction of $VO_{2max}$. Yet, our findings concerning the validity of the Balke protocol seem to be quite similar to results from previous studies of cross-validation of the 20 meter shuttle run test and maximal cycle ergometer tests in adults (Aandstad et al., 2011; Malek et al., 2004; Lockwood et al., 1997; Santtila et al., 2013; Green et al., 2013). However, samples varying with respect to size, gender, age, training status and health, along with disparate ways of reporting the findings, make comparisons between studies difficult. We found minimal biases in all groups, although somewhat larger biases were observed in the oldest age groups (being statistically significant in the 55 – 85 year old men), compared to the younger groups. A plausible explanation is that the oldest age groups were more prone to bias, because the smaller sample size of both the training- and the testing dataset (n less than one-half of that of the younger age groups) lead to more unstable results than in

the younger age groups. These results indicate that TTE, on a group level, was an accurate predictor of $VO_{2max}$ in the present study. However, the biases varied between test centers with a magnitude quite similar to biases reported in previous external validation studies (~5 ml $kg^{-1}$ $min^{-1}$) (Aandstad et al., 2011; Green et al., 2013; Malek et al., 2004). Given the measurement error of $O_2$-analyzers (Hodges et al., 2005), variation in execution of protocols, and often relatively small and homogeneous samples used for deriving or testing equations, biases in predictions are expected. The somewhat greater cluster effect found for prediction models for $VO_{2max}$ (ICC = 0.08 – 0.19) compared to TTE (ICC = 0.02 – 0.14) in the presents study (results not shown), despite rigid corrections of $VO_{2max}$-results using a mechanical lung among the nine test centers involved in the study, supports a possible lack of agreement between $O_2$-analyzers.

Despite good accuracy, the precision of the prediction of $VO_{2max}$ from TTE on an individual level was clearly less than perfect, as individual errors (i.e., LoA) of up to approximately ± 10 and ± 8 ml $kg^{-1}$ $min^{-1}$ must be expected in the younger and older age groups, respectively. Across the age- and gender groups this error equals ± 1.0 to 1.3 times the sample SDs for $VO_{2max}$. Random error in previous external cross-validation studies varies from SEEs ~2.5 to 4.2 ml $kg^{-1}$ $min^{-1}$ (Green et al., 2013; Malek et al., 2004; Lockwood et al., 1997) or LoA ~6 – 9 ml $kg^{-1}$ $min^{-1}$ (Aandstad et al., 2011; Santtila et al., 2013), thus our results seem to be in the middle- to upper range of these previous studies. Contrary to most previous studies which have used trained subjects or athletes, however, we report on a representative sample of men and women 20 – 85 years of age whose fitness and physical activity levels varies greatly. As reliability is a premise for validity, and athletes are more reliable than non-athletes (Hopkins et al., 2001), the lower and more variable training status of the subjects comprising the present data is a plausible explanation for the slightly poorer validity found. It is also reasonable to believe that the greater variation in training status might have introduced greater variation in work economy compared to previous studies (Lacour and Bourdin, 2015; Morgan et al., 1995). Accounting for work economy (and a plateau in $VO_{2max}$) has previously been shown to decrease the SEE of a prediction model for $VO_{2max}$ from TTE on a graded maximal treadmill test from 4.82 to 3.27 ml $kg^{-1}$ $min^{-1}$ in children (Aadland et al., 2014). Moreover, as discussed above, variation in the measurement of $O_2$-consumption between test centers will clearly influence our estimates of precision, as an ordinary linear regression model was used for the cross-validation. Actually, if data was tested using a mixed model (thus accounting for the differences between the test centers), the precision of the model improved by almost 20% (SEE = 3.76 vs. 4.55 ml $kg^{-1}$ $min^{-1}$, results not shown).

The magnitude of the random error found in the present- and previous studies, clearly shows that predicted levels of $VO_{2max}$ on an individual level must be interpreted carefully. Large errors on an individual level are a great problem in epidemiology, as it increases the probability of performing

type II errors (Hutcheon et al., 2010). Still, as hypothesized previously (Aadland et al., 2014), TTE and $VO_{2max}$ as obtained from a maximal graded treadmill protocol are different measures of aerobic fitness, of which both are measured with error, and both performs well as markers of health status (Goel et al., 2011; Kavanagh et al., 2002) and sport performance (Hawley and Noakes, 1992; Noakes et al., 1990). Therefore, we caution against use of $VO_{2max}$ values predicted from TTE on an individual level.

Reliance on TTE, as opposed to determination of $VO_{2max}$, causes a loss of two frequently used criteria to verify a maximal exertion from graded maximal protocols; the obtainment of a plateau in $O_2$-consumption despite an increased work load, and RER-levels above a certain threshold. Yet, which criteria that should be applied and how they should be defined are highly debated (Edvarsen et al, 2014). Determination of blood lactate concentration could be an alternative, however, in many settings where measurement of $VO_{2max}$ is excluded for not being feasible, measurement of the lactate concentration would probably also be inconvenient. Maximal heart rate is simple to measure, but exhibit great variation at maximal effort, thus, despite frequently being applied, it is not recommended as a criterion to verify maximal exertion (Fletcher et al., 2013; Edvardsen et al., 2014). The American Heart Association recommends using RPE ($Borg_{6-20}$ scale) ≥ 18 as a threshold for a valid test (Fletcher et al., 2013). We found a significantly lower performance of those reporting Borg values of 15 or 16 compared to those reporting values ≥ 17, but did not find any evidence of a decreased performance of those reporting a Borg value of 17 compared to ≥ 18 (TTE reduced by ~11 seconds). Thus, guided by our findings, in lack of measurement of $O_2$-consumption we recommend a Borg value of 17 being used as a criterion to verify a maximal exertion. Importantly, ~34% of the present sample obtained a RPE of 17, thus applying a stricter criterion will cause a substantial loss of subjects having a valid test result.

Strengths and limitations

A strength of the present study is the inclusion of a large sample of adults and older men and women. Reference values based on performance on a Balke protocol has not previously been determined in subjects older than 54 years (Balke and Ware, 1959; Wolthuis et al., 1977). A limitation, though, is that the present results is based on a subsample exhibiting a higher physical activity level and a lower BMI compared to the larger sample of 3485 subjects from where the sample was randomly drawn from (Hansen et al., 2012). Also, education level and prevalence of diabetes type 2 and smoking indicate that, especially the older age group, were a selected group comparing favorably to Norwegian population estimates. Finally, the BMI of both men and women

were ~1 – 2 kg $m^{-2}$ lower compared with data from a large population study in Norway (the HUNT 3 study) (Midthjell et al., 2013), and ethnic groups other than Caucasians were under-sampled. Taken together, this implies that our reference values might be somewhat overestimated. To determine the impact of the higher physical activity level and lower BMI in the present sample compared to current population estimates, we re-analyzed the TTE reference values adding BMI and physical activity level (overall cpm as obtained by accelerometry) as predictors to the models ($p \leq .002$ for both predictors across all models). Replacing mean values from the present sample with physical activity levels reduced by 10% (~-35 cpm) and BMI increased by 5% (~1.3 units) resulted in a decreased TTE of 26 to 35 seconds across the groups (results not shown). Thus, it is reasonable to believe that the Norwegian population level of TTE is approximately one-half a minute less than the reference values suggested by the present study.

Another limitation of the present study is the lack of data on reliability of the modified Balke protocol investigated. There is, however, little reason to believe that the reliability of TTE would be weaker than that of $VO_{2max}$ (Hopkins et al., 2001; Bosquet et al., 2008; Scott et al., 2015). According to Hopkins (Hopkins et al., 2001), who reviewed different measures of reliability, incremental tests provided coefficients of variation (CV) < 8% for all outcomes, with CVs for performance measures being 50% lower than those of $VO_{2max}$. As previously discussed, discrepancies might be due to challenges related to calibration of equipment (Hodges et al., 2005; Hopkins et al., 2001). Though, it should be mentioned that some studies have found a significant learning effect of TTE, but not for $VO_{2max}$ (Bensimhon et al., 2008), however, other studies have indicated the opposite (Scott et al., 2015), or no learning effect for either measure (Bosquet et al., 2008). In any case, appropriate study designs would minimize challenges related to possible test-retest biases.

We regard the use of national data collected from nine test centers from all parts of Norway a strength of the study. Thus, the results are obtained by many technicians, using nine different treadmills and three different types of $O_2$-analyzers. As discussed, this has probably caused conservative estimates of the measurement properties, thus we believe the reported findings should be representative to most new test settings.

## Perspectives

Performance on the graded maximal treadmill protocol used in the present study, is a feasible and good measure of aerobic fitness that is suitable in many settings (e.g., clinical, within fitness centers and in large population based studies). The reliance on the Balke protocol in the CCLS being the

largest study cohort that include measurement of aerobic fitness and unambiguously showing clear relationships between aerobic fitness and health (Blair et al., 1996; Blair et al., 1989; Wei et al., 1999), are strong arguments to support its use in large studies. Because the protocol do not require running (for those having a poor fitness level) and has a slow increase in work load, it is well suited to evaluate aerobic fitness in populations heterogeneous in terms of fitness level, age, adiposity or health status in general (Fletcher et al., 2013). To provide an easy way of using and comparing results from the protocol to Norwegian reference values for TTE and $VO_{2max}$, the equations provided in this paper is made available through the following web-page (a beta version): http://fitness.hisf.no. We hope the reference values provided can make the suggested protocol more easily applied, and found superior to other submaximal and/or indirect tests having poor or unknown measurement properties.

**Conclusion**

We provide reference values for TTE obtained from a modified Balke treadmill protocol in a large sample of Norwegian men and women, and we cross-validated performance on the protocol as a measure of $VO_{2max}$. Despite precautions should be taken when aiming to predict $VO_{2max}$ on an individual level, we suggest that using TTE from the Balke protocol presented herein is a feasible and good measure of aerobic fitness in adults across a range of settings. Still, as for all exercise testing procedures, precautions must be taken when applying the test in clinical settings (Fletcher et al., 2013).

Acknowledgement

We thank all the technicians for their contribution during the data collection at the nine institutions involved in the study: Finnmark University College, Hedmark University College, Norwegian University of Science and Technology Social Research, Sogn og Fjordane University College, University of Agder, University of Nordland, University of Stavanger, Telemark University College, and Oslo University Hospital, Ullevål. We also thank Morten Simonsen for development of the web-page provided. Finally, we could not have completed our study without financial support from the Norwegian Directorate of Health, the Norwegian School of Sport Sciences and the Sogn og Fjordane County Council.

**References**


Aadland E, Skrede T, Mamen A & Resaland GK, 2014. Validity of time to exhaustion on a fixed incremental treadmill protocol as a measure of aerobic fitness in 10-year old children. *Sport SPA,* 11**,** 5-13.

Aandstad A, Holme I, Berntsen S & Anderssen SA, 2011. Validity and reliability of the 20 meter shuttle run test in military personnel. *Mil Med,* 176**,** 513-518.

Balke B & Ware RW, 1959. An experimental study of physical fitness of air force personnel. *U S Armed Forced Med J,* 10**,** 675-88.

Barry VW, Baruth M, Beets MW, Durstine JL, Liu JH & Blair SN, 2014. Fitness vs. fatness on all-cause mortality: a meta-analysis. *Prog Cardiovasc Dis,* 56**,** 382-390.

Bensimhon DR, Leifer ES, Ellis SJ, Fleg JL, Keteyian SJ, Pina IL, Kitzman DW, Mckelvie RS, Kraus WE, Forman DE, Kao AJ, Whellan DJ, O'Connor CM, Russell SD & HF-ACTION Trial Investigators, 2008. Reproducibility of peak oxygen uptake and other cardiopulmonary exercise testing parameters in patients with heart failure (from the heart failure and a controlled trial investigating outcomes of exercise training). *Am J Cardiol,* 102**,** 712-717.

Blair SN, Kampert JB, Kohl HW, Barlow CE, Macera CA, Paffenbarger RS & Gibbons IW, 1996. Influences of cardiorespiratory fitness and other precursors on cardiovascular disease and all-cause mortality in men and women. *JAMA,* 276**,** 205-210.

Blair SN, Kohl HW, Paffenbarger RS, Clark DG, Cooper KH & Gibbons IW, 1989. Physical fitness and all-cause mortality. A prospective study of healthy men and women. *JAMA,* 262**,** 2395-2401.

Bland JM & Altman DG, 1986. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet,* 1**,** 307-310.

Borg G, 1970. Perceived exertion as an indicator of somatic stress. *Scand J Rehab Med,* 2-3**,** 92-98.

Bosquet l, Gamelin FX & Berthoin S, 2008. Reliability of postexercise heart rate recovery. *Int J Sports Med,* 29**,** 238-243.

Buchfuhrer MJ, Hansen JE, Robinson TE, Sue DY, Wasserman K & Whipp BJ, 1983. Optimizing the exercise protocol for cardiopulmonary assessment. *J Appl Physiol,* 55**,** 1558-1564.

Edvardsen E, Hansen BH, Holme IM, Dyrstad SM & Anderssen SA, 2013. Reference values for cardiorespiratory response and fitness on the treadmill in a 20-85-year-old population. *Chest, 144(1), 241-8*.

Edvardsen E, Hem E & Anderssen SA, 2014. End criteria for reaching maximal oxygen uptake must be strict and adjusted to sex and age: a cross-sectional study. *Plos One,* 9(1): e85276.

Fletcher GF, Ades PA, Kligfield P, Arena R, Balady GJ, Bittner VA, Coke IA, Fleg JL., Forman DE, Gerber TC, Gulati M, Madan K, Rhodes J, Thompson PD, Williams MA, 2013. Exercise standards for

testing and training: a scientific statement from the american heart association. *Circulation,* 128**,** 873-934.

Froelich VF & Lancaste MC, 1974. Prediction of maximal oxygen-consumption from a continuous exercise treadmill protocol. *Am Heart J,* 87**,** 445-449.

Froelicher VF, Thompson AJ, Noguera I, Davis G, Stewart AJ & Triebwasser JH, 1975. Prediction of maximal oxygen-consumption - comparison of Bruce and Balke treadmill protocols. *Chest,* 68**,** 331-336.

Goel K, Thomas RJ, Squires RW, Coutinho T, Trejo-Gutierrez JF, Somers VK, Miles JM & Lopez-Jimenez F, 2011. Combined effect of cardiorespiratory fitness and adiposity on mortality in patients with coronary artery disease. *Am Heart J,* 161**,** 590-597.

Green MS, Esco MR, Martin TD, Pritchett RC, Mchugh AN & Williford HN, 2013. Crossvalidation of two 20-m shuttle-run tests for predicting $VO_{2max}$ in female collegiate soccer players. *J Strength Cond Res,* 27**,** 1520-1528.

Hansen BH, Kolle E, Dyrstad SM, Holme I & Anderssen SA, 2012. Accelerometer-determined physical activity in adults and older people. *Med Sci Sports Exerc,* 44**,** 266-272.

Hawley JA & Noakes TD, 1992. Peak power output predicts maximal oxygen uptake and performance time in trained cyclists. *Eur J Appl Physiol,* 65**,** 79-83.

Hodges ID, Brodie DA & Bromley PD, 2005. Validity and reliability of selected commercially available metabolic analyzer systems. *Scand J Med Sci Sports,* 15**,** 271-279.

Hopkins WG, Schabort EJ & Hawley JA, 2001. Reliability of power in physical performance tests. *Sports Med,* 31**,** 211-234.

Hutcheon JA, Chiolero A & Hanley JA, 2010. Random measurement error and regression dilution bias. *BMJ,* 340.

Jorgensen T, Andersen IB, Froberg K, Maeder U, Smith IV & Aadahl M, 2009. Position statement: testing physical condition in a population - how good are the methods? *Eur J Sport Sci,* 9**,** 257-267.

Kavanagh T, Mertens DJ, Hamm IF, Beyene J, Kennedy J, Corey P & Shephard RJ, 2002. Prediction of long-term prognosis in 12 169 men referred for cardiac rehabilitation. *Circulation,* 106**,** 666-671.

Kodama S, Saito K, Tanaka S, Maki M, Yachi Y, Asumi M, Sugawara A, Totsuka K, Shimano H, Ohashi Y, Yamada N & Sone H, 2009. Cardiorespiratory fitness as a quantitative predictor of all-cause mortality and cardiovascular events in healthy men and women a meta-analysis. *JAMA,* 301**,** 2024-2035.

Lacour JR & Bourdin M, 2015. Factors affecting the energy cost of level running at submaximal speed. *Eur J Appl Physiol,* 115**,** 651-673.

Lockwood PA, Yoder JE & Deuster DA, 1997. Comparison and cross-validation of cycle ergometry estimates of $VO_{2max}$. *Med Sci Sports Exerc,* 29**,** 1513-1520.

Malek MH, Berger DE, Housh TJ, Coburn JW & Beck TW, 2004. Validity of $VO_{2max}$ equations for aerobically trained males and females. *Med Sci Sports Exerc,* 36**,** 1427-1432.

Midthjell K, Lee CMY, Langhammer A, Krokstad S, Holmen TL, Hveem K, Colagiuri S & Holmen J, 2013. Trends in overweight and obesity over 22 years in a large adult population: the HUNT Study, Norway. *Clin Obes,* 3**,** 12-20

Morgan DW, Bransford DR, Costill DL, Daniels JT, Howley ET & Krahenbuhl GS, 1995. Variation in the aerobic demand of running among trained and untrained subjects. *Med Sci Sports Exerc,* 27**,** 404-409.

Noakes TD, Myburgh KH & Schall R, 1990. Peak treadmill running velocity during the $VO_2$ max test predicts running performance. *J Sports Sci,* 8**,** 35-45.

Pollock ML, Bohannon RL, Cooper KH, Ayres JJ, Ward A, White SR & Linnerud AC, 1976. Comparative analysis of 4 protocols for maximal treadmill stress testing. *Am Heart J,* 92**,** 39-46.

Pollock ML, Foste C, Schmidt D, Hellman C, Linnerud AC & Ward A, 1982. Comparative-analysis of physiologic responses to 3 different maximal graded-exercise test protocols in healthy women. *Am Heart J,* 103**,** 363-373.

Santtila M, Hakkinen K, Pihlainen K & Kyrolainen H, 2013. Comparison between direct and predicted maximal oxygen uptake measurement during cycling. *Mil Med,* 178**,** 234-238.

Scott JM, Hornsby WE, Lane A, Kenjale AA, Eves ND & Jones IW, 2015. Reliability of maximal cardiopulmonary exercise testing in men with prostate cancer. *Med Sci Sports Exerc,* 47**,** 27-32.

Solbraa AK, Mamen A, Resaland GK, Steene-johannessen J, Ylvisåker E, Holme IM & Anderssen SA, 2011. Level of physical activity, cardiorespiratory fitness and cardiovascular disease risk factors in a rural adult population in sogn og fjordane. *Norsk epidemiologi,* 20**,** 179-188.

Solway S, Brooks D, Lacasse Y & Thomas S, 2001. A qualitative systematic overview of the measurement properties of functional walk tests used in the cardiorespiratory domain. *Chest,* 119**,** 256-270.

Wei M, Kampert JB, Barlow CE, Nichaman MZ, Gibbons IW, Paffenbarger RS & Blair SN, 1999. Relationship between low cardiorespiratory fitness and mortality in normal-weight, overweight, and obese men. *JAMA,* 282**,** 1547.

Wolthuis RA, Froelicher VF, Fischer J & Triebwasser JH, 1977. Response of healthy men to treadmill exercise. *Circulation,* 55**,** 153-157.

*Table 1. Subject characteristics for the national sample (n = 765) by sex and age. All values are mean (SD) or frequency.*

| | Men | | Women | |
|---|---|---|---|---|
| Age group (years) | **20-54** | **55-85** | **20-54** | **55-85** |
| n* | 242 | 160 | 223 | 140 |
| Age (years) | 40.4 (8.8) | 64.4 (6.2) | 40.7 (9.2) | 65.8 (7.9) |
| Height (cm) | 180.3 (6.3) | 178.1 (7.0) | 167.4 (6.3) | 164.6 (6.1) |
| Body mass (kg) | 84.6 (12.2) | 84.9 (11.5) | 70.2 (12.9) | 68.2 (10.3) |
| BMI (kg $m^{-2}$) | 26.0 (3.5) | 26.7 (3.1) | 25.0 (4.4) | 25.2 (3.7) |
| Overweight/obese (%) | 43.0/12.8 | 57.5/11.9 | 25.1/14.8 | 35.7/9.3 |
| Physical activity (cpm) | 370 (128) | 345 (134) | 363 (119) | 337 (123) |
| $VO_{2max}$ (ml $kg^{-1}$ $min^{-1}$) | 43.4 (9.3) | 33.1 (6.3) | 35.0 (7.2) | 27.4 (5.9) |
| $VO_{2max}$ (l $min^{-1}$) | 3.61 (0.63) | 2.79 (0.51) | 2.41 (0.44) | 1.84 (0.38) |
| TTE (min) | 15.3 (2.4) | 14.2 (2.5) | 13.1 (2.4) | 11.9 (2.9) |
| RER (ratio) | 1.23 (0.09) | 1.17 (0.11) | 1.21 (0.10) | 1.17 (0.12) |
| RPE (Borg $_{6\text{-}20\ scale}$) | 17.8 (1.2) | 17.4 (1.3) | 17.5 (1.4) | 17.5 (1.1) |
| $HR_{max}$ (beats $min^{-1}$) | 184.8 (12.9) | 162.8 (15.5) | 181.8 (10.2) | 164.8 (13.4) |
| Plateau (%) | 44 | 43 | 44 | 43 |
| Education level (%) | | | | |
| < High school | 14.5 | 7.0 | 10.5 | 14.1 |
| High school | 38.6 | 28.7 | 37.9 | 31.9 |
| University < 4 years | 21.2 | 24.8 | 25.6 | 23.7 |
| University ≥ 4 years | 25.7 | 39.5 | 26.0 | 30.4 |
| Caucasian origin (%) | 99.1 | 100 | 99.0 | 100 |
| Smokers (%) | 12.4 | 16.5 | 17.3 | 16.5 |
| Diabetes type 2 (%) | 4.2 | 3.9 | 1.8 | 2.2 |

BMI = body mass index; MVPA = moderate to vigorous physical activity; $VO_{2max}$ = maximal oxygen consumption; TTE = time to exhaustion; RER = respiratory exchange ratio; RPE = rate of perceived exertion; $HR_{max}$ = maximal heart rate; *total n = 710 – 765 across variables.

*Table 2. Reference values and 80% reference range for test duration (minutes) on a modified Balke treadmill protocol across age-groups in men and women.*

| Age group (years) | Men | | Women | |
|---|---|---|---|---|
| | **Age-predicted mean** | **10 – 90 percentile** | **Age-predicted mean** | **10 – 90 percentile** |
| | | | | |
| | **Protocol; start speed 4.8 km $h^{-1}$** | | | |
| **20-24** | 15.8 | 12.9 – 18.7 | 14.0 | 11.1 – 16.8 |
| **25-29** | 16.0 | 13.2 – 18.9 | 14.1 | 11.3 – 17.0 |
| **30-34** | 16.0 | 13.2 – 18.9 | 14.1 | 11.3 – 16.9 |
| **35-39** | 15.8 | 13.0 – 18.7 | 13.8 | 11.0 – 16.7 |
| **40-44** | 15.4 | 12.5 – 18.2 | 13.4 | 10.5 – 16.2 |
| **45-49** | 14.7 | 11.8 – 17.6 | 12.7 | 9.8 – 15.5 |
| **50-54** | 13.8 | 11.0 – 16.7 | 11.7 | 8.9 – 14.6 |
| | | | | |
| | **Protocol; start speed 3.8 km $h^{-1}$** | | | |
| **55-59** | 15.1 | 12.3 – 17.8 | 13.3 | 10.1 – 16.5 |
| **60-64** | 14.5 | 11.8 – 17.2 | 12.4 | 9.2 – 15.6 |
| **65-69** | 14.0 | 11.2 – 16.7 | 11.6 | 8.4 – 14.8 |
| **70-74** | 13.4 | 10.7 – 16.1 | 10.7 | 7.5 – 13.9 |
| **75-79** | 12.9 | 10.1 – 15.6 | 9.9 | 6.7 – 13.1 |
| **80-85** | 12.3 | 9.6 – 15.0 | 9.0 | 5.8 – 12.2 |

*Table 3. Final equations to estimate $VO_{2max}$ (ml $kg^{-1}$ $min^{-1}$) from TTE and body mass by sex and age based on the total dataset. Regression coefficients are reported as estimate (95% CI).*

| | **Men** | | **Women** | |
|---|---|---|---|---|
| | **Age < 55** | **Age ≥ 55** | **Age < 55** | **Age ≥ 55** |
| n | 288 | 160 | 295 | 141 |
| Intercept | 22.04 (13.28 – 30.79) | 40.05 (26.96 – 53.13) | 23.03 (16.49 – 29.56) | 39.67 (30.06 – 49.28) |
| Age (years) | -0.18 (-0.25 – -0.12) | -0.27 (-0.39 – -0.16) | -0.15 (-0.21 – -0.09) | -0.25 (-0.34 – -0.17) |
| Body mass (kg) | -0.13 (-0.18 – -0.08) | -0.13 (-0.19 – -0.07) | -0.10 (-0.14 – -0.06) | -0.13 (-0.19 – -0.07) |
| TTE (minutes) | 2.61 (2.33 – 2.89) | 1.49 (1.17 – 1.82) | 1.95 (1.71 – 2.20) | 1.07 (0.82 – 1.31) |
| SEE (ml $kg^{-1}$ $min^{-1}$) | 4.46 | 3.91 | 3.87 | 3.19 |

*TTE = time to exhaustion; SEE = standard error of the estimate*

*Supporting information table 1. Equations to estimate $VO_{2max}$ (ml $kg^{-1}$ $min^{-1}$) from TTE and body mass by sex and age based on the training dataset. Regression coefficients are reported as estimate (95% CI).*

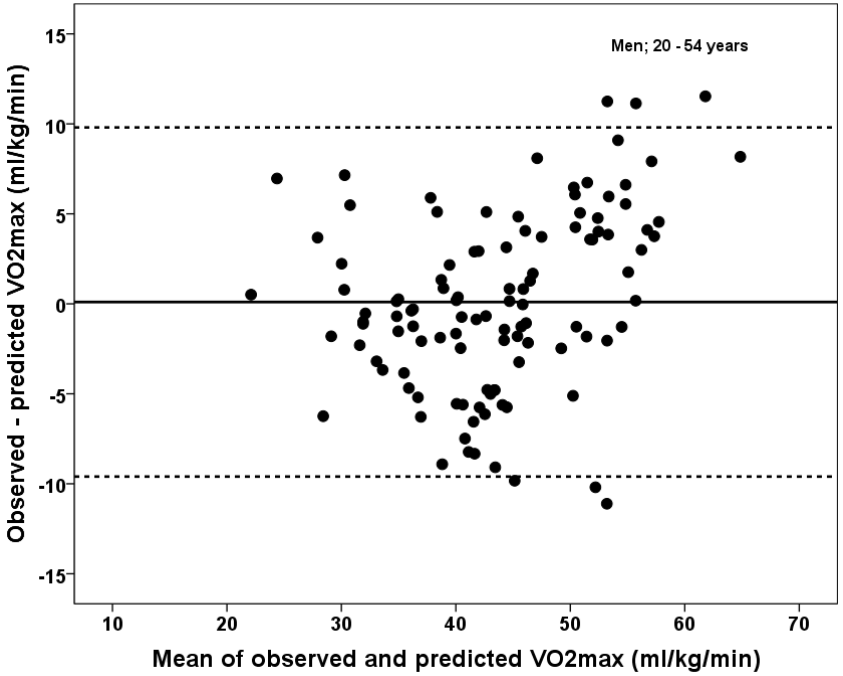
| | **Men** | | **Women** | |
|---|---|---|---|---|
| | **Age < 55** | **Age ≥ 55** | **Age < 55** | **Age ≥ 55** |
| n | 177 | 109 | 186 | 95 |
| Intercept | 25.21 (12.88 – 37.54) | 39.74 (22.13 – 57.35) | 22.81 (14.32 – 31.30) | 37.50 (24.95 – 50.05) |
| Age (years) | -0.20 (-0.29 – -0.11) | -0.28 (-0.44 – -0.12) | -0.15 (-0.23 – -0.08) | -0.24 (-0.36 – -0.13) |
| Body mass (kg) | -0.14 (-0.21 – -0.07) | -0.12 (-0.20 – -0.03) | -0.12 (-0.17 – -0.07) | -0.13 (-0.20 – -0.05) |
| TTE (minutes) | 2.52 (2.11 – 2.93) | 1.47 (1.06 – 1.88) | 2.09 (1.77 – 2.41) | 1.15 (0.82 – 1.47) |
| SEE (ml $kg^{-1}$ $min^{-1}$) | 4.71 | 4.05 | 3.87 | 3.36 |

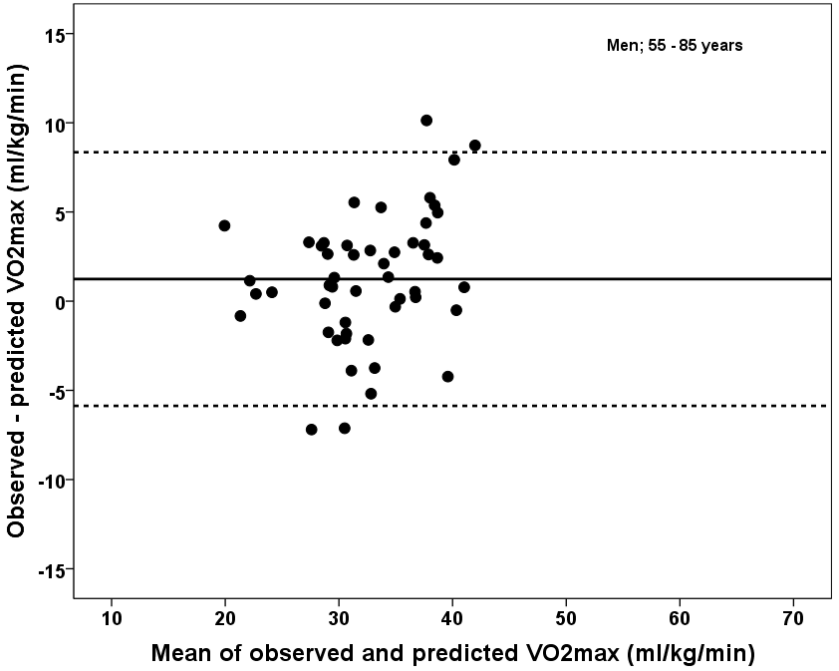
TTE = time to exhaustion; SEE = standard error of the estimate

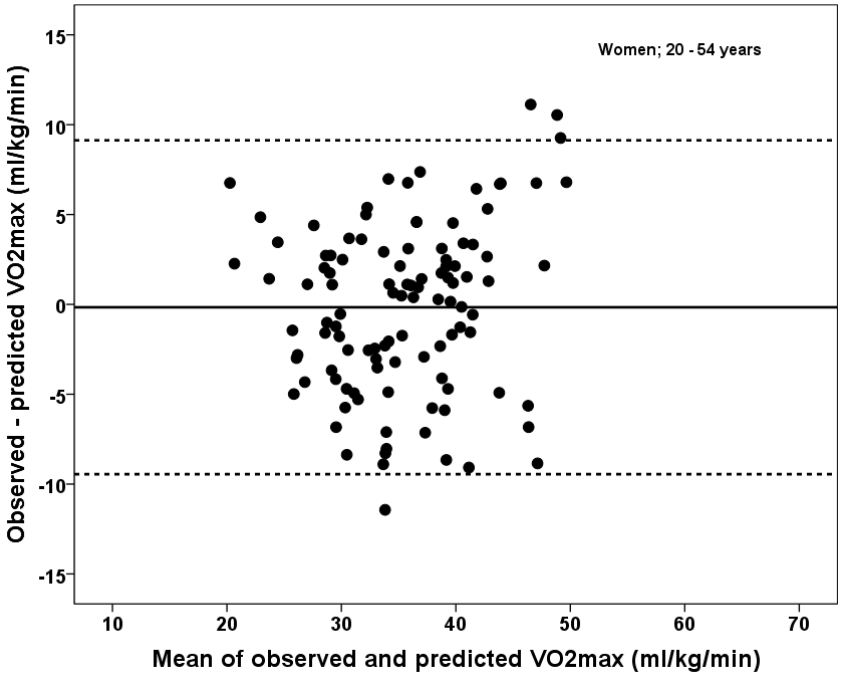
## Figure legends

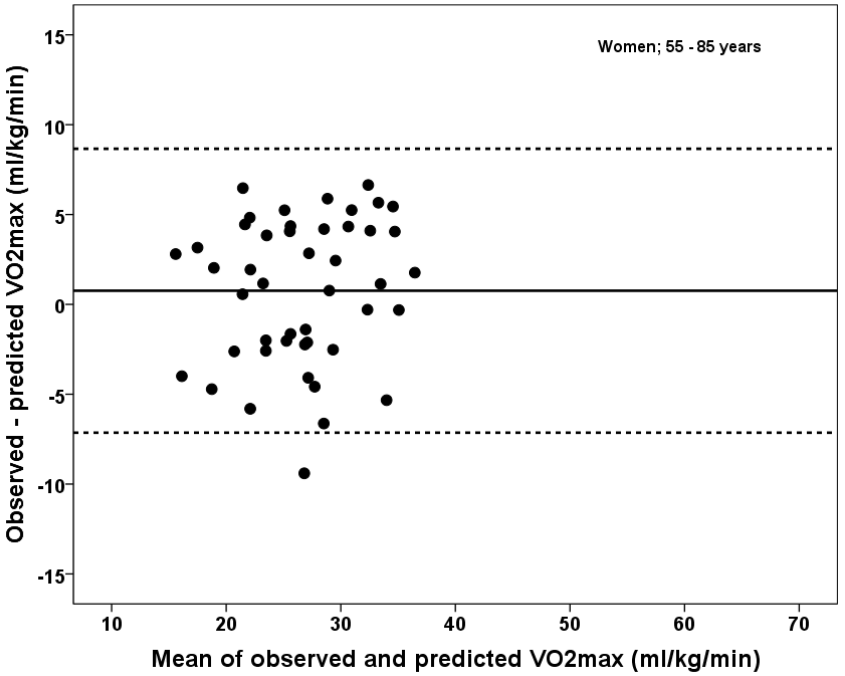
*Figure 1.* ***The relationship between age and TTE****. a) 20 – 54 year old men: TTE = 11.99 (6.33 – 17.65) + 0.27 (-0.02 – 0.57) · age - 0.00448 (-0.00819 – -0.00078) · age$^2$ (SEE = 2.23, ICC = 0.03, n = 242); Women 20 – 54 years: TTE = 10.37 (5.14 – 15.60) + 0.26 (-0.02 – 0.53) · age - 0.00446 (-0.00798 – -0.00094) · age$^2$ (SEE = 2.21, ICC = 0.01, n = 223); Men 55 – 85 years: TTE = 21.38 (17.70 – 25.05) -0.11 (-0.16 – -0.05) · age (SEE = 2.13, ICC = 0.14, n = 160); Women 55 – 85 years: TTE = 23.06 (19.41 – 26.71) -0.17 (-0.22 – -0.11) · age (SEE = 2.50, ICC = 0.03, n = 140).The bold line shows the trend for age, whereas the dotted lines show the 80% reference range (the 10 – 90 percentile). TTE = time to exhaustion; SEE = standard error of the estimate; ICC = intra-class correlation*

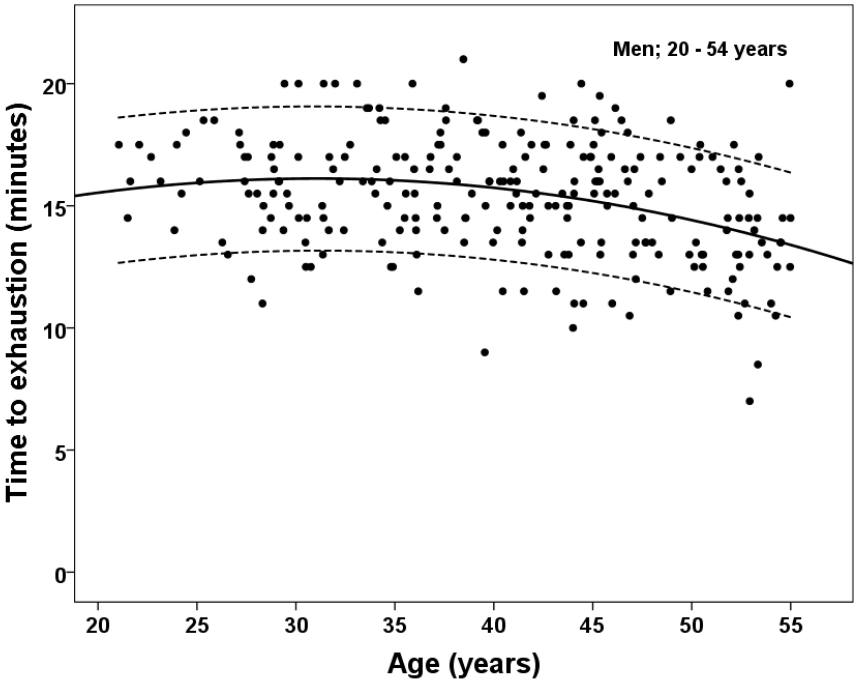
*Figure 2.* ***Agreement for observed versus predicted $VO_{2max}$.*** *Bland Altman plot showing the observed - predicted $VO_{2max}$ from TTE as a function of the mean of the two values in the testing dataset based on the regression equations derived from the training dataset (see supporting information table 1). The solid line is the mean difference; the dotted line is the 95% limits of agreement (bias ± 1.96 SD of the difference). $VO_{2max}$ = maximal oxygen consumption*

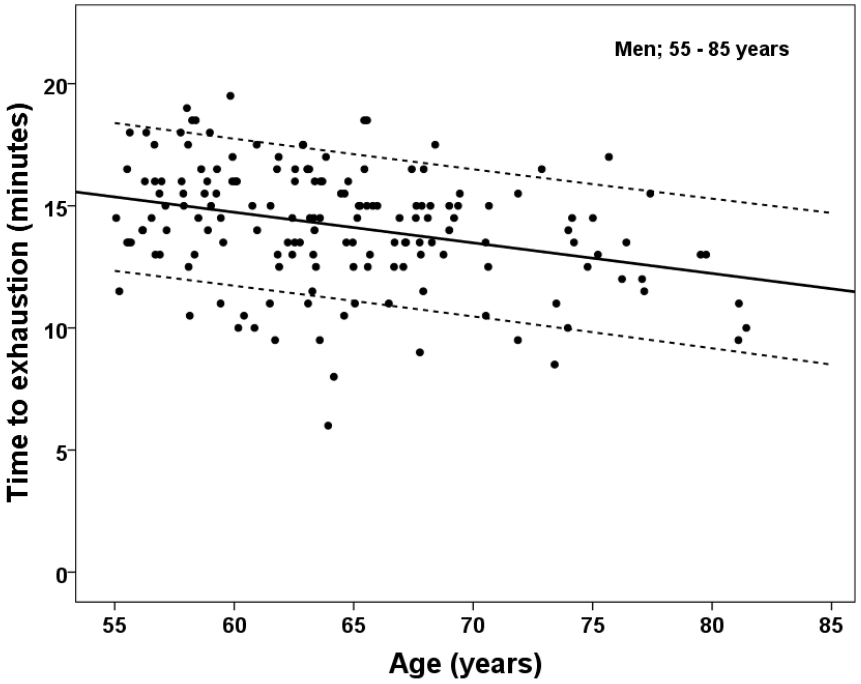

Men; 20 - 54 years
Observed - predicted VO2max (ml/kg/min)
Mean of observed and predicted VO2max (ml/kg/min)

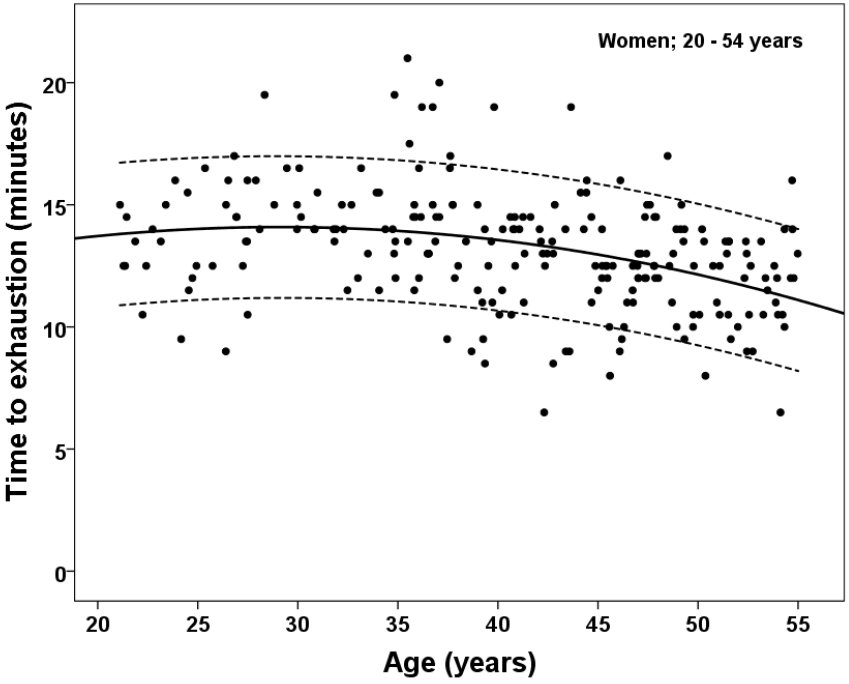

Men; 55 - 85 years
Observed - predicted VO2max (ml/kg/min)
Mean of observed and predicted VO2max (ml/kg/min)

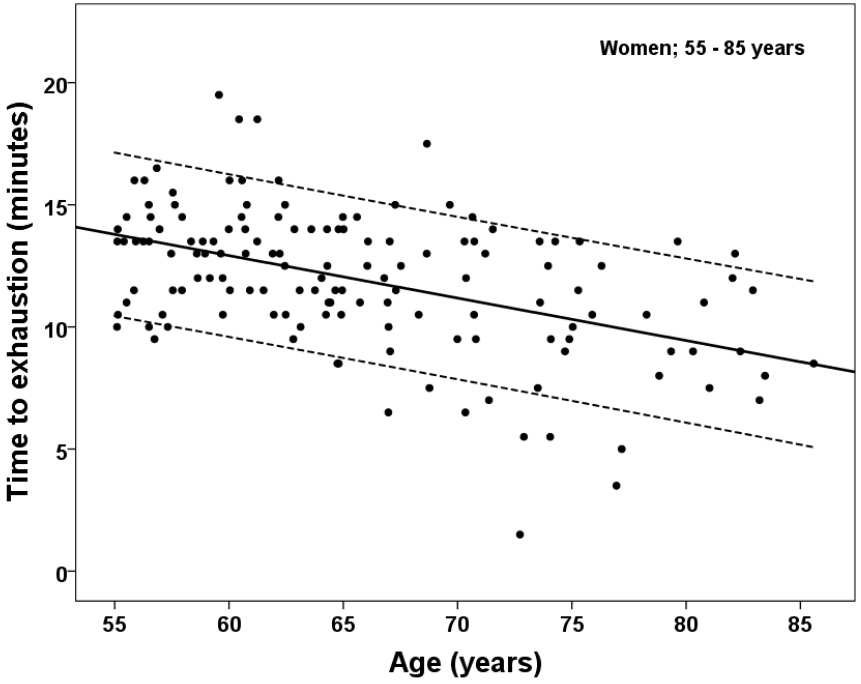

Women; 20 - 54 years
Observed - predicted VO2max (ml/kg/min)
Mean of observed and predicted VO2max (ml/kg/min)


Women; 55 - 85 years
Observed - predicted VO2max (ml/kg/min)
Mean of observed and predicted VO2max (ml/kg/min)


Men; 20 - 54 years
Time to exhaustion (minutes)
Age (years)


Men; 55 - 85 years
Time to exhaustion (minutes)
Age (years)


Women; 20 - 54 years
Time to exhaustion (minutes)
Age (years)


Women; 55 - 85 years
Time to exhaustion (minutes)
Age (years)