Nevill, A. M., & Cooke, C. B. (2017). The Dangers of Estimating V̇O2max Using Linear, Nonexercise Prediction Models. *Medicine and Science in Sports and Exercise*, *49*(5), 1036-1042. https://doi.org/10.1249/MSS.0000000000001178

© 2016 American College of Sports Medicine. This is a non-final version of an article published in final form in Nevill, A.M and Cooke, C.B. (2016) The Dangers of Estimating V˙O2max Using Linear, Nonexercise Prediction Models. *Medicine and Science in Sports and Exercise.* DOI: 10.1249/MSS.0000000000001178

The dangers of estimating $\dot{V}O_2$ max using linear, non-exercise prediction models.

Running Title: Dangers of estimating $\dot{V}O_2$ max.

Authors: Alan M. Nevill[1] and Carlton B. Cooke[2]

1. Faculty of Education, Health and Wellbeing, University of Wolverhampton, Walsall Campus, Walsall, WS1 3BD, UK
2. School of Social and Health Sciences, Leeds Trinity University, Leeds, LS18 5HD, UK

Address for correspondence:

Professor Alan M. Nevill, Ph.D.

University of Wolverhampton,

Faculty of Education, Health and Wellbeing,

Walsall Campus

Gorway Road

Walsall, WS1 3BD

Tel: +44 (0)1902 322838

Fax: +44 (0)1902 322894

Email: a.m.nevill@wlv.ac.uk

Word Count: 4972

Abstract

Purpose: To compare the accuracy and goodness-of-fit of two competing models (linear versus allometric) when estimating $\dot{V}O_2$max (ml.kg^{-1}.min^{-1}) using non-exercise prediction models. **Methods**: The two competing models were fitted to the $\dot{V}O_2$max (ml.kg^{-1}.min^{-1}) data taken from two previously published studies. Study 1 (the Allied Dunbar National Fitness Survey, ADNFS), recruited 1732 randomly selected healthy participants, aged 16 years and over, from thirty English parliamentary constituencies. Estimates of $\dot{V}O_2$max were obtained using a progressive incremental test on a motorized treadmill. In Study 2 (3), maximal oxygen uptake was measured directly during a fatigue limited treadmill test in older men (n = 152) and women (n = 146) aged 55 to 86 years. **Results**: In both studies, the quality-of-fit associated with estimating $\dot{V}O_2$max (ml.kg^{-1}.min^{-1}) was superior using allometric rather than linear (additive) models based on all criteria (R^2, maximum log-likelihood and AIC). Results suggest that linear models will systematically over-estimate $\dot{V}O_2$max for participants in their 20’s and under-estimate $\dot{V}O_2$max for participants in their 60’s and older. The residuals saved from the linear models were neither normally distributed, nor independent of the predicted values nor age. This will probably explain the absence of a key quadratic age^2 term in the linear models, crucially identified using allometric models. Not only does the curvilinear age decline within an exponential function follow a more realistic age decline (the right-hand side of a bell-shaped curve), but the allometric models identified either a stature-to-body-mass ratio (study 1) or a fat-free-mass-to-body-mass ratio (study 2), both associated with leanness when estimating $\dot{V}O_2$max. **Conclusions**: Adopting allometric models will provide more accurate predictions of $\dot{V}O_2$max (ml.kg^{-1}.min^{-1}) using plausible, biologically sound and interpretable models.

Keywords: Curvilinear age decline, bell-shaped curve, quality of fit, residuals.

Introduction

The value of accurately estimating $\dot{V}O_2$ max (ml.kg^{-1}.min^{-1}) has been highlighted in a recent large, population-based cohort study (14) from the Jebsen Center for Exercise in Medicine at the Norwegian University of Science and Technology. The study demonstrated that a simple estimation of $\dot{V}O_2$ max can predict long-term cardiovascular disease and all-cause mortality. Hence the accuracy and validity of estimating $\dot{V}O_2$ max is paramount in reporting the association/link between $\dot{V}O_2$ max and all-cause mortality.

Several studied have reported non-linear associations between $\dot{V}O_2$ max and age, and $\dot{V}O_2$ max and body mass (5, 10, 16). Hence it was surprising that Nes et al. (13) adopted a linear model to estimate $\dot{V}O_2$ max (ml.kg^{-1}.min^{-1}) that was subsequently used by Nes et al. (14) to predict long-term all-cause mortality and cardiovascular disease (CVD). The authors reported the following linear regression models to estimate $\dot{V}O_2$ max (ml.kg^{-1}.min^{-1}) for men; 100.27 - (0.296 · age) - (0.369 · WC) - (0.155 · RHR) + (0.226 · PA-index), and for women; 74.74 - (0.247· age) - (0.259· WC) - (0.114· RHR) + (0.198· PA-index), where WC=waist circumference; RHR=resting hear rate; PA index=physical activity index. The authors reported that their models were unable to detect any interaction or polynomial terms, i.e. the inclusion of such terms was unable “to influence the R^2 of the models appreciably”.

There are at least three major concerns with these linear, additive models. Firstly, both models suggest a linear decline in age that has the same rate (same slope parameter) for participants in the twenties, as in their fifties or sixties and in their eighties. However, there is evidence in the literature that indicates a curvilinear decline in $\dot{V}O_2$ max with age, suggesting the need for a non-linear or quadratic age term to be incorporated into the model, see Astrand and Rodahl (5) Figure 7-15 on page 337, and Hawkins (10). The second concern is the absence of a weight/body-mass term in both models. Nevill et al. (19) and Astrand and Rodahl (5) in their Figure 9-4 on page 400, reported a strong negative association between $\dot{V}O_2$max (ml.kg^{-1}.min^{-1}) and body mass. This is because absolute $\dot{V}O_2$max (l.min^{-1}) scales to, or is associated with body mass (M$^{0.67}$), and hence when researchers calculate $\dot{V}O_2$max (ml.kg^{-1}.min^{-1}), by dividing $\dot{V}O_2$max (l.min^{-1}) by body mass (M), the resulting ratio "over-scales" leaving $\dot{V}O_2$max (ml.kg^{-1}.min^{-1}) proportional to M$^{-0.33}$. This non-linear association with mass should have been considered by Nes et al. (13). Incorporating a power-function body-mass term as a predictor in both models is likely to improve the accuracy when predicting $\dot{V}O_2$max (ml.kg^{-1}.min^{-1}).

Another major concern with these fitted models is the fact that the residuals from both linear models are unlikely to be, a) normally distributed (16) and b) independent of the predictor variables (in particular age). If the residuals demonstrate a lack of normality and independence, then the validity of the models (i.e., the statistical significance of the estimated parameters) will be questionable. For example, we cannot be confident that the decline in age is linear, as discussed above, and that by fitting an alternative biologically-sound allometric model, that a non-linear or curvilinear decline in age and a curvilinear power-function term in body mass might have been detected. For a brief and concise history of allometric modeling, see Winter and Nevill (23).

Hence the purpose of this study was to fit the same linear, additive model adopted by Nes et al. (13) to both estimated and directly measure $\dot{V}O_2$max (ml.kg^{-1}.min^{-1}) data from two previously published studies, Study 1 the Allied Dunbar National Fitness Survey (ADNFS) (2, 16), and Study 2, data reported by Amara et al. (3), to compare a linear model with an alternative, proportional allometric model to discover whether the latter provides, 1) a superior quality of fit (using R^2, maximum log-likelihood and AIC criterion), 2) more normally distributed residuals and, 3) a more plausible, biologically sound and interpretable model.

Methods (study 1)

All variables and measurement used in the current study have been previously described and published (16) or reported in a technical report (7). Cardiopulmonary fitness or $\dot{V}O_2$ max was assessed using a progressive incremental test on a motorized treadmill. In reality, the $\dot{V}O_2$ max measurements are estimates based on the linear relationship (for each subject) between the oxygen cost and heart rate, recorded breath-by-breath (n>50) during a sub-maximal exercise test using an automated respiratory gas analyzer (Quinton Q-plex) and a diagnostic electrocardiogram (Quinton Q4000). The test continued until the end of a one-minute stage in which the subject's heart rate had reached 85% of estimated maximum for age (210 - 0.65·age, beats min^{-1}). For a given individual, the estimated $\dot{V}O_2$ max is the predicted oxygen cost at an assumed maximum heart rate, taken to be 210-0.65·age (11). All submaximal tests used to estimate $\dot{V}O_2$ max are associated with a standard error of prediction which is typically in the range of 10% - 15% (5). One advantage of the protocol used in the Allied Dunbar National Fitness Survey (2, 7) is that the $\dot{V}O_2$ of each stage was directly measured, which eliminates variations in mechanical efficiency associated with the use of workload. However, the accuracy of the method is still dependent on the variability in predicted maximum heart rate, which in normal adult participants has been shown to have a standard deviation of 10-12 beats.min^{-1} (4). The validity of the linear extrapolation method described by Lange-Anderson et al. (12) to predict $\dot{V}O_2$ max using measured submaximal $\dot{V}O_2$ values to a predicted maximimum heart rate has been assessed against directly determined treadmill $\dot{V}O_2$ max, where it was shown to under-predict by 13% with an SE of 1.4 ml.kg.$^{-1}$min^{-1} (9), which is within the range typically reported for estimations of $\dot{V}O_2$ max.

For our measure of physical activity, we adopted the number of 20 min bouts of vigorous exercise (VIGEX), defined as activities that were > 7.5 kcal.min^{-1} or >60% of aerobic capacity reported during the four weeks prior to the exercise test. There are well established limitations to methods of physical activity assessment that rely on self-report which have been shown to introduce measurement error and bias (1). However, in a preliminary study for the Allied Dunbar National Fitness Survey, the recall of participants was shown to be consistent in over 80% of repeat interviews that were completed one month apart (technical report (7) page 11).

Waist girth measurements were obtained using a standardized protocol (see the technical report (7) page 54). From behind the subject, the administrator identifies the iliac crest and the 12th rib, keeping the second (index) and fourth fingers on the sites. A mark, using a demographic pencil, was put on the skin midway between two sites using the third (middle) finger as an indicator. This was repeated on the other side of the body. The tape was placed around the waist to cover the two marked spots and to lie in a horizontal plane around the body. The subject was instructed to stand upright in the standard anatomical position and to breathe normally. The reading was noted at the onset of inhalation and of exhalation and a mean value was recorded to the nearest millimeter.

Resting heart rate (RHR) measurements were also obtained using a standard protocol for obtaining blood pressure and resting heart rate using an automated sphygmomanometer

(Accutorr 1, Data Corporation, Cambridge, UK; see technical report (7) page 57). Measurements were carried out after the anthropometry and flexibility test but before any strenuous tests. At least three measurements were recorded at one minute intervals, after the participants had been seated with their legs uncrossed for at least three minutes. The value used for resting heart rate was that associated with the lowest diastolic blood pressure measurement.

Methods (study 2)

A detailed description of subject selection and recruitment are provided in a previous study see Amara et al. (3). Briefly, the subjects were independently living women (n = 146) and men (n = 152) who volunteered to participate in the study and indicated verbally that they were able to walk a distance of 80 m (self-paced walk test). Body mass (M) was assessed to the nearest 0.1 kg using calibrated Leverbalance scales (HealthOMeter, Inc., Bridgeview, IL, USA) and body height was measured using a stadiometer to the nearest 0.1 cm with the subject standing, lightly clothed and without footwear. Harpenden skinfold calipers (Harpenden, British Indicators Ltd, UK) were used to measure skinfold thickness at four sites (biceps, triceps, suprailiac and subscapular) on the right side of the body. Total body density was estimated from the log of the sum of four skinfold measurements with the equation from Durnin & Womersley (6) for adults 50 years of age and older. Percentage body fat and subsequent fat-free mass were estimated using Siri's equation (21).

The methods for determining $\dot{V}O_2$ max are also described by Amara et al. (3). In brief, while breathing through a mouthpiece with nose clips, subjects performed an incremental ramp test to volitional or symptom-limited fatigue on a motorised treadmill. The protocol consisted of a 4 min warmup at 0.76 m s^{-1} 1.7 mph) and a 0% gradient followed by gradient and/or speed changes such that oxygen uptake increased each minute by 1-3 $ml.kg^{-1}.min^{-1}$ and the total duration of the test was between 8 and 12 min. Subjects were encouraged verbally throughout the test to perform to the limit of their tolerance. Gas exchange and ventilatory variables were analysed using a calibrated mass spectrometer (PerkinElmer MGA110) and a bidirectional turbine and volume transducer (SensorMedics VMM2A), respectively. Heart rate (HR) was monitored throughout the test using a bipolar chest lead (CM5).

The physical activity of the participants in study 2 was assessed by the Minnesota Leisure Time Physical Activity (MLTA) questionnaire (22). Amara et al. (3) chose to include only the heavy intensity activity scores in their analysis since they should theoretically provide the greatest cardiorespiratory stimulus. The heavy intensity activities were those requiring >6 METS (1 metabolic equivalent (MET) = 3·5 ml.kg^{-1}.min^{-1}). This value was age adjusted based on previous data (D. H. Paterson, unpublished) from their laboratory to account for the age associated decline in $\dot{V}O_2$max such that the male heavy intensity activity code decreased by 1.00% per year and the female heavy intensity activity code decreased by 1.04% per year above age 55 years. Each subject's heavy intensity physical activity was determined as time spent and energy expenditure (METS.year^{-1}).

Statistical methods

As discussed above, given that body mass (M) is likely to be strongly (albeit negatively) associated with $\dot{V}O_2$max (ml.kg^{-1}.min^{-1}) and allowing the possibility of a non-linear association with age, we adopted the following multiplicative model with allometric body size components for study 1 as proposed by Amara et al. (3), Nevill and Holder (17) and Nevill et al.(18),

$\dot{V}O_2$ max (ml.kg^{-1}.min^{-1}) = $M^{k_1} \cdot H^{k_2} \cdot \exp(a + b_1 \cdot age + b_2 \cdot age^2 + b_3 \cdot WC + b_4 \cdot RHR + b_5 \cdot VIGEX) \cdot \varepsilon$, (Eq1)

where 'ε' is a multiplicative, error ratio that assumes the error will be in proportion to $\dot{V}O_2$ max (ml.kg^{-1}.min^{-1}), see Figure 1.

The model (Eq. 1) can be linearized with a log transformation. A linear regression analysis on log($\dot{V}O_2$max) can then be used to estimate the unknown parameters in the log transformed model i.e., the transformed model (Eq2) is now additive that conforms with the assumptions associated with ordinary least squares:

$\log(\dot{V}O_2 max) = k_1 \cdot \log(M) + k_2 \cdot \log(H) + a + b_1.age + b_2.age^2 + b_3.\ WC + b_4\ RHR + b_5\ VIGEX + \log(\varepsilon)$, (Eq2)

where the residual errors log(ε) are assumed to be normally distributed, and the intercept "a" and the other parameters "b_i" are allowed to vary for various categorical or group differences within the population, e.g. sex.

Study 1 results using linear, additive models

Fitting a similar linear model for $\dot{V}O_2$max (ml.kg^{-1}min^{-1}) as Nes et al. (13), we obtained the following equations for $\dot{V}O_2$max,

$\dot{V}O_2$max (men) = 91.86 - (0.396 · age) - (0.212 · WC) - (0. 177· RHR) + (0.075 · VIGEX) + ε,

$\dot{V}O_2$max (women) = 69.49 - (0.267· age) - (0.212· WC) - (0.108· RHR) + (0.075· VIGEX) + ε, where the residual errors ε are assumed to be normally distributed. Note that the PA index variable, used by Nes et al. (13), has been replaced by VIGEX, the number of 20 min bouts of vigorous exercise (VIGEX), defined as activities that were > 7.5 kcal.min^{-1} or >60% of aerobic capacity reported during the four weeks prior to the exercise test. The R^2 was = 0.638 (Adjusted R^2= 0.636).

The residuals saved from the above analysis were neither normally distributed (Kolmogorov-Smirnov statistic 0.031; P<0.001; Shapiro-Wilk statistic=0.983) nor independent of either the predicted values (see Figure 1) or the key predictor variable age, i.e., the correlation between the absolute residuals vs predicted values was (r=0.173; P<0.001) and with age (r=-0.127; P<0.001). The lack of normality and the heteroscedastic residual errors observed in Figure 1 must cast serious doubt regarding the validity of the predictor variables (questioning the statistical significance of some of the fitted variables but more likely the lack of significance or absence of body mass or higher order polynomial terms, in particular an age^2 term. The systematically increasing spread of residuals observed in Figure 1 and the negative correlation between absolute residuals and age, must also cast serious doubt on the accuracy/precision of predicting $\dot{V}O_2$max especially for young/fit participants with high estimates of $\dot{V}O_2$max (where the residual errors are at their widest/greatest, see Figure 1).

` Figure 1 about here

Study 1 results using allometric, multiplicative models

The parsimonious allometric model for $\dot{V}O_2$max (ml.kg^{-1}.min^{-1}) was found to be

$$\dot{V}O_2\text{max (men)} = M^{-.436} \cdot H^{.790} \cdot \exp(5.67 - 0.000106\ \text{age}^2 - 0.0037\ \text{RHR} + 0.0017\ \text{VIGEX}) \cdot \varepsilon,$$

and

$$\dot{V}O_2\text{max (women)} = M^{-.436} \cdot H^{.790} \cdot \exp(5.397 - 0.000106\ \text{age}^2 - 0.0037\ \text{RHR} + 0.0017\ \text{VIGEX}) \cdot \varepsilon.$$

The R^2 was = 0.653 (Adjusted R^2= 0.651). The fitted age^2 parameter was -0.000106 (SE=0.000003; 95% CI -0.000112 to -0.0000995). The age and waist (WC) terms were both not significant (P>0.05). The residuals saved from the above analysis were normally distributed (Kolmogorov-Smirnov statistic 0.021; P=0.064; Shapiro-Wilk statistic 0.997) and acceptably independent of either the predicted values (see Figure 2) and age, i.e. the correlation between the absolute residuals vs predicted values (log-transformed) was (r=-0.048; P=0.044) and vs age (r=0.033; P=0.169).

Figure 2 about here

The negative age^2 term within an exponential function, is now biologically sound. The model now predicts the age decline of $\dot{V}O_2$max will follow the right-hand side of the bell-shaped normal distribution type curve, see Figure 3, where the slope of age decline in $\dot{V}O_2$max is flat/zero at zero years (i.e. it reaches a plateau), and as age increases to old age, $\dot{V}O_2$max tends towards a zero asymptote, i.e., it can never become negative unlike the negative linear age decline proposed and fitted by Nes et al. (13).

Figure 3 about here

Study 2 results using linear, additive models

Fitting a linear model for $\dot{V}O_2$ max (ml.kg^{-1}.min^{-1}) as proposed by Nes et al. (13) but using the variables available to Amara et al. (3) plus body mass (for the reasons described in the introduction), we obtained the following equations for $\dot{V}O_2$ max,

$$\dot{V}O_2 \text{max (men)} = 51.38 - (0.385 \cdot \text{age}) + (0.357 \cdot \text{FFM}) - (0. 298 \cdot \text{M}) + (0.006 \cdot \text{PA}) + \varepsilon,$$

$$\dot{V}O_2 \text{max (women)} = 41.27 - (0.258 \cdot \text{age}) + (0.357 \cdot \text{FFM}) - (0. 298 \cdot \text{M}) + (0.006 \cdot \text{PA}) + \varepsilon,$$

where the residual errors ε are assumed to be normally distributed. Note that the PA index variable, used by Nes et al. (13), has been replaced by the results from the Minnesota Leisure Time Physical Activity (MLTA) questionnaire (22). The R^2 was = 0.469 (Adjusted R^2= 0.456).

As in Study 1, the residuals saved from the linear, additive model were not normally distributed (Kolmogorov-Smirnov statistic 0.067; $P=0.007$; Shapiro-Wilk statistic=0.965; $P<0.001$). The lack of normality must cast doubt regarding the validity of the predictor variables (questioning the statistical significance of some of the fitted variables but more likely the lack of significance or absence of a higher order polynomial terms, in particular the age^2 term.

Study 2 results using allometric, multiplicative models

The parsimonious allometric model for $\dot{V}O_2$ max (ml.kg^{-1}.min^{-1}) was found to be

$$\dot{V}O_2 \text{max (men)} = \text{M}^{-.872} \cdot \text{FFM}^{.679} \cdot \text{PA}^{.025} \cdot \exp(4.57 - 0.00011.\text{age}^2) \cdot \varepsilon.$$

$$\dot{V}O_2 \text{max (women)} = \text{M}^{-.872} \cdot \text{FFM}^{.679} \cdot \text{PA}^{.025} \cdot \exp(4.47 - 0.00011.\text{age}^2) \cdot \varepsilon.$$

The R^2 was = 0.491 (Adjusted R^2= 0.481). The fitted age^2 parameter was -0.00011 (SE=0.00001; 95% CI -0.000124 to -0.000087) and as in Study 1, the linear age term not significant (P>0.05). The residuals saved from the above analysis were acceptably normally distributed (Kolmogorov-Smirnov statistic 0.031; P>0.200; Shapiro-Wilk statistic 0.995; P=0.546).

The goodness of fit of the competing linear and allometric models.

Clearly, since the models are not nested or hierarchical, a direct comparison between two competing model forms (linear vs allometric) is not possible using traditional criteria such as the residual sum-of-squares, the standard error and the coefficient of determination (R^2). However, Nevill and Holder (16) and Nevill et al. (18) chose the maximum likelihood criterion and the Akaike Information Criteria (AIC) as their standard criterion of model assessment (quality of fit) that does not require the competing models to be either nested or hierarchical.

A simple modification of the maximum log likelihood criterion is able to produce the Akaike Information Criteria (AIC= -2×(maximum log-likelihood) + 2×(number of parameters fitted)) that would take into account the different number of fitted parameters in the two model structures to be compared, see goodness-of-fit data from both studies 1 and 2 (Table 1).

Table 1 about here

Discussion

Based on the concerns discussed in the introduction, the results from both studies confirm that the allometric models proposed by Amara et al. (3), Nevill and Holder (17) and Nevill et al. (18) (Eq1) performed better than the linear model proposed by Nes et al. (13) in all three major areas of concern.

The goodness of fit is superior when fitting allometric models. The R^2 was greater but more importantly the maximum log-likelihood (MLL) was also greater, and the Akaike Information Criterion (AIC) was smaller, compared with the linear additive models (see Table 1).

Furthermore the residuals from both studies saved from fitting the linear, additive models violate the assumption of normality and reveal evidence of heteroscedastic errors associated with both the predicted values and age. This will seriously question, 1) the selection (or more importantly the non-selection) of possible predictor variables, and 2) the accuracy when predicting $\dot{V}O_2$max, in particular, of the young and fit individuals in Study 1 (who had the greatest predicted $\dot{V}O_2$max) where the residual errors were at their greatest (see Figure 1). In contrast, the log-transformed allometric model resulted in residuals from both studies that were normally distributed and in the case of study 1, independent of both the predicted values and the key predictor variable age. When we fitted the quadric in age in both studies, the parsimonious solution identified only an age^2 term within an exponential function as the appropriate model to describe the age decline in $\dot{V}O_2$max (i.e. the right-hand side of a normal, bell-shaped frequency distribution curve). Note that since the age^2 parameters in the allometric models fitted to study 1 and study 2 were very similar, the curvilinear decline in age will be almost identical (Figure 3). These models, see Figure 3, are now biologically sound and interpretable. To illustrate this based on the results of Study 1, compare the systematic errors likely if we use the linear model proposed by Nes et al. (13). The linear model predicts the age decline as 2.96 and 2.47 (ml.kg^{-1}.min^{-1}) per decade (for all ages and decades) for men and women respectively. However, the more realistic age decline (see for example Astrand and Rodahl (5) Figure 7-15 on page 337) using the allometric model (see Figure 1) was only 2.58 and 1.80 (ml.kg^{-1}.min^{-1}) for men and women in their 20’s, but almost double that rate, found to be 4.66 and 3.25 (ml.kg^{-1}.min^{-1}) for men and women in their 60’s.

Further support for the allometric model (1) comes from the fitted stature/height and body mass exponents obtained in Study 1, found to be $M^{-436} \cdot H^{.790}$. Nevill et al. (19) anticipated that when researchers calculate $\dot{V}O_2$ max (ml.kg^{-1}.min^{-1}) by dividing $\dot{V}O_2$ max (l.min^{-1}) by body mass (M), the ratio “over scales” leaving $\dot{V}O_2$ max (ml.kg^{-1}.min^{-1}) theoretically proportional to $M^{-0.33}$. The fitted body-mass exponent (-0.436; SE = 0.027) was greater than that anticipated (-0.333) but confirms the need for its inclusion and the concern by its absence from the Nes et al. (13) linear models. However, when taken together, the two allometric body-size components can be re-arranged as $(H^{1.81}.M^{-1})^{0.436}$. This too has a sound biological interpretation, as the resulting index is a stature-to-body mass ratio that closely approximates the inverse BMI (iBMI), thought to be a measure of leanness (15, 20) . Clearly having a greater lean body mass index (LBMI), as described by Nevill and Holder (15), should also be strongly associated with predicting $\dot{V}O_2$ max (ml.kg^{-1}.min^{-1}).

A similar “leanness” ratio was identified in Study 2. The fitted fat-free mass and body mass exponents were found to be $M^{-.872} \cdot FFM^{.679}$. Again taken together, the two allometric body-size components can also be re-arranged as $(FFM^{.779}.M^{-1})^{0.872}$. The resulting fat-free mass-to-body mass ratio is physiologically similar to the ratio reported in study 1, as a greater FFM is a strong determinant of $\dot{V}O_2$ max (ml.kg^{-1}.min^{-1}) (8).

We acknowledge that the current study is not without limitations. The fact that we have been able to demonstrate the benefits of modelling $\dot{V}O_2$ max using allometric models using just two data sets is not ideal. Clearly future research should explore the benefits of allometric models using many more $\dot{V}O_2$ max data sets especially ones where linear, additive models such as those reported by Nes et al. (13) have been adopted/reported.

In summary, the quality of fit associated with predicting $\dot{V}O_2$ max (ml.kg^{-1}.min^{-1}) using allometric models in both studies was superior to linear, additive models based on all criteria (R^2, maximum log-likelihood and AIC). Furthermore, it would appear that by fitting the linear, additive models proposed by Nes et al. (13), systematic errors are likely when predicting $\dot{V}O_2$ max (ml.kg^{-1}.min^{-1}), see Figure 3. The linear models fitted to study 1 will systematically over-estimate $\dot{V}O_2$ max for participants in their 20's and systematically under-estimate $\dot{V}O_2$ max for participants in their 60's. The failure by Nes et al. (13) to identity curvature in their age decline or the presence of a body-mass power function term might well have been explained by examining the residuals saved from their analyses. The residuals from the linear regression analysis from both study 1 and study 2 were neither normally distributed, nor independent of the predicted values and key predictor variables such as age. This will almost certainly explain their possible invalid inclusion of some terms, or more likely the absence of other key variables such as body mass and the quadratic term in age^2, both crucially identified using the allometric models proposed be Nevill and co-workers. Not only does the curvilinear age decline within an exponential function follow a more realistic age decline (right-hand side of the bell-shaped curve, see Astrand and Rodahl (5) Figure 7-15 on page 337), but the allometric models also identified a stature-to-body-mass ratio (study 1) or a fat-free-mass-to-body-mass ratio (study 2), both known to be associated with leanness, new insights that lead to a more plausible, biologically sound and interpretable model when predicting $\dot{V}O_2$ max (ml.kg^{-1}.min^{-1}).

Acknowledgments

The authors have no conflicts of interest. The results of the study are presented clearly, honestly, and without fabrication, falsification, or inappropriate data manipulation. The results of the present study do not constitute endorsement by ACSM.

References

1. Ainsworth BE, Caspersen CJ, Matthews CE, Masse LC, Baranowski T and Zhu W. Recommendations to improve the accuracy of estimates of physical activity derived from self-report. *J.Phys.Act.Health.* 2012; 90: S76 - S84.
2. Allied Dunbar National Fitness Survey. *A Report on Activity Patterns and Fitness Levels: Main Findings*. (Sports Council and Health Education Authority, London) 1992. 1-398 p.
3. Amara CE, Koval JJ, Johnson PJ, Paterson DH, Winter EM and Cunningham DA. Modelling the influence of *fat*-free mass and physical activity on the decline in maximal oxygen uptake with age in older humans. *Exp Physiol* 2000;85: 877-886.
4. Arena R, Myers J, Williams MA, Gulati M, Kligfield P, Balady GJ, Collins E and Fletcher G. Assessment of functional capacity in clinical and research settings. A scientific statement from the American Heart Association Committee on Exercise, Rehabilitation and Prevention of the Council on Clinical Cardiology and the Council on Cardiovascular Nursing. *Circulation*. 2007; 116: 329-343.

5. Astrand P-O and Rodahl K. Textbook of work physiology, 3rd ed. New York: McGraw-Hill; 1986. 337 p.
6. Durnin, J. V. G. A. & Womersley, J. Body fat assessed from total body density and its estimation from skinfold thickness: measurements on 481 men and women aged from 16 to 72 years. British Journal of Nutrition. 1974; 32: 77-97.
7. Fentem PH, Collins MF, Tuxworth W, Walker A, Hoinville E, Cooke C, Harries U J and Rayson M. Allied Dunbar National Fitness Survey. Technical Report. London: Sports Council 1994. 1-294 p.
8. Goran M, Fields DA, Hunter GR, Herd SL and Weinsier RL. Total body fat does not influence maximal aerobic capacity. *Int. J. Obesity.* 2000; 24: 841 - 848.
9. Grant S, Corbett K, Amjad AM, Wilson J and Aitchison T. A comparison of methods of predicting maximum oxygen uptake. *Br. J. Sports. Med.* 1995; 29: 147-152.
10. Hawkins SA and Wiswell RA. Rate and Mechanism of VO_2max Decline with Aging. Sports Med. 2003; 33: 877-88.
11. Jones N and Campbell, E. *Clinical exercise testing*. Philadelphia. WB: Saunders Co., 1975. 214 p.
12. Lange-Anderson K, Shephard RH, Denoln H, Varnauskas E and Masironi R. *Fundamentals of Exercise Testing*. Geneva, World Health Organisation, 1971. 134 p.
13. Nes BM, Janszky I, Vatten LJ, Nilsen TIL, Aspenes ST, Wislkff U. Estimating VO2peak from a Non-exercise Prediction Model: The HUNT Study, Norway. *Med. Sci. Sports Exerc.* 2011; 43: 2024–30.
14. Nes BM, Vatten LJ, Nauman J, Janszky I, and Wislkff U. A Simple Non-exercise Model of Cardiorespiratory Fitness Predicts Long-Term Mortality. Med Sci Sports Exerc. 2014; 46: 1159–65.

15. Nevill AM and Holder RL. Body Mass Index; a measure of fatness or leanness?. The British Journal of Nutrition, 1995; 73: 507-516.

16. Nevill AM and Holder RL. Modelling maximum oxygen uptake - A case study in non-linear regression model formulation and comparison. Journal of the Royal Statistical Society, Series C. 1994; 43: 653-66.

17. Nevill AM, Holder RL. Scaling, Normalizing and 'per ratio' standards; an allometric modeling approach. *J. Appl Physiol.* 1995; 79: 1027-31.

18. Nevill AM, Holder RL, Baxter-Jones A, Round J, and Jones DA. Modeling developmental changes in strength and aerobic power in children. J Appl Physiol. 1998; 84: 963-70.

19. Nevill AM, Ramsbottom R, and Williams C. Scaling physiological measurements for individuals of different body size. Eur J Appl Physiol.1992;65: 110-7.

20. Nevill AM, Stavropoulos-Kalinoglou A., Metsios G.S., Koutedakis Y., Holder R.L., Kitas G.D. Mohammed M.A (2011) Inverted BMI rather than BMI is a better proxy for percentage of body fat. Ann Hum Biol. 2012; 38: 681-4.

21. Siri, W. E. (1961). Body composition from fluid space and density: analysis of methods. In Technique for Measuring Body Composition, ed. Brozek, J. & Henschel, A. National Academy of Sciences, Washington, DC, USA. 223-244 p.

22. Taylor, H. L., Jacobs, D. R. J., Schucker, B., Knudsen, J., Leon, A. S. & Debacker, G. A questionnaire for the assessment of leisure time physical activities. Journal of Chronic Diseases. 1978; 31: 741—755.

23. Winter E.M. and Nevill A.M. Scaling: Adjusting for differences in body size. In "Kinanthropometry and Exercise Physiology Laboratory Manual for Tests, Procedures and Data, 3rd ed." (eds) R. G. Eston and T. Reilly, Abingdon, Oxon, Routledge; 2009. 300-314 p.

Legends to Tables

Table 1. The maximum log-likelihood (MLL) and Akaike Information Criterion (AIC) together with the number of fitted parameters for the competing models to predict $\dot{V}O_2$ max, results from Studies 1 and 2.

Legends to figures

Figure 1. Residuals versus predicted $\dot{V}O_2$ max (ml.kg^{-1}.min^{-1}) obtained using the linear, additive model proposed by Nes et al. (13).

Figure 2. Residuals versus predicted log-transformed $\dot{V}O_2$ max (ml.kg^{-1}.min^{-1}) obtained using the allometric model (Eq1) proposed/adopted from Amara et al. (3), Nevill and Holder (17) and Nevill et al.(18).

Figure 3. The age decline of $\dot{V}O_2$ max (ml.kg^{-1}.min^{-1}) predicted from the allometric model (Eq1) proposed/adopted by Amara et al. (3), Nevill and Holder (17) and Nevill et al.(18).