

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

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

3 Authors: Alan M. Nevill¹ and Carlton B. Cooke²

4

5 1. Faculty of Education, Health and Wellbeing, University of Wolverhampton, Walsall
6 Campus, Walsall, WS1 3BD, UK

7 2. School of Social and Health Sciences, Leeds Trinity University, Leeds, LS18 5HD,
8 UK

9 **Address for correspondence:**

10 Professor Alan M. Nevill, Ph.D.

11 University of Wolverhampton,

12 Faculty of Education, Health and Wellbeing,

13 Walsall Campus

14 Gorway Road

15 Walsall, WS1 3BD

16 Tel: +44 (0)1902 322838

17 Fax: +44 (0)1902 322894

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

19 **Word Count: 4972**

20

21 **Abstract**

22 **Purpose:** To compare the accuracy and goodness-of-fit of two competing models (linear
23 versus allometric) when estimating $\dot{V}O_2 \text{ max}$ ($\text{ml.kg}^{-1}.\text{min}^{-1}$) using non-exercise prediction
24 models. **Methods:** The two competing models were fitted to the $\dot{V}O_2 \text{ max}$ ($\text{ml.kg}^{-1}.\text{min}^{-1}$)
25 data taken from two previously published studies. Study 1 (the Allied Dunbar National
26 Fitness Survey, ADNFS), recruited 1732 randomly selected healthy participants, aged 16
27 years and over, from thirty English parliamentary constituencies. Estimates of $\dot{V}O_2 \text{ max}$ were
28 obtained using a progressive incremental test on a motorized treadmill. In Study 2 (3),
29 maximal oxygen uptake was measured directly during a fatigue limited treadmill test in older
30 men ($n = 152$) and women ($n = 146$) aged 55 to 86 years. **Results:** In both studies, the
31 quality-of-fit associated with estimating $\dot{V}O_2 \text{ max}$ ($\text{ml.kg}^{-1}.\text{min}^{-1}$) was superior using
32 allometric rather than linear (additive) models based on all criteria (R^2 , maximum log-
33 likelihood and AIC). Results suggest that linear models will systematically over-estimate
34 $\dot{V}O_2 \text{ max}$ for participants in their 20's and under-estimate $\dot{V}O_2 \text{ max}$ for participants in their
35 60's and older. The residuals saved from the linear models were neither normally distributed,
36 nor independent of the predicted values nor age. This will probably explain the absence of a
37 key quadratic age^2 term in the linear models, crucially identified using allometric models. Not
38 only does the curvilinear age decline within an exponential function follow a more realistic
39 age decline (the right-hand side of a bell-shaped curve), but the allometric models identified
40 either a stature-to-body-mass ratio (study 1) or a fat-free-mass-to-body-mass ratio (study 2),
41 both associated with leanness when estimating $\dot{V}O_2 \text{ max}$. **Conclusions:** Adopting allometric
42 models will provide more accurate predictions of $\dot{V}O_2 \text{ max}$ ($\text{ml.kg}^{-1}.\text{min}^{-1}$) using plausible,
43 biologically sound and interpretable models.

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

45

46 **Introduction**

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

53 Several studies have reported non-linear associations between $\dot{V}O_2 \text{ max}$ and age, and $\dot{V}O_2$
54 max and body mass (5, 10, 16). Hence it was surprising that Nes et al. (13) adopted a linear
55 model to estimate $\dot{V}O_2 \text{ max}$ ($\text{ml.kg}^{-1}.\text{min}^{-1}$) that was subsequently used by Nes et al. (14) to
56 predict long-term all-cause mortality and cardiovascular disease (CVD). The authors reported
57 the following linear regression models to estimate $\dot{V}O_2 \text{ max}$ ($\text{ml.kg}^{-1}.\text{min}^{-1}$) for men; $100.27 -$
58 $(0.296 \cdot \text{age}) - (0.369 \cdot \text{WC}) - (0.155 \cdot \text{RHR}) + (0.226 \cdot \text{PA-index})$, and for women; $74.74 -$
59 $(0.247 \cdot \text{age}) - (0.259 \cdot \text{WC}) - (0.114 \cdot \text{RHR}) + (0.198 \cdot \text{PA-index})$, where WC=waist
60 circumference; RHR=resting heart rate; PA index=physical activity index. The authors
61 reported that their models were unable to detect any interaction or polynomial terms, i.e. the
62 inclusion of such terms was unable “to influence the R^2 of the models appreciably”.

63 There are at least three major concerns with these linear, additive models. Firstly, both
64 models suggest a linear decline in age that has the same rate (same slope parameter) for
65 participants in the twenties, as in their fifties or sixties and in their eighties. However, there is
66 evidence in the literature that indicates a curvilinear decline in $\dot{V}O_2 \text{ max}$ with age, suggesting
67 the need for a non-linear or quadratic age term to be incorporated into the model, see Astrand
68 and Rodahl (5) Figure 7-15 on page 337, and Hawkins (10). The second concern is the

69 absence of a weight/body-mass term in both models. Nevill et al. (19) and Astrand and
70 Rodahl (5) in their Figure 9-4 on page 400, reported a strong negative association between
71 $\dot{V}O_2 \text{ max (ml.kg}^{-1}.\text{min}^{-1})$ and body mass. This is because absolute $\dot{V}O_2 \text{ max (l.min}^{-1})$ scales
72 to, or is associated with body mass ($M^{0.67}$), and hence when researchers calculate $\dot{V}O_2 \text{ max}$
73 ($\text{ml.kg}^{-1}.\text{min}^{-1}$), by dividing $\dot{V}O_2 \text{ max (l.min}^{-1})$ by body mass (M), the resulting ratio “over-
74 scales” leaving $\dot{V}O_2 \text{ max (ml.kg}^{-1}.\text{min}^{-1})$ proportional to $M^{0.33}$. This non-linear association
75 with mass should have been considered by Nes et al. (13). Incorporating a power-function
76 body-mass term as a predictor in both models is likely to improve the accuracy when
77 predicting $\dot{V}O_2 \text{ max (ml.kg}^{-1}.\text{min}^{-1})$.

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

87 Hence the purpose of this study was to fit the same linear, additive model adopted by Nes et
88 al. (13) to both estimated and directly measure $\dot{V}O_2 \text{ max (ml.kg}^{-1}.\text{min}^{-1})$ data from two
89 previously published studies, Study 1 the Allied Dunbar National Fitness Survey (ADNFS)
90 (2, 16), and Study 2, data reported by Amara et al. (3), to compare a linear model with an
91 alternative, proportional allometric model to discover whether the latter provides, 1) a
92 superior quality of fit (using R^2 , maximum log-likelihood and AIC criterion), 2) more

93 normally distributed residuals and, 3) a more plausible, biologically sound and interpretable
94 model.

95

96

97 **Methods (study 1)**

98 All variables and measurement used in the current study have been previously described and
99 published (16) or reported in a technical report (7). Cardiopulmonary fitness or $\dot{V}O_2$ max was
100 assessed using a progressive incremental test on a motorized treadmill. In reality, the $\dot{V}O_2$
101 max measurements are estimates based on the linear relationship (for each subject) between
102 the oxygen cost and heart rate, recorded breath-by-breath ($n > 50$) during a sub-maximal
103 exercise test using an automated respiratory gas analyzer (Quinton Q-plex) and a diagnostic
104 electrocardiogram (Quinton Q4000). The test continued until the end of a one-minute stage in
105 which the subject's heart rate had reached 85% of estimated maximum for age ($210 -$
106 $0.65 \cdot \text{age}$, beats min^{-1}). For a given individual, the estimated $\dot{V}O_2$ max is the predicted oxygen
107 cost at an assumed maximum heart rate, taken to be $210 - 0.65 \cdot \text{age}$ (11). All submaximal tests
108 used to estimate $\dot{V}O_2$ max are associated with a standard error of prediction which is
109 typically in the range of 10% - 15% (5). One advantage of the protocol used in the Allied
110 Dunbar National Fitness Survey (2, 7) is that the $\dot{V}O_2$ of each stage was directly measured,
111 which eliminates variations in mechanical efficiency associated with the use of workload.
112 However, the accuracy of the method is still dependent on the variability in predicted
113 maximum heart rate, which in normal adult participants has been shown to have a standard
114 deviation of 10-12 $\text{beats} \cdot \text{min}^{-1}$ (4). The validity of the linear extrapolation method described
115 by Lange-Anderson et al. (12) to predict $\dot{V}O_2$ max using measured submaximal $\dot{V}O_2$ values

116 to a predicted maximum heart rate has been assessed against directly determined treadmill
117 $\dot{V}O_2$ max, where it was shown to under-predict by 13% with an SE of $1.4 \text{ ml.kg}^{-1}\text{min}^{-1}$ (9),
118 which is within the range typically reported for estimations of $\dot{V}O_2$ max.

119

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

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

137

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

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

146 **Methods (study 2)**

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

159 The methods for determining $\dot{V}O_2$ max are also described by Amara et al. (3). In brief, while
160 breathing through a mouthpiece with nose clips, subjects performed an incremental ramp test
161 to volitional or symptom-limited fatigue on a motorised treadmill. The protocol consisted of a
162 4 min warmup at 0.76 m s⁻¹ (1.7 mph) and a 0% gradient followed by gradient and/or speed
163 changes such that oxygen uptake increased each minute by 1-3 ml.kg⁻¹.min⁻¹ and the total

164 duration of the test was between 8 and 12 min. Subjects were encouraged verbally throughout
165 the test to perform to the limit of their tolerance. Gas exchange and ventilatory variables were
166 analysed using a calibrated mass spectrometer (PerkinElmer MGA110) and a bidirectional
167 turbine and volume transducer (SensorMedics VMM2A), respectively. Heart rate (HR) was
168 monitored throughout the test using a bipolar chest lead (CM5).

169 The physical activity of the participants in study 2 was assessed by the Minnesota Leisure
170 Time Physical Activity (MLTA) questionnaire (22). Amara et al. (3) chose to include only
171 the heavy intensity activity scores in their analysis since they should theoretically provide the
172 greatest cardiorespiratory stimulus. The heavy intensity activities were those requiring >6
173 METS (1 metabolic equivalent (MET) = $3.5 \text{ ml.kg}^{-1}.\text{min}^{-1}$). This value was age adjusted
174 based on previous data (D. H. Paterson, unpublished) from their laboratory to account for the
175 age associated decline in $\dot{V}O_2 \text{ max}$ such that the male heavy intensity activity code decreased
176 by 1.00% per year and the female heavy intensity activity code decreased by 1.04% per year
177 above age 55 years. Each subject's heavy intensity physical activity was determined as time
178 spent and energy expenditure ($\text{METS}.\text{year}^{-1}$).

179

180 **Statistical methods**

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

186 $\dot{V}O_2 \text{ max (ml.kg}^{-1}.\text{min}^{-1}) = M^{k_1} \cdot H^{k_2} \cdot \exp(a + b_1 \cdot \text{age} + b_2 \cdot \text{age}^2 + b_3 \cdot \text{WC} + b_4 \cdot \text{RHR} +$
 187 $b_5 \cdot \text{VIGEX}) \cdot \varepsilon,$ (Eq1)

188 where ‘ ε ’ is a multiplicative, error ratio that assumes the error will be in proportion to $\dot{V}O_2$
 189 $\text{max (ml.kg}^{-1}.\text{min}^{-1})$, see Figure 1.

190 The model (Eq. 1) can be linearized with a log transformation. A linear regression analysis on
 191 $\log(\dot{V}O_2 \text{ max})$ can then be used to estimate the unknown parameters in the log transformed
 192 model i.e., the transformed model (Eq2) is now additive that conforms with the assumptions
 193 associated with ordinary least squares:

194 $\log(\dot{V}O_2 \text{ max}) = k_1 \cdot \log(M) + k_2 \cdot \log(H) + a + b_1 \cdot \text{age} + b_2 \cdot \text{age}^2 + b_3 \cdot \text{WC} + b_4 \cdot \text{RHR} + b_5 \cdot \text{VIGEX} +$
 195 $\log(\varepsilon),$ (Eq2)

196 where the residual errors $\log(\varepsilon)$ are assumed to be normally distributed, and the intercept “a”
 197 and the other parameters “ b_i ” are allowed to vary for various categorical or group differences
 198 within the population, e.g. sex.

199

200 **Study 1 results using linear, additive models**

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

203 $\dot{V}O_2 \text{ max (men) = 91.86 - (0.396 \cdot \text{age}) - (0.212 \cdot \text{WC}) - (0.177 \cdot \text{RHR}) + (0.075 \cdot \text{VIGEX}) +$
 204 $\varepsilon,$

205 $\dot{V}O_2 \text{ max (women) = 69.49 - (0.267 \cdot \text{age}) - (0.212 \cdot \text{WC}) - (0.108 \cdot \text{RHR}) + (0.075 \cdot \text{VIGEX}) +$
 206 $\varepsilon,$

207 where the residual errors ε are assumed to be normally distributed. Note that the PA index
208 variable, used by Nes et al. (13), has been replaced by VIGEX, the number of 20 min bouts of
209 vigorous exercise (VIGEX), defined as activities that were $> 7.5 \text{ kcal}\cdot\text{min}^{-1}$ or $>60\%$ of
210 aerobic capacity reported during the four weeks prior to the exercise test. The R^2 was = 0.638
211 (Adjusted $R^2= 0.636$).

212

213 The residuals saved from the above analysis were neither normally distributed (Kolmogorov-
214 Smirnov statistic 0.031; $P<0.001$; Shapiro-Wilk statistic=0.983) nor independent of either the
215 predicted values (see Figure 1) or the key predictor variable age, i.e., the correlation between
216 the absolute residuals vs predicted values was ($r=0.173$; $P<0.001$) and with age ($r=-0.127$;
217 $P<0.001$). The lack of normality and the heteroscedastic residual errors observed in Figure 1
218 must cast serious doubt regarding the validity of the predictor variables (questioning the
219 statistical significance of some of the fitted variables but more likely the lack of significance
220 or absence of body mass or higher order polynomial terms, in particular an age^2 term. The
221 systematically increasing spread of residuals observed in Figure 1 and the negative
222 correlation between absolute residuals and age, must also cast serious doubt on the
223 accuracy/precision of predicting $\dot{V}O_2 \text{ max}$ especially for young/fit participants with high
224 estimates of $\dot{V}O_2 \text{ max}$ (where the residual errors are at their widest/greatest, see Figure 1).

225 ` Figure 1 about here

226 **Study 1 results using allometric, multiplicative models**

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

228 $\dot{V}O_2\text{max (men)} = M^{-.436} \cdot H^{.790} \cdot \exp (5.67 - 0.000106 \text{ age}^2 - 0.0037 \text{ RHR} + 0.0017 \text{ VIGEX})$

229 $\cdot \varepsilon,$

230 and

231 $\dot{V}O_2\text{max (women)} = M^{-.436} \cdot H^{.790} \cdot \exp (5.397 - 0.000106 \text{ age}^2 - 0.0037 \text{ RHR} + 0.0017$

232 $\text{VIGEX}) \cdot \varepsilon.$

233

234 The R^2 was = 0.653 (Adjusted R^2 = 0.651). The fitted age^2 parameter was -0.000106

235 (SE=0.000003; 95% CI -0.000112 to -0.0000995). The age and waist (WC) terms were both

236 not significant ($P>0.05$). The residuals saved from the above analysis were normally

237 distributed (Kolmogorov-Smirnov statistic 0.021; $P=0.064$; Shapiro-Wilk statistic 0.997) and

238 acceptably independent of either the predicted values (see Figure 2) and age, i.e. the

239 correlation between the absolute residuals vs predicted values (log-transformed) was ($r=-$

240 0.048; $P=0.044$) and vs age ($r=0.033$; $P=0.169$).

241

242 Figure 2 about here

243

244 The negative age^2 term within an exponential function, is now biologically sound. The model

245 now predicts the age decline of $\dot{V}O_2\text{max}$ will follow the right-hand side of the bell-shaped

246 normal distribution type curve, see Figure 3, where the slope of age decline in $\dot{V}O_2\text{max}$ is

247 flat/zero at zero years (i.e. it reaches a plateau), and as age increases to old age, $\dot{V}O_2\text{max}$

248 tends towards a zero asymptote, i.e., it can never become negative unlike the negative linear

249 age decline proposed and fitted by Nes et al. (13).

250

Figure 3 about here

251 **Study 2 results using linear, additive models**

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

255
$$\dot{V}O_2 \text{ max (men)} = 51.38 - (0.385 \cdot \text{age}) + (0.357 \cdot \text{FFM}) - (0.298 \cdot \text{M}) + (0.006 \cdot \text{PA}) + \varepsilon,$$

256
$$\dot{V}O_2 \text{ max (women)} = 41.27 - (0.258 \cdot \text{age}) + (0.357 \cdot \text{FFM}) - (0.298 \cdot \text{M}) + (0.006 \cdot \text{PA}) + \varepsilon,$$

257 where the residual errors ε are assumed to be normally distributed. Note that the PA index
258 variable, used by Nes et al. (13), has been replaced by the results from the Minnesota Leisure
259 Time Physical Activity (MLTA) questionnaire (22). The R^2 was = 0.469 (Adjusted R^2 =
260 0.456).

261 As in Study 1, the residuals saved from the linear, additive model were not normally
262 distributed (Kolmogorov-Smirnov statistic 0.067; $P=0.007$; Shapiro-Wilk statistic=0.965;
263 $P<0.001$). The lack of normality must cast doubt regarding the validity of the predictor
264 variables (questioning the statistical significance of some of the fitted variables but more
265 likely the lack of significance or absence of a higher order polynomial terms, in particular the
266 age^2 term.

267 **Study 2 results using allometric, multiplicative models**

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

269
$$\dot{V}O_2 \text{ max (men)} = M^{-.872} \cdot \text{FFM}^{.679} \cdot \text{PA}^{.025} \cdot \exp(4.57 - 0.00011 \cdot \text{age}^2) \cdot \varepsilon.$$

270
$$\dot{V}O_2 \text{ max (women)} = M^{-.872} \cdot \text{FFM}^{.679} \cdot \text{PA}^{.025} \cdot \exp(4.47 - 0.00011 \cdot \text{age}^2) \cdot \varepsilon.$$

271 The R^2 was = 0.491 (Adjusted R^2 = 0.481). The fitted age² parameter was -0.00011
272 (SE=0.00001; 95% CI -0.000124 to -0.000087) and as in Study 1, the linear age term not
273 significant ($P>0.05$). The residuals saved from the above analysis were acceptably normally
274 distributed (Kolmogorov-Smirnov statistic 0.031; $P>0.200$; Shapiro-Wilk statistic 0.995;
275 $P=0.546$).

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

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

284

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

289

290 **Table 1 about here**

291

292 **Discussion**

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

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

301 Furthermore the residuals from both studies saved from fitting the linear, additive models
302 violate the assumption of normality and reveal evidence of heteroscedastic errors associated
303 with both the predicted values and age. This will seriously question, 1) the selection (or more
304 importantly the non-selection) of possible predictor variables, and 2) the accuracy when
305 predicting $\dot{V}O_2$ max, in particular, of the young and fit individuals in Study 1 (who had the
306 greatest predicted $\dot{V}O_2$ max) where the residual errors were at their greatest (see Figure 1). In
307 contrast, the log-transformed allometric model resulted in residuals from both studies that
308 were normally distributed and in the case of study 1, independent of both the predicted values
309 and the key predictor variable age. When we fitted the quadric in age in both studies, the
310 parsimonious solution identified only an age^2 term within an exponential function as the
311 appropriate model to describe the age decline in $\dot{V}O_2$ max (i.e. the right-hand side of a
312 normal, bell-shaped frequency distribution curve). Note that since the age^2 parameters in the
313 allometric models fitted to study 1 and study 2 were very similar, the curvilinear decline in
314 age will be almost identical (Figure 3). These models, see Figure 3, are now biologically
315 sound and interpretable. To illustrate this based on the results of Study 1, compare the
316 systematic errors likely if we use the linear model proposed by Nes et al. (13). The linear

317 model predicts the age decline as 2.96 and 2.47 ($\text{ml.kg}^{-1}.\text{min}^{-1}$) per decade (for all ages and
318 decades) for men and women respectively. However, the more realistic age decline (see for
319 example Astrand and Rodahl (5) Figure 7-15 on page 337) using the allometric model (see
320 Figure 1) was only 2.58 and 1.80 ($\text{ml.kg}^{-1}.\text{min}^{-1}$) for men and women in their 20's, but almost
321 double that rate, found to be 4.66 and 3.25 ($\text{ml.kg}^{-1}.\text{min}^{-1}$) for men and women in their 60's.

322

323 Further support for the allometric model (1) comes from the fitted stature/height and body
324 mass exponents obtained in Study 1, found to be $M^{-0.436} \cdot H^{0.790}$. Nevill et al. (19) anticipated
325 that when researchers calculate $\dot{V}O_2 \text{ max}$ ($\text{ml.kg}^{-1}.\text{min}^{-1}$) by dividing $\dot{V}O_2 \text{ max}$ (l.min^{-1}) by
326 body mass (M), the ratio “over scales” leaving $\dot{V}O_2 \text{ max}$ ($\text{ml.kg}^{-1}.\text{min}^{-1}$) theoretically
327 proportional to $M^{-0.33}$. The fitted body-mass exponent (-0.436; SE = 0.027) was greater than
328 that anticipated (-0.333) but confirms the need for its inclusion and the concern by its absence
329 from the Nes et al. (13) linear models. However, when taken together, the two allometric
330 body-size components can be re-arranged as $(H^{1.81} \cdot M^{-1})^{0.436}$. This too has a sound biological
331 interpretation, as the resulting index is a stature-to-body mass ratio that closely approximates
332 the inverse BMI (iBMI), thought to be a measure of leanness (15, 20). Clearly having a
333 greater lean body mass index (LBMI), as described by Nevill and Holder (15), should also be
334 strongly associated with predicting $\dot{V}O_2 \text{ max}$ ($\text{ml.kg}^{-1}.\text{min}^{-1}$).

335 A similar “leanness” ratio was identified in Study 2. The fitted fat-free mass and body mass
336 exponents were found to be $M^{-0.872} \cdot \text{FFM}^{0.679}$. Again taken together, the two allometric body-
337 size components can also be re-arranged as $(\text{FFM}^{0.779} \cdot M^{-1})^{0.872}$. The resulting fat-free mass-to-
338 body mass ratio is physiologically similar to the ratio reported in study 1, as a greater FFM is
339 a strong determinant of $\dot{V}O_2 \text{ max}$ ($\text{ml.kg}^{-1}.\text{min}^{-1}$) (8).

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

345

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

364 new insights that lead to a more plausible, biologically sound and interpretable model when
365 predicting $\dot{V}O_2$ max (ml.kg⁻¹.min⁻¹).

366 **Acknowledgments**

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

370

371

372 **References**

- 373 1. Ainsworth BE, Caspersen CJ, Matthews CE, Masse LC, Baranowski T and Zhu W.
374 Recommendations to improve the accuracy of estimates of physical activity derived
375 from self-report. *J.Phys.Act.Health.* 2012; 90: S76 - S84.
- 376 2. Allied Dunbar National Fitness Survey. *A Report on Activity Patterns and Fitness*
377 *Levels: Main Findings.* (Sports Council and Health Education Authority, London)
378 1992. 1-398 p.
- 379 3. Amara CE, Koval JJ, Johnson PJ, Paterson DH, Winter EM and Cunningham DA.
380 Modelling the influence of *fat*-free mass and physical activity on the decline in
381 maximal oxygen uptake with age in older humans. *Exp Physiol* 2000;85: 877-886.
- 382 4. Arena R, Myers J, Williams MA, Gulati M, Kligfield P, Balady GJ, Collins E and
383 Fletcher G. Assessment of functional capacity in clinical and research settings. A
384 scientific statement from the American Heart Association Committee on Exercise,
385 Rehabilitation and Prevention of the Council on Clinical Cardiology and the Council
386 on Cardiovascular Nursing. *Circulation.* 2007; 116: 329-343.

- 387 5. Astrand P-O and Rodahl K. Textbook of work physiology, 3rd ed. New York:
388 McGraw-Hill; 1986. 337 p.
- 389 6. Durnin, J. V. G. A. & Womersley, J. Body fat assessed from total body density and its
390 estimation from skinfold thickness: measurements on 481 men and women aged from
391 16 to 72 years. *British Journal of Nutrition*. 1974; 32: 77-97.
- 392 7. Fentem PH, Collins MF, Tuxworth W, Walker A, Hoinville E, Cooke C, Harries U J
393 and Rayson M. Allied Dunbar National Fitness Survey. Technical Report. London:
394 Sports Council 1994. 1-294 p.
- 395 8. Goran M, Fields DA, Hunter GR, Herd SL and Weinsier RL. Total body fat does not
396 influence maximal aerobic capacity. *Int. J. Obesity*. 2000; 24: 841 - 848.
- 397 9. Grant S, Corbett K, Amjad AM, Wilson J and Aitchison T. A comparison of methods
398 of predicting maximum oxygen uptake. *Br. J. Sports. Med.* 1995; 29: 147-152.
- 399 10. Hawkins SA and Wiswell RA. Rate and Mechanism of VO₂max Decline with Aging.
400 *Sports Med.* 2003; 33: 877-88.
- 401 11. Jones N and Campbell, E. *Clinical exercise testing*. Philadelphia. WB: Saunders Co.,
402 1975. 214 p.
- 403 12. Lange-Anderson K, Shephard RH, Denolin H, Varnauskas E and Masironi R.
404 *Fundamentals of Exercise Testing*. Geneva, World Health Organisation, 1971. 134 p.
- 405 13. Nes BM, Janszky I, Vatten LJ, Nilsen TIL, Aspenes ST, Wislckff U. Estimating
406 VO₂peak from a Non-exercise Prediction Model: The HUNT Study, Norway. *Med.*
407 *Sci. Sports Exerc.* 2011; 43: 2024–30.
- 408 14. Nes BM, Vatten LJ, Nauman J, Janszky I, and Wislckff U. A Simple Non-exercise
409 Model of Cardiorespiratory Fitness Predicts Long-Term Mortality. *Med Sci Sports*
410 *Exerc.* 2014; 46: 1159–65.

- 411 15. Nevill AM and Holder RL. Body Mass Index; a measure of fatness or leanness?. The
412 British Journal of Nutrition, 1995; 73: 507-516.
- 413 16. Nevill AM and Holder RL. Modelling maximum oxygen uptake - A case study in
414 non-linear regression model formulation and comparison. Journal of the Royal
415 Statistical Society, Series C. 1994; 43: 653-66.
- 416 17. Nevill AM, Holder RL. Scaling, Normalizing and 'per ratio' standards; an allometric
417 modeling approach. *J. Appl Physiol.* 1995; 79: 1027-31.
- 418 18. Nevill AM, Holder RL, Baxter-Jones A, Round J, and Jones DA. Modeling
419 developmental changes in strength and aerobic power in children. *J Appl Physiol.*
420 1998; 84: 963-70.
- 421 19. Nevill AM, Ramsbottom R, and Williams C. Scaling physiological measurements for
422 individuals of different body size. *Eur J Appl Physiol.* 1992; 65: 110-7.
- 423 20. Nevill AM, Stavropoulos-Kalinoglou A., Metsios G.S., Koutedakis Y., Holder R.L.,
424 Kitas G.D. Mohammed M.A (2011) Inverted BMI rather than BMI is a better proxy
425 for percentage of body fat. *Ann Hum Biol.* 2012; 38: 681-4.
- 426 21. Siri, W. E. (1961). Body composition from fluid space and density: analysis of
427 methods. In *Technique for Measuring Body Composition*, ed. Brozek, J. & Henschel,
428 A. National Academy of Sciences, Washington, DC, USA. 223-244 p.
- 429 22. Taylor, H. L., Jacobs, D. R. J., Schucker, B., Knudsen, J., Leon, A. S. & Debacker, G.
430 A questionnaire for the assessment of leisure time physical activities. *Journal of*
431 *Chronic Diseases.* 1978; 31: 741—755.
- 432 23. Winter E.M. and Nevill A.M. Scaling: Adjusting for differences in body size. In
433 "Kinanthropometry and Exercise Physiology Laboratory Manual for Tests,
434 Procedures and Data, 3rd ed." (eds) R. G. Eston and T. Reilly, Abingdon, Oxon,
435 Routledge; 2009. 300-314 p.

436

437

438 **Legends to Tables**

439

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

443

444

445 **Legends to figures**

446

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

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

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

454

455

456

457