Scaling physiological measurements for individuals of different body size

Alan M. Nevill[1], Roger Ramsbottom[2], and Clyde Williams[3]

[1] School of Sport and Exercise Sciences, University of Birmingham, Birmingham, England
[2] Department of Sports Studies, Roehampton Institute, London, England
[3] Department of Physical Education and Sports Science, University of Technology, Loughborough, England

Summary. This paper examines how selected physiological performance variables, such as maximal oxygen uptake, strength and power, might best be scaled for subject differences in body size. The apparent dilemma between using either ratio standards or a linear adjustment method to scale was investigated by considering how maximal oxygen uptake ($l \cdot min^{-1}$), peak and mean power output (W) might best be adjusted for differences in body mass (kg). A curvilinear power function model was shown to be theoretically, physiologically and empirically superior to the linear models. Based on the fitted power functions, the best method of scaling maximum oxygen uptake, peak and mean power output, required these variables to be divided by body mass, recorded in the units $kg^{2/3}$. Hence, the power function ratio standards ($ml \cdot kg^{-2/3} \cdot min^{-1}$) and ($W \cdot kg^{-2/3}$) were best able to describe a wide range of subjects in terms of their physiological capacity, i.e. their ability to utilise oxygen or record power maximally, independent of body size. The simple ratio standards ($ml \cdot kg^{-1} \cdot min^{-1}$) and ($W \cdot kg^{-1}$) were found to best describe the same subjects according to their performance capacities or ability to run which are highly dependent on body size. The appropriate model to explain the experimental design effects on such ratio standards was shown to be log-normal rather than normal. Simply by taking logarithms of the power function ratio standard, identical solutions for the design effects are obtained using either ANOVA or, by taking the unscaled physiological variable as the dependent variable and the body size variable as the covariate, ANCOVA methods.

Key words: Ratio standards – Physiological capacity – Performance capacity – Experimental design effects – Log-linear models

Correspondence to: A. M. Nevill

Introduction

Measurements in physiology and medicine are often recorded a per-body mass (m) or per-surface area ratio standards, e.g. maximum oxygen uptake ($ml \cdot kg^{-1} \cdot min^{-1}$) or peak power per cross-sectional area of muscle ($W \cdot cm^{-2}$). These simple ratio standards have been used to facilitate the comparison of measurements recorded from individuals of different sizes, since, by dividing by an "appropriate" body size variable, it is assumed that differences in the physiological variable due to the subject's size will have been removed. Hence, the scaled physiological performance variables are assumed to be independent of the subjects' body size dimension.

However, these simple ratio standards have come under strong criticism by authors such as Tanner (1949), Katch (1972, 1973), Katch and Katch (1974) and more recently by Winter et al. (1991). These authors' criticism is based on several observations. Firstly, when simple ratio standards, e.g. maximum oxygen uptake ($ml \cdot kg^{-1} \cdot min^{-1}$), are correlated with a body size dimensions, e.g. m (kg), the correlations are found to be negative, i.e. the simple ratio standards fail to produce a dimensionless physiological performance variable. On the contrary, it would appear that these simple ratio standards "over-scale" by converting the positive correlation, between the physiological performance variable and the body size variable, to a negative one. Secondly, the authors assume that the relationship between the two ratio variables is linear and then argue that if the model that describes this relationship is a true linear proportion, then the least-squares linear regression line should pass close to, if not through, the origin. When this fails to occur in most of the studies cited, the authors propose using an additive adjustment, based on the regression line, to correct the numerator variable for differences in the body-size denominator variable [see Katch and Katch (1974) for details].

In contrast, Kleiber (1950) proposed using a power function model, the simple ratio standard being a special case, to explain the relationship between a physiol-

ogical variable and a body size variable, since such variables are known to be proportional to a power of the subjects' linear dimension (L), e.g. m is proportional to L^3 (Astrand and Rodahl 1986). Hence, theoretically the relationship between such variables is more likely to be a power function model rather than linear model, even if the empirical least-squares regression line may happen to fit the data better (i.e. produce a lower residual sum-of-squares). Additional support for the use of a power function model comes from Schmidt-Nielsen (1984) when scaling physiological variables for animals of different sizes. When metabolic rate was plotted against m for a large variety of animal species, using a log-log scale, a straight line relationship was obtained, a well-known characteristic of a power function model. A similar straight line was obtained by McMahon (1984) when plotting the relationship between log[maximum oxygen uptake $(ml \cdot s^{-1})$] against the $\log[m$ (kg)] for a similar range in animal scale, once again confirming the underlying power function model.

These methods of scaling will be examined by investigating the results of two studies. The first study examines how the maximal oxygen uptake values $(l \cdot min^{-1})$ of 308 subjects might best be scaled for differences in their m (kg). The second study investigates how 16 subjects' peak and mean power output scores (W) might best be scaled for differences in m (kg).

Scaling models and statistical methods

The methods of scaling described by the authors above (Tanner 1949; Katch 1972, 1973; Katch and Katch 1974) were all presented without reference to the errors associated with each model. However, different methods of scaling will make quite different assumptions concerning the relationship that exists between the physiological performance variable Y_i, the relevant body size variable X_i and their associated residual errors.

(i) The simple ratio standard (Y_i/X_i) implies the following mathematical relationship between the physiological performance and the body size variables,

$$Y_i = a \cdot X_i + e_i, \tag{1}$$

where e_i is an additive error term. The model assumptions, given in Eq. 1, assume the groups' ratio standard parameter a is calculated using least-squares linear regression but by omitting the intercept constant.

(ii) The least-squares linear adjustment method of scaling assumes the following model:

$$Y_i = a + b \cdot X_i + e_i \tag{2}$$

where e_i denotes the residual errors from the least-squares regression line of best fit. Similar to the simple ratio standard model, least-squares linear regression assumes a constant error term throughout the range of the variables Y_i and X_i, an assumption that would appear to be questionable when modelling human performance variables recorded from subjects of different body size. Assuming a linear model with an additive error term, the residual errors are obtained by re-arranging Eq. 1:

$$e_i = Y_i - a - b X_i.$$

The method of adjustment, outlined by Katch and Katch (1974), requires this residual error to be added to the group's arithmetic mean, i.e. $\bar{Y} + e_i$, to represent the individual's "adjusted" or scaled score.

(iii) The power function ratio standard (Y_i/X_i^b) assumes the following model between the physiological performance variable (Y_i) and the body size variable (X_i)

$$Y_i = a \cdot X_i^b \cdot e_i \tag{3}$$

where e_i is a multiplicative error ratio term. This multiplicative error term is a particularly attractive feature of this model since it will accommodate a spread in the subjects' Y_i and X_i values when plotted against each other, provided these scores diverge at a constant proportion to each other. The parameters a and b are obtained assuming a log-linear model, i.e. by taking logs of Eq. 3 and fitting the least-squares regression line to $\log(Y_i)$ using $\log(X_i)$ as the predictor variable,

$$\log(Y_i) = \log(a) + b \cdot \log(X_i) + \log(e_i) \tag{4}$$

The residual error ratios are obtained by re-arranging Eq. 3 to obtain

$$e_i = Y_i/(a \cdot X_i^b) = (Y_i/X_i^b)/a. \tag{5}$$

Hence, the power function of best fit would suggest the appropriate method of scaling should be the power function ratio standard Y_i/X_i^b, a familiar equation to supporters of the simple ratio standard when the power function parameter equals unity (i.e. $b = 1$).

The statistical modelling of the data from studies 1 and 2, given below, was carried out on the University of Birmingham's IBM 3090 mainframe computer, using the statistical package GLIM [Generalized Linear Interactive Modelling; Baker and Nelder (1978)]. Much of the descriptive analyses were confirmed on an IBM PS/2 using the statistical package MINITAB (Minitab Reference Manual 1989).

Study 1: scaling maximum oxygen uptake for differences in m

Subjects and methods. Over a 7-year period 308 recreationally active subjects (men, $n = 179$; women, $n = 129$) gave their informed consent and volunteered to take part in experiments at Loughborough University which required the direct determination of maximum oxygen uptake. All the experiments were conducted in keeping with the principles embodied in the Declaration of Helsinki for experiments involving human subjects. Of the 308 subjects, 204 subjects (men, $n = 112$; women, $n = 92$), completed both the maximum oxygen uptake test and a 5-km performance test on the running track. The subject's maximum oxygen uptake was determined during inclined treadmill running using a modification of procedures originally described by Taylor et al. (1955). The speed of the treadmill was constant throughout the test, but the elevation was increased by 2.5% every 3 min. Voluntary exhaustion was usually reached within 8–12 min. Samples of expired air were collected in 150-l Douglas bags during the last minute of each 3-min period and during the last minute of the test. Oxygen uptake was determined as previously described by Williams and Nute (1983).

Table 1. The residual sum-of-squares from the three scaling models relating maximum oxygen uptake $(l \cdot min^{-1})$ to body mass (kg)

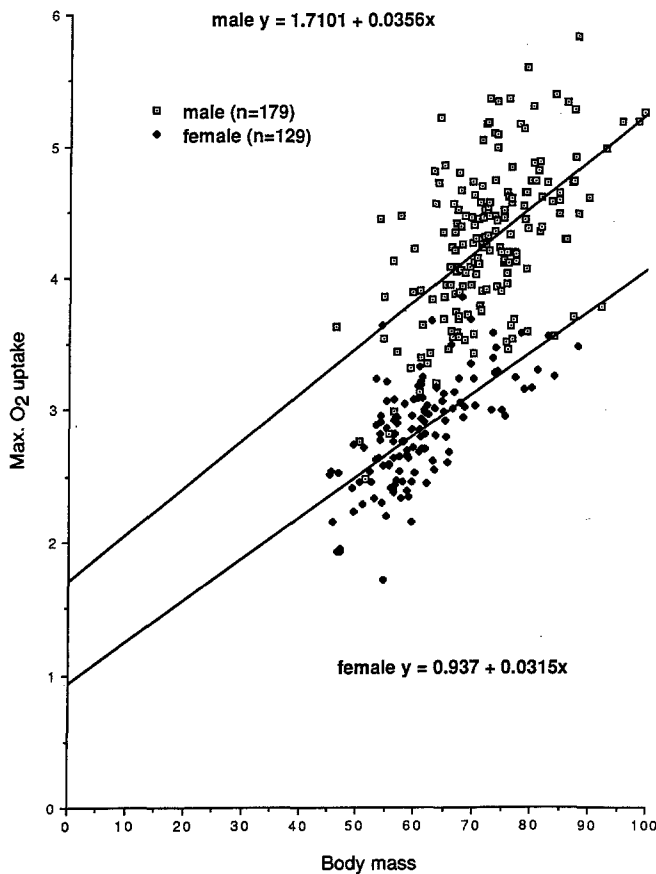|	Residual sum-of-squares (SEM)	
Scaling models	Men $(n = 179)$	Women $(n = 129)$
Simple ratio standard	49.6 (0.53)	13.6 (0.330)
Linear regression	41.8 (0.49)	11.5 (0.301)
Power function	41.6 (0.48)	11.4 (0.298)

Fig. 1. The least-squares linear relationships between maximum oxygen uptake and body mass for both men ($\square$; $n = 179$; $y = 1.7101 + 0.0356x$) and women ($\blacklozenge$; $n = 129$; $y = 0.937 + 0.0315x$)

Fig. 2. The power functional relationships between maximum oxygen uptake and body mass for both men ($\square$; $n = 179$; $y = 0.2876 \cdot x_{0.6297}$) and women ($\blacklozenge$; $n = 129$; $y = 0.1498 \cdot x_{0.7163}$)

Modelling the relationship between maximum oxygen uptake and m. The results of modelling maximum oxygen uptake ($l \cdot min^{-1}$) and m (kg) for men and women using all three models are summarised in Table 1.

As anticipated, the correlation between the simple ratio standard, maximum oxygen uptake ($ml \cdot kg^{-1} \cdot min^{-1}$), and m (kg) for men and women was negative $r = -0.385$ ($P < 0.01$; $n = 179$) and $r = -0.347$ ($P < 0.01$; $n = 129$) respectively. In contrast, when maximum oxygen uptake ($l \cdot min^{-1}$) was scaled according to the linear adjustment methods outlined in Katch and Katch (1974) and correlated with m (kg), the correlation was found to be zero ($r = 0.00$) for both men and women. When the least-squares regression lines were fitted to the data for the men and women separately (Fig. 1) both the intercept parameters were found to be significantly greater than zero. This result agrees with the findings of Ribisl and Kachadorian (1969) based on 24 men, presented in Katch (1973).

However, the best fit was found when these data were modelled using power function, assuming a log-linear model to estimate the parameters. The resulting power functions for men and women are given in Fig. 2.

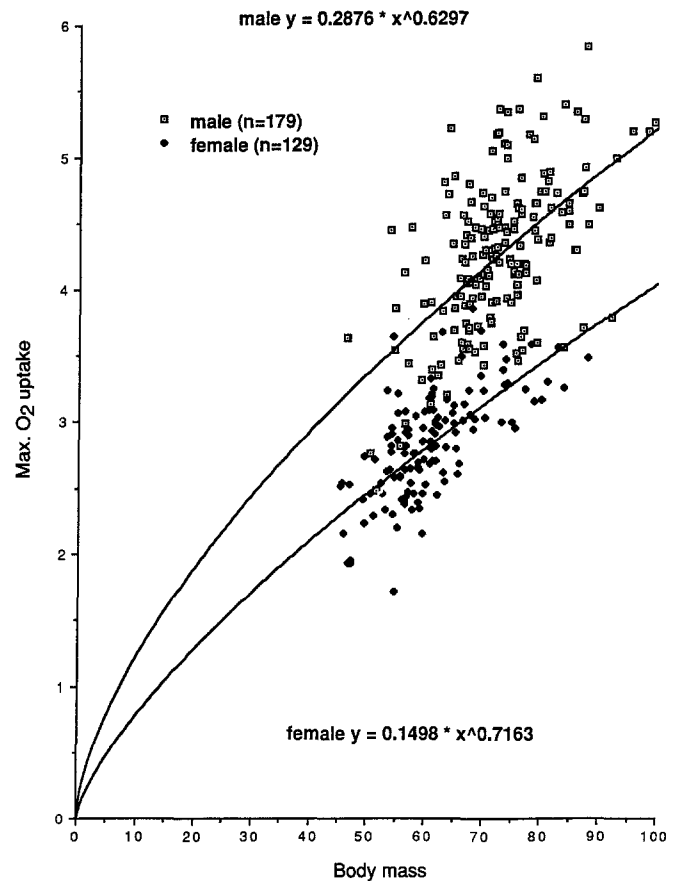No statistically significant differences were found between the power function parameters (0.63 and 0.72),

but a significant difference was found between the constant multipliers (0.29 and 0.15). Hence, when these data were analysed together, allowing separate constant multipliers but a common power function parameter, the following power functions were obtained:

$$Y(\text{men}) = 0.244 \cdot X^{0.669} \qquad (6)$$

and

$$Y(\text{women}) = 0.183 \cdot X^{0.669} \qquad (7)$$

It is interesting to note that the combined power function parameter b is almost exactly 2/3. Assuming that maximum oxygen uptake ($l \cdot min^{-1}$) should be scaled using the model that gives the best fit (least residual sum-of-squares), the most appropriate method of scaling is the power function ratio standard, that would require the subjects' maximum oxygen uptake ($l \cdot min^{-1}$) to be divided by a proportion of the subjects' m, recorded in the units $kg^{2/3}$, i.e. ($l \cdot kg^{2/3} \cdot min^{-1}$) or ($ml \cdot kg^{-2/3} \cdot min^{-1}$). Note that the correlation between maximum oxygen uptake ($ml \cdot kg^{-2/3} \cdot min^{-1}$) and m (kg) for men and women was not statistically significant $r = 0.034$ ($P > 0.05$; $n = 179$) and $r = 0.017$ ($P > 0.05$; $n = 129$) respectively.

"Scaling" maximum oxygen uptake to predict 5-km running performance. Nevill et al. (1990) showed that 5-km

Table 2. The residual sum-of-squares from predicting 5-km running performance

| Methods of scaling | Residual sum-of-squares (SEM) | |
	Men ($n = 112$)	Women ($n = 192$)
Simple ratio standard (ml·kg^{-1}·min^{-1})	7.44 (0.260)	5.47 (0.247)
Linear adjustment method of scaling	12.83 (0.342)	7.59 (0.291)
Power function (ml·kg$^{-2/3}$·min^{-1})	12.04 (0.331)	7.51 (0.289)

running performance was better related (correlated) to maximum oxygen uptake if running performance was recorded as an average speed (m·s^{-1}) rather than an absolute run time (min). Hence, when the average 5-km run speeds were predicted using the scaled power function ratio, maximum oxygen uptake (ml·kg$^{-2/3}$·min^{-1}), the residual sum-of-squares was found to be a little less than the linear adjustment method of scaling for both men and women (Table 2).

However, using this ratio standard limits the power function parameter to $b = 2/3$ when trying to explain average 5-km running speed. On the other hand, if we assume that the relationship between running performance (5-km run speed m·s^{-1}), maximum oxygen uptake (l·min^{-1}) and m (kg) is better described by a power function, but not subject to the restriction imposed by the ratio (ml·kg$^{-2/3}$·min^{-1}), multiple regression can be used to optimally combine maximum oxygen uptake (l·min^{-1}) and m (kg) to predict running performance (5-km run speed m·s^{-1}). Assuming a log-linear model, no statistically significant differences were found between the male and female models for either the maximum oxygen uptake or m parameters. Hence, the common power function model relating 5-km running speed, Z(m·s^{-1}), to maximum oxygen Y(l·min^{-1}) and m X(kg) is given by

$$Z(\text{m·s}^{-1}) = 84.3 \cdot Y^{1.01} \cdot X^{-1.03} . \tag{8}$$

The power function model, given by Eq. 8, is both simple and meaningful. The best predictor of 5-km run times, when recorded as a rate of performance, i.e. mean running speed (m·s^{-1}), is almost exactly proportional to the ratio standard maximum oxygen uptake (l·min^{-1}) divided by m (kg) or ml·kg^{-1}·min^{-1}. Hence, when the average 5-km run speeds were predicted for men using maximum oxygen uptake (ml·kg^{-1}·min^{-1}), the residual sum-of-squares was found to be considerably reduced for both men and women (Table 2). The relationship between average 5-km run speed and maximum oxygen uptake (ml·kg^{-1}·min^{-1}) is plotted in Fig. 3.
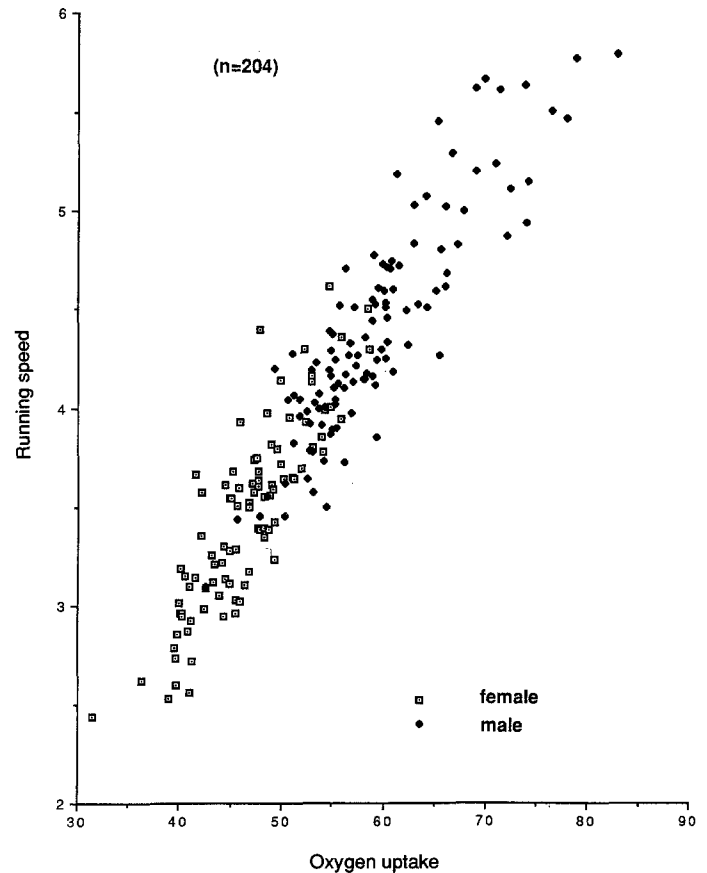
Fig. 3. Running speed vs maximum oxygen uptake in men (◆) and women (□); $n = 204$

Study 2: scaling peak and mean power output for differences in m

Subjects and methods. The peak and mean power output (W) of 16 recreationally active male subjects were recorded on a non-motorised treadmill as part of a training study carried out at Loughborough University, the results of which are described in Holmyard et al. (1991). In addition to the laboratory-based treadmill tests, the subjects also undertook a 50-m and a 30-s sprint test on an outdoor running track under carefully standardised test procedures and conditions. The subjects' mean (SD) age, height, and weight were 22.6 (2.7) years, 1.74 (0.1) m, and 68.0 (11.0) kg respectively. All the data presented below were collected prior to the sprint training programme.
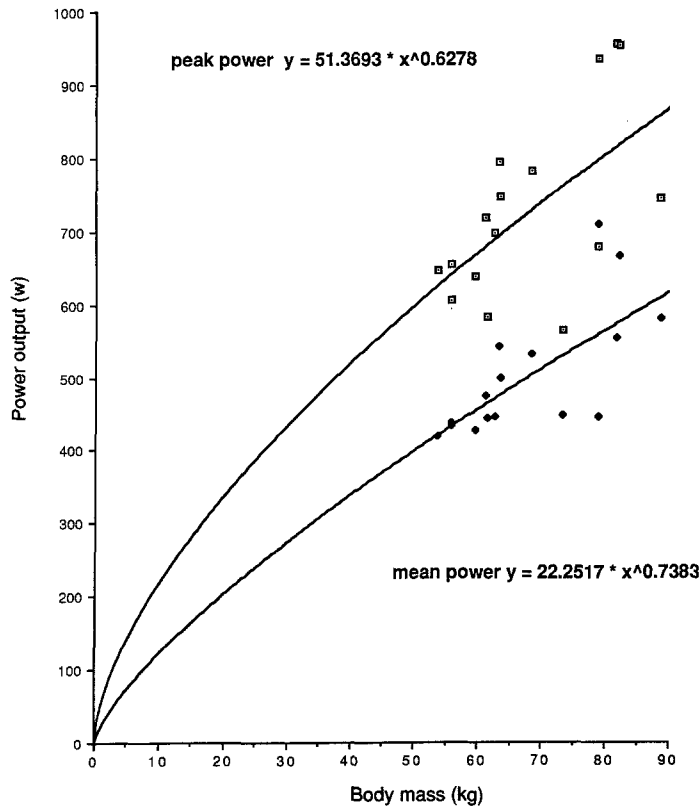
Modelling the relationship between power output and m. As anticipated, a moderately strong positive relationship was found between peak power (W) and m (kg) $r = 0.608$ and between mean power output (W) and m (kg) $r = 0.703$. The results of modelling peak and mean power output (W) against m (kg) using all three models are summarised in Table 3.

Once again, the curvilinear power function models resulted in the best fit to both the peak and mean power output data (Fig. 4), having estimated the exponents to be 0.628 and 0.738 respectively using the log-linear regression model.

Further inspection of Fig. 4 reveals the scores diverge with an increase in scale, questioning the constant error

Table 3. The residual sum-of-squares from the three scaling models relating peak and mean power output (W) to body mass (kg)

Scaling models	Residual sum-of-squares (SEM)	
	Peak power ($n = 16$)	Mean power ($n = 16$)
Simple ratio standard	175 434 (108.1)	64 137 (65.39)
Linear regression	150 386 (103.6)	58 527 (64.66)
Power function	150 086 (100.0)	58 481 (62.44)

Table 4. The residual sum-of-squares from predicting 50-m and 30-s sprint running performance

Methods of scaling	Residual sum-of-squares (SEM)	
	50 m (m·s^{-1})	30 s (m·s^{-1})
Simple ratio standard (ml·kg^{-1}·min^{-1})	0.550 (0.198)	1.181 (0.290)
Linear adjustment method of scaling	0.562 (0.200)	1.202 (0.293)
Power function (ml·kg$^{-2/3}$·min^{-1})	0.574 (0.203)	1.259 (0.300)

Fig. 4. The power functional relationships between peak power and body mass and between mean power and body mass. □, Men; ◆, Women. Peak power $y = 51.3693^* \ x_{0.6278}$; mean power $y = 22.2517^* \ x_{0.7383}$

"Scaling" peak and mean power output to predict sprint running performance. In order to relate the peak and mean power performances (W), recorded on the non-motorised treadmill, to running performance, the subjects were asked to complete a 50-m and a 30-s sprint test. When peak and mean power output was scaled as the ratio standard W·kg$^{-2/3}$ and used to predict the 50-m and 30-s sprint (m·s^{-1}) respectively, the quality of fit was found to be marginally worse than the fit obtained using the linear adjustment method of scaling peak and mean power (Table 4). However, as mentioned earlier, using this ratio standard limits the power function parameter to $b = 2/3$. If we assume that the relationship between sprinting performance, power output (W) and m (kg) is better described by a power function model, but not subject to the restriction imposed by the ratio W·kg$^{-2/3}$, multiple regression can be used to optimally combine power output (W) and m (kg) to predict sprinting performance. Assuming a log-linear model to estimate the parameters, the optimal predictors of 50-m and 30-s sprint performance were found to be almost exactly ratio standards (W·kg^{-1})$^{0.14}$ and (W·kg^{-1})$^{0.3}$ respectively. However, if the simple ratio standards, peak and mean power (W·kg^{-1}), was used to predict 50-m and 30-s sprint performance respectively, the residual sum-of-squares increased by only a small amount in both cases, suggesting an equally satisfactory and optimal fit (Table 4).

assumption required by the linear scaling model (ii) and supporting the multiplicative error term in the power function model (iii). No statistically significant differences were found between the power function parameters (0.628 and 0.738) but a significant difference was found between the constant multipliers (51.8 and 22.4). Hence, when these data were analysed together, allowing separate constant multipliers but a common power function parameter, the following power functions were obtained:

Y(peak power) = $41.6 \cdot X^{0.68}$
and
Y(mean power) = $28.7 \cdot X^{0.68}$.

As before, the combined power function parameter 'b' is almost exactly 2/3.

Investigating experimental design effects

The researcher will often with to investigate how these ratio variables respond to various experimental design conditions. Consider the classical "split-plot" experimental design frequently used to investigate the effectiveness of a training study, e.g. the design used by Holmyard et al. (1991). Typically, the researcher would divide the group of volunteers at random into two groups, an experimental group and a control group. Performance measurements would be taken for all subjects before and after a period of time in which only the experimental group would have completed a training programme, i.e. the control group would just continue with their normal lifestyle. The linear model used to describe such a design, assuming additive effects and an additive error term, is given by the following equation:

$$Y_{ijk}/X_{ijk} = \mu + \alpha_i + \pi_{k(i)} + \beta_j + \alpha\beta_{ij} + \beta\pi_{jk(i)} + \varepsilon_{m(ijk)} \quad (11)$$

where Y_{ijk}/X_{ijk} is the ratio standard, α_i denotes the group effect (e.g. control vs experimental), β_j represents the within-subject or trial effect (e.g. before vs after training), $\pi_{k(i)}$ indicates the effect of the kth subject in group α_i and the residual error $\varepsilon_{n(ijk)}$ is assumed to be a constant additive terrm throughout the range of measurements (see Winer 1971, pp 519 for full details of this design). Some authors, for example, Winter et al. (1991) recommend that the body size variable X_{ijk} should be included as a covariate, rather than used as the denominator in the ratio standard, when analysing the results of their experiments, i.e. in the model described above they would recommend that the physiological variable Y_{ijk} should be analysed assuming the following model:

$$Y_{ijk} = b(X_{ijk} - \overline{X_i}) + \mu + \alpha_i + \pi_{k(i)} + \beta_j + \alpha\beta_{ij} + \\ + \beta\pi_{k(i)} + \varepsilon_{m(ijk)} \quad (12)$$

where X_{ijk} will automatically adjust the physiological variable Y_{ijk} assuming a linear relationship (see Winer 1971, p 781 for full details of this design). Clearly the models given by Eqs. 11 and 12 are incompatible and are likely to yield contradictory results when investigating experimental design effects, i.e. multiplying both sides of Eq. 11 by X_{ijk} will result in a over-complicated model involving interaction terms between the covariate and the experimental design parameters.

As discussed earlier, the assumption that physiological and body size variables are linearly related with a constant additive error term is unreasonable. By observing Figs. 1, 2 or 4, the data have a tendency to diverge as the scale of both maximum oxygen uptake ($l \cdot min^{-1}$) and power (W) increase with m (kg), implying a multiplicative error term rather than a constant additive error term assumed by the models in Eqs. 11 and 12. In addition, any changes in these ratio standards, due to the experimental design, are also more likely to be described by a multiplicative model. In the training study, for example, the researcher is often more interested in comparing the percentage change in the ratio variable, from before to after training, between the control and experimental groups, rather than the absolute change in the ratio variable. When experimenter is interested in percentage changes rather than absolute changes in the dependent variable, a multiplicative model should be used to explain such effects.

A multiplicative model, used to describe a split-plot experimental design when a ratio standard is to be used as the dependent variable, is given by the following equation:

$$Y_{ijk}/X_{ijk} = \mu \cdot \alpha_i \cdot \pi_{k(i)} \cdot \beta_j \cdot \alpha\beta_{ij} \cdot \beta\pi_{jk(i)} \cdot \varepsilon_{m(ijk)} \quad (13)$$

where, as before, Y_{ijk}/X_{ijk} is the ratio standard, α_i denotes the group ratio effect, β_j represents the trial ratio effect, $\pi_{k(i)}$ indicates the ratio effect of the kth subject in group α_i and the residual error ratio $\varepsilon_{m(ijk)}$ is assumed to be a constant multiplicative error ratio. Provided the ratio standard in the left hand side of Eq. 13 has been defined by the power function of best fit, i.e. Y_{ijk}/X_{ijk}^b, by simply taking logarithms of this ratio and using an

appropriate re-parameterization, the experimental design effects can be investigated using traditional ANOVA methods. This does, of course, assume that the error ratio term in Eq. 13 has a log-normal distribution. However, since many of the physiological and body size variables have been shown to be proportional to a power (>1) of L, e.g. $m\alpha L^3$, the distribution of such variables are likely to be positively skewed, a characteristic that will be corrected by a logarithmic transformation. The same experimental design effects can be obtained by analysing $\log(Y_{ijk})$ as the dependent variable, using the analysis of covariance (ANCOVA), where $\log(X_{ijk})$ will automatically adjust the numerator dependent variable to provide the same design effects as the ANOVA solution described above.

As mentioned in our examples, if the researcher wishes to use the ratio standard that best reflects running performance, he or she should use m related variables, e.g. maximum oxygen uptake ($ml \cdot kg^{-1} \cdot min^{-1}$), peak or mean power ($W \cdot kg^{-1}$) rather than the ratio standard defined by the power function of best fit ($ml \cdot kg^{-2/3} \cdot min^{-1}$) or ($W \cdot kg^{-2/3}$). Under these circumstances the appropriate design effects can only be found by analysing the logarithms of the ratio standard, m related maximum oxygen uptake ($ml \cdot kg^{-1} \cdot min^{-1}$), peak or mean power output ($W \cdot kg^{-1}$), as the dependent variable using traditional ANOVA methods.

Discussion

An important necessary although not sufficient requirement of a scaling technique is, that having applied the scaling method to the physiological performance variable, the scaled variable should be independent of body size. This is clearly not the case for the simple ratio scales maximum oxygen uptake ($ml \cdot kg^{-1} \cdot min^{-1}$), peak and mean power ($W \cdot kg^{-1}$), since, in the examples presented in either studies 1 or 2, the simple ratio scales correlated negatively with m, although, due to the reduced sample size in study 2, we can only report the anticipated significant negative correlation for both men and women in study 1. In contrast, both the linear adjustment and power function methods of scaling would appear to have successfully removed the body size dimension, as suggested by the insignificant correlations found between the scaled variables and m for all the examples given.

However, if we assume that physiological performance variables, such as maximum oxygen uptake ($l \cdot min^{-1}$) peak and mean power output (W), should be scaled for individuals of different body size using the most appropriate model (based on the residual sum-of-squares and the more appropriate multiplicative error term), rather than using the linear adjustment methods, the best method of scaling is given by the power function ratio standards, maximum oxygen uptake ($ml \cdot kg^{-2/3} \cdot min^{-1}$) and peak and mean power output ($W \cdot kg^{-2/3}$). In all four examples described above, the residual sum-of-squares was less for the power function models when compared with the simple ratio scale mod-

els or the linear adjustment models. Hence, the scaling models that best divided a population in terms of their physiological capacity, i.e. ability to utilise oxygen or provide power output maximally for a wide range of subjects, independent of body size, was shown to be the power function ratio standards (ml·kg$^{-2/3}$·min^{-1}) and (W·kg$^{-2/3}$) respectively.

The same conclusion was reached by Astrand and Rodahl (1986), when modelling maximum oxygen uptake (l·min^{-1}) of top Norwegian athletes trained in a variety of events (data by Vaage and Hermansen) and by Secher et al. (1983) and Secher (1990) when describing the maximum oxygen uptake of oarsmen. Indeed, the findings that peak and mean power should be scaled by dividing the power output (W) by m (kg$^{-2/3}$) is entirely plausible, once again, based on the work of Astrand and Rodahl (1986). If, as the authors show, human power output is proportional to L^2 then the appropriate scaling parameter is $m^{2/3}$, where L represents a linear function of body size. Additional support for this result comes from Nevill et al. (1991), when scaling leg peak power (W) of 34 men and 47 women using lean leg volume (l). Not surprisingly, the power function ratio standard proved to be the best method of scaling peak power output (W) with the body size denominator variable, l, raised to the power 0.63, (approximately 2/3). This finding suggests that the appropriate method of scaling peak power is to divide the power output (W) by $l^{2/3}$, i.e. a surface area dimension such as the cross-sectional area of the leg muscle. This result might have been anticipated by many muscle physiologists.

An implicit assumption behind scaling physiological performance measurements such as maximum oxygen uptake (l·min^{-1}) for m (kg), is that having adjusted the variable of interest, any remaining "residual" difference in the scaled variable should help to explain differences in other independent performance variables, e.g. 5-km run times, without being confounded by differences in the subjects' body size. When the power function ratio standards, maximum oxygen uptake (ml·kg$^{-2/3}$·min^{-1}) and peak and mean power output (W·kg$^{-2/3}$), were used to predict running performance, the quality of fit was good but not optimal. In both studies investigated, the simple ratio standard, maximum oxygen uptake (ml·kg^{-1}·min^{-1}), produced the optimal predictor of average 5-km running speed and the simple ratio standards, peak and mean power output (W·kg^{-1}), were found to be optimal at predicting 50-m and 30-s track running speeds respectively.

However, it is essential that the distinction between, and the implications associated with, these two "scaling" methods are clearly understood. The simple ratio standards, maximum oxygen uptake (ml·kg^{-1}·min^{-1}), peak and mean power (W·kg^{-1}), are not true scaling methods. On the contrary, they result in a "scaled" index that over-adjusts the subjects' physiological performance variables for m but nevertheless produces a ratio or index that best relates to running performance. Hence, if the researcher wishes to obtain a index that reflects the performance capacity of a sample, he or she should use the ratio derived from the multiple regression fit, assum-

ing a log-linear model, using a valid measure of performance as the dependent variable and the physiological numerator variable and the body size denominator variable as the predictor variables. The ratio standard (ml·kg^{-1}·min^{-1}) was the model that was able to divide the subjects according to their performance capacity or ability to run 5 km. The same was true for "scaling" peak and mean power output to reflect sprinting performance. The index that was best able to reflect the subjects' sprinting performance on the track was found to be peak and mean power output (W·kg^{-1}).

Clearly the performance capacity of participants in different sporting events is likely to vary according to the nature of the sport. Sporting events that require the performer to carry his or her own body weight, such as running, the likely body size denominator variable will be the entire m (kg). Alternatively, if the sporting event is weight supported, such as cycling, rowing or canoeing, the body size denominator component of a "scaled" index is like likely to be considerably less if not absent, e.g. unscaled maximum oxygen uptake (l·min^{-1}).

However, if the researcher wishes to investigate the results of experiments on physiological variables that are known to vary with body size, such as maximum oxygen uptake (l·min^{-1}), peak and mean power output (W), but with the body-size dimension removed, the appropriate design effects will be obtained by analysing the logarithms of the power function ratio standard, i.e. (ml·kg$^{-2/3}$·min^{-1}) and (W·kg$^{-2/3}$), as the dependent variable using traditional ANOVA methods. Alternatively, the identical conclusions will be obtained simply by setting the logarithms of the physiological numerator variable as the dependent variable and using the logarithms of body size denominator variable as the covariate in the analysis of covariance (ANCOVA). This will automatically adjust or scale the physiological variable for differences in the subjects' body size to provide the same design effects as the ANOVA solution.

References

Astrand P-O, Rodahl K (1986) Textbook of work physiology, 3rd edn. McGraw-Hill, New York

Baker RJ, Nelder JA (1978) Generalized linear interactive modelling, release 3. Numerical Algorithms Group, Oxford

Holmyard DJ, Chia YHM, Lakomy HKA, Nevill ME (1991) Effects of training on treadmill sprinting and recovery. J Sports Sci 9:417–418

Katch VL (1972) Correlation v ratio adjustment of body weight in exercise-oxygen studies. Ergonomics 15:671–680

Katch VL (1973) Use of the oxygen/body weight ratio in correlational analyses: spurious correlations and statistical considerations. Med Sci Sports Exerc 5:253–257

Katch VL, Katch FI (1974) Use of weight-adjusted oxygen uptake scores that avoid spurious correlations. Res Q 4:447–451

Kleiber M (1950) Physiological meaning of regression equations. J Appl Physiol 2:417–423

McMahon TA (1984) Muscles, reflexes and locomotion. Princeton University Press, New Jersey

Minitab Reference Manual (1989) Minitab Inc. 3081, Enterprise Drive, State Collage, PA 16801, USA

Nevill AM, Ramsbottom R, Williams C (1990) The relationship between athletic performance and maximal oxygen uptake. J Sports Sci 8:290–292

Nevill AM, Ramsbottom R, Williams C, Winter EM (1991) Scaling physiological performance measurements for individuals of different body size. J Sports Sci 9:427–428

Ribisl PM, Kachadorian WA (1969) Maximal oxygen intake prediction in young and middle aged males. J Sports Med Phys Fitness 9:17–22

Schmidt-Nielsen K (1984) Scaling: why is animal size so important? Cambridge University Press, Cambridge

Secher N (1990) Rowing. In: Reilly T, Sencher N, Snell P, Williams CE, Spoon FN (eds) Physiology of sports. Chapman and Hall, London, pp 259–286

Secher N, Vaage O, Jensen K, Jackson RC (1983) Maximal aerobic power in oarsmen. Eur J Appl Physiol 51:155–162

Tanner JM (1949) Fallacy of per-weight and per-surface area standards and their relation to spurious correlations. J Appl Physiol 2:1–15

Taylor HL, Buskirk E, Henschel A (1955) Maximum oxygen intake as an objective measure of cardiorespiratory performance. J Appl Physiol 8:73–80

Williams C, Nute MLG (1983) Some physiological demands of a half-marathon race on recreational runners. Br J Sports Med 17:152–161

Winer BJ (1971) Statistical principles in experimental design. McGraw-Hill, New York

Winter EM, Brookes FBC, Hamley EJ (1991) Maximal exercise performance and lean leg volume in men and women. J Sports Sci 9:3–13