

Interpreting Aerobic Fitness in Youth: The Fallacy of Ratio Scaling

Jo Welsman and Neil Armstrong
University of Exeter

In this paper, we draw on cross-sectional, treadmill-determined, peak oxygen uptake data, collected in our laboratory over a 20-year period, to examine whether traditional per body mass (ratio) scaling appropriately controls for body size differences in youth. From an examination of the work of pioneering scientists and the earliest studies of peak oxygen uptake, we show how ratio scaling appears to have no sound scientific or statistical rationale. Using simple methods based on correlation and regression, we demonstrate that the statistical relationships, which are assumed in ratio scaling, are not met in groups of similar aged young people. We also demonstrate how sample size and composition can influence relationships between body mass and peak oxygen uptake and show that mass exponents derived from log-linear regression effectively remove the effect of body mass. Indiscriminate use of ratio scaling to interpret young people's fitness, to raise "Clinical Red Flags", and to assess clinical populations concerns us greatly, as recommendations and conclusions based upon this method are likely to be spurious. We urge those involved with investigating youth fitness to reconsider how data are routinely scaled for body size.

Keywords: peak oxygen uptake, allometry, adolescent, exercise performance, children

Peak oxygen uptake (peak $\dot{V}O_2$) is internationally recognized as the criterion measure of youth aerobic fitness, and laboratory-determined peak $\dot{V}O_2$ data from children and adolescents have been available for over 80 years. However, throughout this period, the development of peak $\dot{V}O_2$ with growth and maturation has been persistently misinterpreted through the uncritical use of ratio scaling of data ($\text{mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$). More appropriate means of accommodating body size were well-established in the general biological sciences (eg, 19) before the initial studies of children's aerobic fitness but were either not considered, or noted and ignored, by the pioneers of pediatric exercise physiology. This set a trend which has continued to the present day and which has not only clouded our current understanding of the growth and development of youth aerobic fitness (and other size-related physiological variables) but also misrepresented the role of aerobic fitness in subsequent analyses. For example, the generation of fitness norms for young people's health and well-being, and the assessment of fitness in young people with serious long-term health conditions.

Robinson's (28) laboratory-based study of physical fitness in 6- to 91-year-old males, in 1938, was the first to include boys. In accord with earlier studies of men, $\dot{V}O_2$ data were initially reported as $\text{L}\cdot\text{min}^{-1}$ before, without a rationale, being "referred to body weight" (p. 280) and presented as a ratio in $\text{mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$. Robinson compared the ratio-scaled data of the men in his study with his own unpublished data and with values he calculated from individual values of $\dot{V}O_2\text{max}$ (in $\text{L}\cdot\text{min}^{-1}$) and body mass (in kg) published in classical papers which had themselves not ratio-scaled their data (16,17,29). In the second laboratory-based investigation of boys' peak $\dot{V}O_2$, data were not even presented in $\text{L}\cdot\text{min}^{-1}$ but, without an underpinning rationale or statistical justification, only reported in ratio with body mass (25).

In 1952, the first study to include girls, Åstrand (10) presented and discussed his data in both absolute ($\text{L}\cdot\text{min}^{-1}$) and ratio-scaled ($\text{mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$) terms but, insightfully, expressed reservations about whether this approach was appropriate with children. Specifically, he noted that group comparisons of peak $\dot{V}O_2$ should be scaled for active muscle mass rather than total body mass but, given the complications of measuring this, at least scaled to lean body mass. Subsequently, in his inspirational 1970 "Textbook of Work Physiology," he devoted a chapter to interpreting "Body dimensions and muscular work" and addressed the interpretation of "maximal aerobic power in children" (11). Interestingly between these 2 publications, and before any further significant studies of young people, von Döbeln (37) had postulated, and subsequently demonstrated, that maximal aerobic power in young men and women did not scale in direct ratio with body mass and should be expressed relative to (body mass – fat mass)^{0.67}.

The influential Saskatchewan Growth and Development study of the 1970s–1980s was the first longitudinal study to evaluate physiological parameters in both boys and girls over an extended period (24). By the time of the study, Tanner's (36) seminal paper, subsequently reexplored with specific reference to oxygen uptake (20) had unequivocally established that expressing physiological functions such as $\dot{V}O_2$ in relation to body mass is, "theoretically fallacious and in practice (except for a special case) misleading" (36, p. 14). The Saskatchewan team noted, "the most common way of expressing $\dot{V}O_2\text{max}$ has been relative to body weight. The use of such an index is open to criticism because no account is taken of the spurious correlation between indices (Tanner, 1949) and considerable difficulties in interpretation can arise" (24, p. 19–20) and reported data in both $\text{L}\cdot\text{min}^{-1}$ and $\text{mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$.

In his landmark book, Bar-Or (12) stated that, "Although theoretically not the method of choice, the most common way of expressing maximal O_2 uptake for comparative purposes has been per kilogram body weight" (p. 4) but then proceeded to discuss ratio-scaled data in relation to age. Over the next 20 years, regular critical reviews and empirical reports (eg, 6,42,43,49)

The authors are with Children's Health and Exercise Research Centre, University of Exeter, Exeter, United Kingdom. Armstrong (N.Armstrong@exeter.ac.uk) is corresponding author.

clearly demonstrated the inadequacy of ratio scaling for interpreting physiological data in relation to body size, and new textbooks of pediatric exercise physiology devoted whole chapters to the issue (eg, 5,31,39,48). In the second edition of Bar-Or's text, coauthored with Rowland and now including an appendix reviewing "Scaling for size differences," it was reiterated that, "ratio scaling is not the ideal way to compare the maximal aerobic power of people who differ in body size" (13, p. 6) but noted that as, "the most convenient and traditionally accepted way . . . a majority of studies still express maximal O₂ uptake per kilogram body mass" (p. 7). In the 15 years since this publication, incredibly, nothing appears to have changed. Second editions and new textbooks (eg, 30,38,40,41) have updated tutorial chapters which clearly demonstrate the fallacy of ratio scaling and advocate the use of alternative methodology. The vast majority of authors, however, persist in failing to present a scientific evidence-based rationale and/or statistical justification for the use of ratio scaling but "conveniently" report peak $\dot{V}O_2$ in mL·kg⁻¹·min⁻¹ as the primary (or often the only) descriptor of youth aerobic fitness regardless of context. We have even been advised ourselves by both reviewers and editors of reputable journals to erroneously ratio scale our data and present it in this manner, "if it is to be considered further for publication"!

With reference to interpreting growth-related data, we agree with Nobel laureate Andre Gide who allegedly began many of his lectures with, "Everything has already been said, but since nobody was listening, we have to start again." The objective of this paper is therefore to question the "convenient and traditional" use of mL·kg⁻¹·min⁻¹ to describe youth peak $\dot{V}O_2$ without a scientific evidence-based rationale and statistical justification. It is not the aim of this paper to provide a detailed statistical tutorial for scaling methods, as these are available elsewhere (38,40,41) or to extensively review the literature. In this paper, we draw on our own data sets collected over a 20-year period to examine, using simple statistical techniques, in how many of these a simple ratio standard is an appropriate and statistically justified method for accounting for size differences. We also consider whether theoretically derived alternative standards might be applied as a routine alternative and examine some of the factors which affect the interpretation of size-related physiological function.

Sources of Data for This Paper

Over the 20-year period between the first laboratory measurements of children's peak $\dot{V}O_2$ within the Children's Health and Exercise Research Centre and 2006, over 1700 treadmill (TM) peak $\dot{V}O_2$ determinations were made with 9- to 18-year-old males and females in a combination of individual cross-sectional and longitudinal studies. The vast majority of these individuals were volunteers recruited from schools and colleges local to the Children's Health and Exercise Research Centre and comprise a normal, healthy population with a consequent range of body size, body fatness, biological maturity, and levels of physical activity.

In this paper, we draw on all of our cross-sectional data from this period totaling 958 TM peak $\dot{V}O_2$ determinations from 13 different studies which themselves comprise 20 individual male or female groups. We have not combined children of the same age from different studies as, for the purposes of this paper, we want to examine how factors such as different sample sizes and sample composition may affect the interpretation of size-adjusted peak $\dot{V}O_2$ in typical studies of young people. In many of the original studies, investigating sex differences was a primary concern. We are not concerned with sex differences here and so

provide separate illustrative analyses for males and females. Table 1 summarizes the original references and relevant descriptive statistics. Participant numbers, and hence mean values for peak $\dot{V}O_2$, may vary from those reported in the published data as, for our present purposes, we did not require complete study data sets, merely basic anthropometry and an initial peak $\dot{V}O_2$ assessment.

Although in this paper we focus on the interpretation of TM-determined peak $\dot{V}O_2$ from cross-sectional data, the principles and analyses apply equally to any size-related physiological variable measured either directly, such as peak and mean power and muscle strength, or predicted from performance tests.

For complete details of our exercise protocols and measurement procedures over this period, interested readers are referred to previous publications (eg, 6–8). To briefly summarize, all TM peak $\dot{V}O_2$ determinations reported here were made using a protocol with 2- or 3-minute stages, some interspersed with a rest period of a minute. In addition to a rigorous, experienced subjective evaluation of maximal effort, we also applied the objective criteria of a heart rate which was leveling-off over the final stages of the test at a value within 5% of the mean peak heart rate we have previously reported for youth of these ages (8), and a respiratory exchange ratio that exceeded 1.0. Some participants completed verification tests (4) as part of the study protocol, and we are confident that our methods elicited maximal efforts.

We have collated and reanalyzed data from all of these studies to examine and illustrate the limitations of simple ratio scaling, but no additional data "cleaning" has been undertaken. If we accepted the child's effort as maximal in the original study and entered their data into the Centre's database then they were included in the reanalyses presented here.

Why Do We Need to Scale?

Measures of exercise performance such as peak $\dot{V}O_2$ are strongly related to body size, reflecting their underlying dependency on the amount of active muscle mass, and so controlling for size differences is essential if we want to compare individuals or groups. Katch (20), summarizing the contemporary adult literature, reported correlation coefficients approximating .71 for $\dot{V}O_2$ max and body mass. Drawn from our entire cross-sectional data set, Figures 1A and 1B show the relationship between absolute peak $\dot{V}O_2$ (L·min⁻¹) and body mass for boys and girls respectively. Within the 20 individual cross-sectional groups of young people aged 9–18 years, we identified significant ($P < .05$) correlation coefficients (Pearson's r) ranging from .58 to .84 with mean values of $r = .73$ for boys and $r = .75$ for girls. Individual group correlation coefficients are also summarized in Table 1.

The data in Figure 1 confirm that the relationship between body mass and peak $\dot{V}O_2$ is statistically strong, necessitating some size adjustment for comparative purposes. However, the strength of the relationship provides us with the first indication that simple division by body mass will not "remove" the effect of body mass. As Katch stated, ". . . the ratio standard assumes a perfect linear correlation between oxygen uptake (l/min) and body weight . . ." (20, p. 253).

When Can the Ratio Standard Be Used: Testing Assumptions

Simple division by body mass is the usual approach to normalizing data for body size. However, this measure is crude at best.

Table 1 Summary of Age Range, Mean Values for Peak Oxygen Uptake Expressed in Absolute, Relative, and Allometric Terms and the Correlation Coefficients Between These and Body Mass in 20 Individual Study Groups

| Study (Reference) | Age, y | Sex | n | Peak $\dot{V}O_2$, $L \cdot \text{min}^{-1}$ | Correlation coefficients for peak $\dot{V}O_2$ ($L \cdot \text{min}^{-1}$) and body mass | Peak $\dot{V}O_2$ ($\text{mL} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$) | Correlation coefficients for peak $\dot{V}O_2$ ($\text{mL} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$) and body mass | Allometric (b) exponent \pm standard error | Correlation coefficients for peak $\dot{V}O_2$ ($\text{mL} \cdot \text{kg}^{-b} \cdot \text{min}^{-1}$) and body mass |
|-------------------|--------|-----|-----|---|--|--|---|--|---|
| 1 (8) | 11–16 | M | 113 | 2.31 (0.60) | .84, $P < .01$ | 49 (7) | ns | $b = 0.94 \pm 0.06$ | ns |
| 2 (4) | 9–10 | M | 17 | 1.94 (0.24) | .83, $P < .01$ | 62 (5) | -.63, $P < .01$ | $b = 0.64 \pm 0.11^*$ | ns |
| 3 (35) | 9–10 | M | 19 | 1.78 (0.12) | .71, $P < .01$ | 58 (8) | -.87, $P < .01$ | $b = 0.37 \pm 0.09^*$ | ns |
| 4 (47) | 9–10 | M | 16 | 1.68 (0.22) | .61, $P < .01$ | 57 (7) | -.51, $P < .05$ | $b = 0.53 \pm 0.20^*$ | ns |
| 5 (45) | 9–10 | M | 45 | 1.85 (0.25) | .70, $P < .01$ | 53 (7) | -.71, $P < .01$ | $b = 0.47 \pm 0.08^*$ | ns |
| 6 (7) | 10–11 | M | 125 | 1.81 (0.26) | .70, $P < .01$ | 50 (6) | -.54, $P < .01$ | $b = 0.62 \pm 0.05^*$ | ns |
| 7 (3) | 11–12 | M | 36 | 2.06 (0.36) | .74, $P < .01$ | 53 (7) | -.55, $P < .01$ | $b = 0.63 \pm 0.11^*$ | ns |
| 8 (3) | 13–14 | M | 36 | 2.76 (0.51) | .75, $P < .01$ | 54 (6) | -.43, $P < .01$ | $b = 0.76 \pm 0.10^*$ | ns |
| 9 (3) | 16–18 | M | 19 | 3.99 (0.80) | .69, $P < .01$ | 54 (9) | ns | $b = 0.79 \pm 0.19$ | ns |
| 10 (8) | 11–16 | F | 107 | 1.90 (0.36) | .66, $P < .01$ | 41 (6) | -.50, $P < .01$ | $b = 0.61 \pm 0.06^*$ | ns |
| 11 (35) | 9–10 | F | 17 | 1.53 (0.20) | .71, $P < .01$ | 49 (6) | -.62, $P < .01$ | $b = 0.54 \pm 0.12^*$ | ns |
| 12 (4) | 9–10 | F | 17 | 1.80 (0.18) | .80, $P < .01$ | 52 (6) | -.85, $P < .01$ | $b = 0.45 \pm 0.09^*$ | ns |
| 13 (44) | 9–10 | F | 54 | 1.68 (0.27) | .77, $P < .01$ | 49 (6) | -.57, $P < .05$ | $b = 0.65 \pm 0.07^*$ | ns |
| 14 (23) | 9–10 | F | 30 | 1.43 (0.22) | .79, $P < .01$ | 46 (5) | -.37, $P < .04$ | $b = 0.76 \pm 0.10^*$ | ns |
| 15 (7) | 10–11 | F | 128 | 1.62 (0.27) | .83, $P < .01$ | 43 (5) | -.50, $P < .01$ | $b = 0.72 \pm 0.04^*$ | ns |
| 16 (3) | 11–12 | F | 35 | 1.97 (0.31) | .85, $P < .01$ | 47 (5) | -.61, $P < .01$ | $b = 0.68 \pm 0.07^*$ | ns |
| 17 (3) | 12–13 | F | 47 | 2.20 (0.35) | .83, $P < .01$ | 47 (5) | -.69, $P < .01$ | $b = 0.64 \pm 0.06^*$ | ns |
| 18 (34) | 13–14 | F | 38 | 2.32 (0.34) | .69, $P < .01$ | 41 (5) | -.39, $P < .02$ | $b = 0.68 \pm 0.12^*$ | ns |
| 19 (3) | 13–14 | F | 29 | 2.35 (0.28) | .58, $P < .01$ | 46 (5) | ns | $b = 0.75 \pm 0.19$ | ns |
| 20 (3) | 16–18 | F | 25 | 2.59 (0.32) | .69, $P < .01$ | 44 (5) | -.48, $P < .02$ | $b = 0.61 \pm 0.15^*$ | ns |

*Exponent is significantly ($P < .05$) different from $b = 1.0$. ns signifies not significant ($P > .05$).

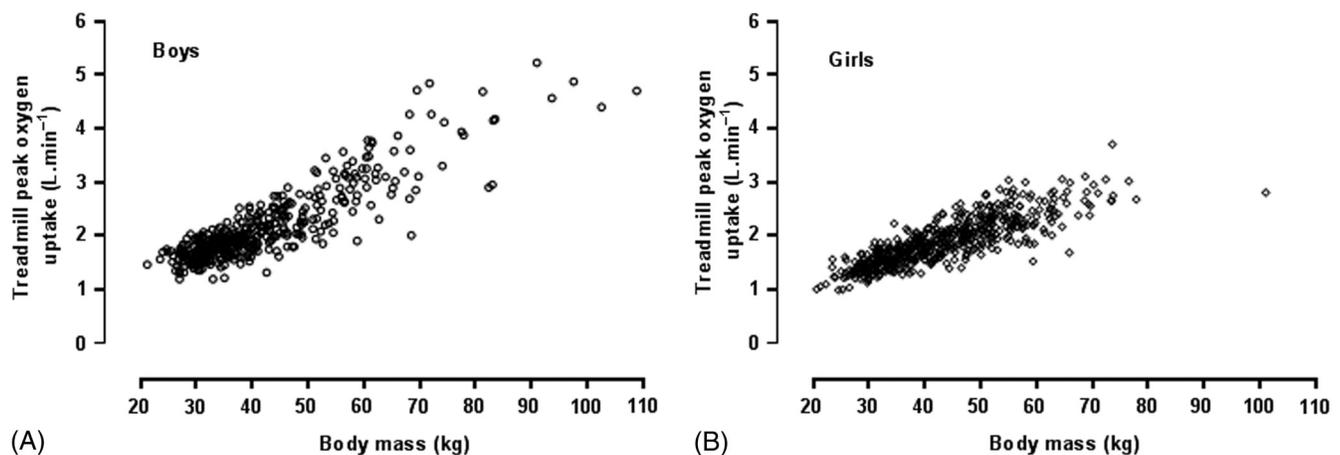


Figure 1 — (A and B) The cross-sectional relationship between peak $\dot{V}O_2$ and body mass in 426 males and 532 females aged 9–18 years from 13 studies comprising 20 individual groups of males and females. Individual group correlation coefficients ranged from $r = .58$ to $r = .84$ with a mean value across the 20 groups of $r = .73$ for boys and $r = .75$ for girls.

We know that additional factors affect peak $\dot{V}O_2$ and individual differences relate best to the relative proportions of lean body mass to fat mass and, in youth, physiological and biochemical changes related to age and biological maturity status. These factors are rarely considered or explored in data sets.

As both Tanner (36) and Katch (20) highlighted, subsequently reiterated by contemporary authors (26,38), application of the ratio standard assumes an underlying set of specific statistical assumptions. We have seen that assumption 1, that of a perfect correlation ($r = 1.0$) between variables is not valid,

and we go on to examine further assumptions in the following sections.

Assumption 2

Expressing peak $\dot{V}O_2$ in $\text{mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ removes the influence of body mass to enable valid comparison of individuals or groups.

Evidence. If the simple ratio, peak $\dot{V}O_2$ divided by body mass, effectively removes the effect of body mass then the ratio standard, $\text{mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$, should not remain significantly correlated with body mass. Table 1 includes correlation coefficients describing this relationship for the individual study groups. In all but 3 of the 20 studies analyzed, a statistically significant ($P < .05$) negative relationship remained with coefficients ranging from $r = -.37$ to $r = -.87$. Computing this relationship offers the researcher a simple technique by which to verify if ratio scaling has removed the influence of body mass.

Assumption 3

The relationship between peak $\dot{V}O_2$ and body mass is best described by a simple linear relationship: Peak $\dot{V}O_2 = b \cdot \text{body mass}$.

Evidence. In statistical terms, when we express peak $\dot{V}O_2$ in $\text{mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ we are making the, usually untested, assumption that the line of best fit is one that extends from 0 (the origin) through the point at which the mean values for body mass and peak $\dot{V}O_2$ intersect. This can be expressed as the simple linear relationship: Peak $\dot{V}O_2 = b \cdot \text{body mass}$, where b is the slope of the line.

We illustrate this in Figure 2 using as an example the data from study number 13 in Table 1 (44), comprising $n = 54$, 9- to 10-year old girls. The dashed line is the line assumed by the ratio standard but a visual inspection of the data suggests that this is not the best fit, and the data would be better represented by a line which *intercepts with the y axis somewhat above 0*. Rather than assuming the existence of a statistical relationship, computing a linear regression with peak $\dot{V}O_2$ and body mass as the independent and dependent variables, respectively, reveals that this does indeed better describe the data.

Although this analysis allows for individual mass-regressed scores for peak $\dot{V}O_2$ to be computed, this method very often does not provide a statistically appropriate analysis for size-related exercise data. Observation of the individual data points in Figures 1 and 2 reveals a feature that is typically observed in size-related

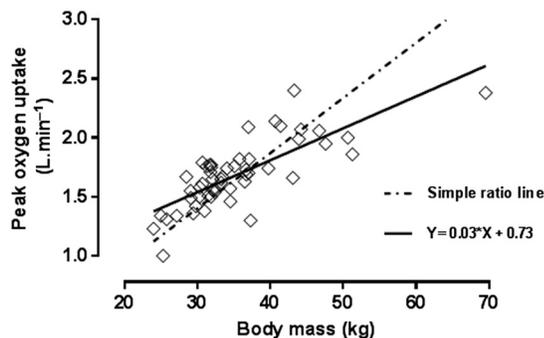


Figure 2 — The assumed simple linear relationship of the ratio standard $\text{mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ differs from the actual computed linear regression of body mass on peak $\dot{V}O_2$ which includes a positive intercept term. Data are from study 13, Table 1: 9- to 10-year-old girls, $n = 54$ (44).

physiological function: data tend to cluster around the regression line at the lower values for both variables and progressively fan out as values increase toward the maximum observed. This indicates that the linear regression we applied to the data previously mentioned does not provide a good fit for the data throughout the range of points measured.

This phenomenon is relatively straightforward to rectify by fitting the linear regression line to data after they have been natural log-transformed:

$$\log_e \text{ peak } \dot{V}O_2 = \log_e a + b \cdot \log_e \text{ mass}$$

This is also the linear form of the curvilinear allometric relationship used extensively to analyze size-related function in biological sciences as we described in the introduction:

$$\text{Peak } \dot{V}O_2 = a \cdot \text{mass}^b$$

Fitting a linear regression to log-transformed data allows us to test the final assumption underlying the validity of the simple ratio, $\text{mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$, because the per-body-mass ratio is an allometric relationship where the b exponent equals 1.0, that is, peak $\dot{V}O_2/\text{mass}^{1.0}$.

Assumption 4

The relationship between body mass and peak $\dot{V}O_2$ is one of direct proportionality, that is, the allometric body mass exponent (b) = 1.0 (and ratio scaling applies).

Evidence. Table 1 shows the values for the mass exponents plus or minus SE computed for the individual groups that comprise our cross-sectional data set. We also indicate the extent of $2 \times \text{SE}$ approximating the 95% confidence intervals. If the confidence intervals exclude 1.0, we can state that the exponent is significantly different from that assumed by the ratio standard. Of the 20 data sets analyzed, in only 3 cases was the mass exponent not significantly different from 1.0. Therefore, in 20 years of measuring peak $\dot{V}O_2$ in almost 1000 young people, our conclusion must be that the weight of evidence does *not* support the assumption that the relationship between body mass and peak $\dot{V}O_2$ is one of direct proportionality.

As we described previously, a simple way to check whether a computed ratio has removed the influence of the size variable is to correlate that ratio with the size variable. The mass exponent derived in the log-linear regression described previously can be used to compute a ratio by dividing peak $\dot{V}O_2$ (in mL) by body mass raised to the value of the b exponent, for example, for study 2 in Table 1 (4) this would be $\text{mass}^{0.64}$. The resulting ratio would be in $\text{mL}\cdot\text{kg}^{-0.64}\cdot\text{min}^{-1}$. We computed these allometrically adjusted ratios using the group-specific mass exponents and correlated them with body mass. The resulting coefficients are presented in Table 1. In all cases, these were nonsignificant ($P > .05$) confirming that this form of adjustment successfully removed the influence of body mass.

To summarize so far, using our data sets as originally collected, we have demonstrated that:

- (1) Peak $\dot{V}O_2$ is significantly correlated with body mass with the strength of that relationship, in groups of similar chronological age, comparable to the strength of the relationship observed in adults ($r \sim .75$).
- (2) The traditional means of controlling for individual and group differences in body mass by computing the simple ratio standard ($\text{mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$) does not remove the

influence of body mass if the ratio remains significantly, negatively correlated with body mass.

- (3) Data are not best fit by the simple linear relationship that underpins ratio scaling, but a log-linear regression is needed to accommodate the spread that is typically evident in size-related exercise data. Using this analysis, mass exponents were, in all but 3 cases, significantly different from the value of 1.0 assumed by the ratio standard.
- (4) Ratios computed using the mass exponents derived from the log-linear regression analyses successfully remove the effect of body mass as evidenced by no residual correlation with body mass.

Variation in Mass Exponents

Table 1 summarizes a large series of cross-sectional data on TM-determined peak $\dot{V}O_2$ collected in the same laboratory over a 20-year period. Within the groups represented, we identified mass exponents ranging from 0.37 to 0.94 in boys and 0.45 to 0.76 in girls.

We would not wish to overinterpret the individual values for exponents obtained, but it is worth highlighting what factors might underlie the variation observed. For example, in the groups of 9- to 10-year-old boys (studies 2–5 in Table 1 [4,35,45,47]), mass exponents ranged from 0.37 to 0.64. Sample size and composition are likely to play key roles in this variation, particularly as they were volunteer groups. Clearly, there are differences in fitness between the groups, and this is underpinned by a large spread in body size even within this predominantly prepubertal population—a fact which Tanner (36) specifically highlighted when cautioning against the uncritical adoption of ratio standards for basal metabolism in children “. . . the variability of weight and surface area for some single-year age groups considerably exceeds that of adults.” (p. 9). In study 5 (45), in Table 1 for example, body mass ranged from under 25 to 58 kg and included 4 overweight (15) boys. This variation in body size, unsurprisingly, becomes more noted in the age groups associated with peak pubertal growth. It is likely that a combination of these factors means that mass exponents will, as demonstrated in our data set, be sample specific.

It is interesting to consider in more detail one example where the log-linear derived mass exponent appeared to offer no significant improvement to traditional ratio scaling (study 1 in Table 1 [8]). In 1991, in what was our initial venture into alternative scaling methods, we presented data on 10- and 15-year-old boys showing a positive effect for age or maturation on peak $\dot{V}O_2$, an effect previously obscured by ratio scaling (46). Therefore, it is likely that the exponent of 0.94, generated on a mixed group of boys aged 11–16 years, is artifactually increased by failure to account for the independent effects of age. Adding age into the log-linear regression equation yielded a significant term, and the value of the mass exponent reduced to $b = 0.74$, $SE = 0.07$, a value now significantly different from 1.0.

In a longitudinal study of peak $\dot{V}O_2$ elsewhere in this issue (2), we observed that a significant age effect masks a more specific effect for changes in fatness relative to overall body mass, that is, reflecting lean body mass. If we also include sum of triceps and subscapular skinfold thickness as a covariate in our example of 11- to 16-year-old boys (study 1, Table 1 [8]), the effect of age becomes nonsignificant $P > .05$, a significant effect of fatness is observed (-0.291 , $SE = 0.04$) and the exponent for mass is increased to 1.08 ($SE = 0.05$).

Interestingly, an inflated exponent was not observed for the girls with a similar within-study age range (study 10, Table 1 [8]), where the mass exponent was computed as $b = 0.61$. When added to the log-linear regression analysis here, and in contrast to the boys, age did not add significantly to the model; however, the addition of skinfold thickness did yield a significant, negative exponent as also observed in our recent longitudinal analysis of peak $\dot{V}O_2$ in similar-aged girls (2).

Is There a Universal Alternative to Ratio Scaling?

Perhaps one of the drivers underpinning the continued use of the ratio standard is the absence of a universally applicable alternative. As ourselves and others have discussed, the mass exponent most often suggested for adoption is $b = 0.67$, a value derived from theories of geometric similarity (see 11,31). We have observed values very close to this in large sample longitudinal data, but only after adjusting for other significant covariates such as age, maturity status, and skinfold thickness (7). Although we could generalize from our cross-sectional data presented here and say that the computed mass exponents are closer to 0.67 than they are to 1.0—indeed the value of 0.67 falls within the 95% confidence intervals of all but 3 of the 20 exponents computed in Table 1—we do not recommend the universal adoption of this as a value.

We have seen that sample size and composition affects the value of computed exponents, and we know that the effects of age and maturity status may need to be considered, particularly in boys. In addition, from our own and others' analyses, we know that peak $\dot{V}O_2$ is influenced by other morphological covariates, notably stature and skinfold thickness. In our most recent longitudinal analyses of 10- to 18-year-old young people (2), we have demonstrated that both body mass and skinfold thickness should be controlled for—perhaps unsurprising given that, in combination, this gives us a better estimate of lean body mass than just controlling for total body mass alone.

Does it Really Matter?

Over 30 years ago we raised concerns with the use of field tests, notably the 20-m shuttle run, to predict children's aerobic fitness (9). Recent years, however, have seen an explosion in the use of this test with data reported on literally hundreds of thousands of children, including for those as young as 3 years old. Such tests have been recommended and accepted for inclusion in fitness test batteries for children for population level surveillance (21,33) and as the basis for classifying youth fitness, with levels of 42 and 35 $mL \cdot kg^{-1} \cdot min^{-1}$ for boys and girls, respectively, identified as “Clinical Red Flags”—potentially warranting intervention (21,33).

While this explosion stems from a desire to promote youth fitness for the prevention of current and future disease, based upon relationships between ratio-scaled measures of fitness and a range of health indicators (32), our genuine concerns lie with the spurious relationships and recommendations that might potentially emerge from laboratory or field tests that predict peak $\dot{V}O_2$ in $mL \cdot kg^{-1} \cdot min^{-1}$ and adversely or inappropriately impact on the promotion of young people's health and well-being.

We also have serious concerns regarding the increasing use of maximal exercise testing with ratio scaling of data in the fitness assessment of children with a range of severe or potentially life-limiting diseases of, for example, the lungs (22,27) muscles

(14) or heart (1,18). As we have already mentioned, not only did Tanner (36) explicitly explain why caution should be applied to the generation of per body mass standards in children, he also demonstrated that the use of such standards would lead to inappropriate conclusions when subsequently used for investigating relationships with other lifestyle, physiological, or behavioral factors. Although the statistical analyses within which the ratio-scaled and/or predicted values are being used may be somewhat more sophisticated these days, the underlying principles remain: If the use of $\text{mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ cannot be demonstrated to appropriately describe children's aerobic fitness, then any comparisons, conclusions, or recommendations based upon per body mass standards are likely to be spurious.

Conclusion

Interpretation of peak $\dot{V}\text{O}_2$ in youth is not straightforward, and based upon the evidence discussed within this paper, indiscriminate use of $\text{mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ cannot be sanctioned. We know of no other scientific discipline where an assumed relationship albeit "convenient and traditional" has become the acceptable alternative to rigorous scientific justification.

The issues we raise are not new, having been eloquently and completely described by visionary scientists since 1949, and which we initially applied to children's data in 1991 and have continued to champion for almost 30 years. We strongly urge exercise scientists, medics, and population health experts to consider the issues raised in this paper and to use the simple statistical techniques suggested to check whether a simple per body mass ratio is justified in all studies of children's fitness. Tanner's "tongue in cheek" comment from nearly 70 years ago still resonates today ... rather than raising a "Clinical Red Flag" perhaps children with levels of fitness below the international standards recommended above should be investigated not for cardiovascular risk but for "no more formidable a disease than statistical artefact" (36, p. 3).

Acknowledgment

The authors gratefully acknowledge the support of the Darlington Trust.

References

1. Amedro P, Gavotto A, Guillaumont S, et al. Cardiopulmonary fitness in children with congenital heart diseases versus healthy children. *Heart*. 2018;104:1026–36. PubMed ID: [29170358](#) doi:[10.1136/heartjnl-2017-312339](#)
2. Armstrong N, Welsman J. Sex-specific longitudinal modeling of youth peak oxygen uptake. *Pediatr Exerc Sci*. 2019;31. doi:[10.1123/pes.2018-0175](#)
3. Armstrong N, Welsman J. *Unpublished Studies of Peak Oxygen Uptake*. Exeter, UK: Children's Health and Exercise Research Centre, University of Exeter; 1986–2006.
4. Armstrong N, Welsman J, Winsley R. Is peak VO_2 a maximal index of children's aerobic fitness? *Int J Sports Med*. 1996;17:356–9. PubMed ID: [8858407](#) doi:[10.1055/s-2007-972860](#)
5. Armstrong N, Welsman JR. *Young People and Physical Activity*. Oxford, UK: Oxford University Press; 1997:6–55.
6. Armstrong N, Welsman JR, Kirby BJ. Peak oxygen uptake and maturation in 12-yr-olds. *Med Sci Sports Exerc*. 1998;30:165–9. PubMed ID: [9475659](#) doi:[10.1097/00005768-199801000-00023](#)
7. Armstrong N, Welsman JR, Nevill AM, Kirby BJ. Modeling growth and maturation changes in peak oxygen uptake in 11–13 yr olds. *J Appl Physiol*. 1999;87:2230–6. PubMed ID: [10601172](#) doi:[10.1152/jappl.1999.87.6.2230](#)
8. Armstrong N, Williams J, Balding J, Gentle P, Kirby B. The peak oxygen uptake of British children with reference to age, sex and sexual maturity. *Eur J Appl Physiol Occup Physiol*. 1991;62:369–75. doi:[10.1007/BF00634975](#)
9. Armstrong N, Williams J, Ringham D. Peak oxygen uptake and progressive shuttle-run performance in boys aged 11–14 years. *Br J Phys Educ Res Suppl*. 1988;4:10–11.
10. Åstrand PO. *Experimental Studies of Physical Working Capacity in Relation to Sex and Age*. Copenhagen, Denmark: Munksgaard; 1952.
11. Åstrand PO, Rodahl K. *Textbook of Work Physiology*. New York, NY: McGraw-Hill; 1970:319–40.
12. Bar-Or O. *Pediatric Sports Medicine for the Practitioner*. New York, NY: Springer-Verlag; 1983.
13. Bar-Or O, Rowland TW. *Pediatric Exercise Medicine*. Champaign, IL: Human Kinetics; 2004.
14. Bartels B, Takken T, Blank AC, van Moorsel H, van der Pol WL, de Groot JF. Cardiopulmonary exercise testing in children and adolescents with dystrophinopathies: a pilot study. *Pediatr Phys Ther*. 2015;27:227–34. PubMed ID: [26102164](#) doi:[10.1097/PEP.0000000000000159](#)
15. Cole TJ, Bellizzi MC, Flegal KM, Dietz WH. Establishing a standard definition for child overweight and obesity worldwide: international survey. *BMJ*. 2000;320:1240. PubMed ID: [10797032](#) doi:[10.1136/bmj.320.7244.1240](#)
16. Henderson Y, Haggard HW. The maximum of human power and its fuel. *Am J Physiol*. 1925;72:264–82. doi:[10.1152/ajplegacy.1925.72.2.264](#)
17. Hill AV, Lupton H. Muscular exercise, lactic acid, and the supply and utilization of oxygen. *Q J Med*. 1923;os-16:135–71. doi:[10.1093/qjmed/os-16.62.135](#)
18. Hock J, Reiner B, Neidenbach RC, et al. Functional outcome in contemporary children with total cavopulmonary connection—health-related physical fitness, exercise capacity and health-related quality of life. *Int J Cardiol*. 2018;255:50–4. PubMed ID: [29425566](#) doi:[10.1016/j.ijcard.2017.11.092](#)
19. Huxley JS. On the relation between egg-weight and body-weight in birds. *Zool J Linn Soc*. 1927;36:457–66. doi:[10.1111/j.1096-3642.1927.tb02180.x](#)
20. Katch VL. Use of the oxygen/body weight ratio in correlational analyses: spurious correlations and statistical considerations. *Med Sci Sports*. 1973;5(4):253–7. PubMed ID: [4774204](#)
21. Lang JJ, Tomkinson GR, Janssen I, et al. Making a case for cardiorespiratory fitness surveillance among children and youth. *Exerc Sport Sci Rev*. 2018;46:66–75. PubMed ID: [29346159](#) doi:[10.1249/JES.0000000000000138](#)
22. Madsen A, Green K, Buchvald F, Hanel B, Neilsen KG. Aerobic fitness in children and young adults with primary ciliary dyskinesia. *PLoS ONE*. 2013;8(8):e71409. doi:[10.1371/journal.pone.0071409](#)
23. McManus AM, Armstrong N, Williams CA. Effect of training on the aerobic power and anaerobic performance of prepubertal girls. *Acta Paediatr*. 1997;86:456–9. PubMed ID: [9183481](#) doi:[10.1111/j.1651-2227.1997.tb08912.x](#)
24. Mirwald RL, Bailey DA. *Maximal Aerobic Power*. London, Canada: Sports Dynamics; 1986.
25. Morse M, Schlutz FW, Cassels DE. Relation of age to physiological responses of the older boy (10–17 years) to exercise. *J Appl Physiol*. 1949;1:683–709. PubMed ID: [18118960](#) doi:[10.1152/jappl.1949.1.10.683](#)

