Traditional and New Perspectives on Youth Cardiorespiratory Fitness

NEIL ARMSTRONG, and JO WELSMAN

Children's Health and Exercise Research Centre, University of Exeter, Exeter, UNITED KINGDOM

ABSTRACT

ARMSTRONG, N., and J. WELSMAN. Traditional and New Perspectives on Youth Cardiorespiratory Fitness. *Med. Sci. Sports Exerc.*, Vol. 52, No. 12, pp. 2563–2573, 2020. **Purpose:** This study aimed to review traditional and new perspectives in the interpretation of the development of youth cardiorespiratory fitness (CRF). **Methods:** We analyzed data from (i) the literature which for 80 yr has been traditionally based on interpretations of peak oxygen uptake ($\dot{V}O_2$) in ratio with body mass (BM) and (ii) recent multilevel allometric models founded on 994 (475 from girls) determinations of 10- to 16-yr-olds' peak $\dot{V}O_2$ with measures of age, maturity status, and morphological covariates (BM and fat-free mass), and from 10 to 13 yr, 110 peak $\dot{V}O_2$ determinations of maximum cardiovascular covariates (stroke volume, cardiac output, and arteriovenous oxygen difference). **Results:** The application of ratio scaling of physiological variables requires satisfying specific statistical assumptions that are seldom met. In direct conflict with the ratio-scaled data interpretation of CRF, multilevel allometric modeling shows that with BM controlled, peak $\dot{V}O_2$ increases with age but the effect is smaller in girls than boys. Maturity status exerts a positive effect on peak $\dot{V}O_2$, in addition to those of age and BM. Changes in maximum cardiovascular covariates contribute to explaining the development of CRF, but fat-free mass (as a surrogate for active muscle mass) is the most powerful single influence. With age, maturity status, morphological covariates, and maximum cardiovascular covariates controlled, there remains an unexplained ~4% to ~9% sex difference in peak $\dot{V}O_2$. **Conclusions:** The traditional interpretation of peak $\dot{V}O_2$ in ratio with BM is fallacious and leads to spurious correlations with other health-related variables. Studies of the development of CRF require analyses of sex-specific, concurrent changes in age- and maturation-driven morphological and maximum cardiovascular covariates. Multilevel allometric modeling provides a rigorous, flexible, and sensitive method of data analysis. **Key Words:** CHILDREN, FAT-FREE MASS, MAXIMUM CARDIOVASCULAR VARIABLES, PEAK OXYGEN UPTAKE, MULTILEVEL ALLOMETRIC MODELING

Cardiorespiratory fitness (CRF) defines the integrated ability of the body to deliver oxygen from the atmosphere to the muscles and to use it to generate energy to support muscle activity during exercise. In adults, CRF is considered by the American Heart Association as a "reflection of total body health" ([1], p. e654). The health-related benefits of CRF in youth are also widely recognized (2) but often misinterpreted through fallacious assessments and spurious interpretations (3–5).

Address for correspondence: Neil Armstrong Ph.D., D.Sc., F.A.C.S.M., Children's Health and Exercise Research Centre, University of Exeter, St Lukes Campus, Heavitree Road, Exeter EX1 2LU, United Kingdom; E-mail: N.Armstrong@exeter.ac.uk.
Submitted for publication March 2020.
Accepted for publication May 2020.

0195-9131/20/5212-2563/0
MEDICINE & SCIENCE IN SPORTS & EXERCISE®

DOI: 10.1249/MSS.0000000000002418

No single physiological variable describes fully youth CRF (for reviews, see [6–8]), but peak oxygen uptake ($\dot{V}O_2$) the highest rate of oxygen consumed during a progressive exercise test to exhaustion, limits the capacity to perform aerobic exercise and is internationally recognized as the "gold standard" measure of youth CRF (6,9,10). Peak $\dot{V}O_2$ is the most comprehensively researched physiological variable in the history of developmental exercise physiology (11), but despite over eight decades of intensive study understanding of the development of peak $\dot{V}O_2$ with growth and maturation is both confused and shrouded in controversy.

The present paper uses the terms peak $\dot{V}O_2$ and CRF synonymously and challenges the "convenient and traditional" (12) interpretation of the development of CRF in youth. It uses multilevel allometric modeling of longitudinal data sets to elucidate the sex-specific development of peak $\dot{V}O_2$ in 10- to 16-yr-olds, in relation to concurrent changes in age, maturity status, morphological covariates, and maximum cardiovascular covariates (13–16).

TRADITIONAL PERSPECTIVES

The first laboratory-based "experimental studies of physical fitness" to include measurements of boys' $\dot{V}O_2$ were published in 1938 (17) and 1949 (18). These pioneers assessed

the "physical fitness" of boys as young as 6 yr of age but readily acknowledged the difficulty in achieving true maximum values as "the youngest boys were unwilling to continue work after it ceased to be fun" ([17], p. 281). There remain few secure data on young children (for reviews, see [6,19,20]), and the present paper is therefore focused on children 10 yr old and above.

Peak Oxygen Uptake and Age

Cross-sectional studies. Robinson reported the "highest $\dot{V}O_2$ attained in maximal work" ([17], p. 279) to increase with age. Åstrand (21) in the first study to include girls, observed a similar trend in both sexes with boys' mean peak $\dot{V}O_2$ higher than girls' mean peak $\dot{V}O_2$ at all ages studied. He reported the sex difference in peak $\dot{V}O_2$ to increase from 10 to 16 yr and to reach ~36% at age 16/17 yr. These sex-specific, age-related trends in mean peak $\dot{V}O_2$ have been confirmed in numerous subsequent cross-sectional studies (see [6,12,19] for reviews).

Longitudinal studies. Cross-sectional studies only offer a single "snapshot" in time. Longitudinal studies where the same individuals are measured repeatedly over time, offer an opportunity to monitor and analyze developmental changes in CRF. There are, however, remarkably few rigorous longitudinal investigations of both boys and girls. Reports generally consist of annual analyses ("snapshots") of mean peak $\dot{V}O_2$ at specific ages and collectively longitudinal data reported in this manner unsurprisingly reflect cross-sectional trends (see [6,13] for reviews).

Pediatric norms. Compilations of data relating mean values of peak $\dot{V}O_2$ to chronological age have been used to construct pediatric norms (22) but mean age-related values of peak $\dot{V}O_2$ do not consider changes in body mass (BM), body composition, or maturity status which progress in accord with individual biological clocks (20). Age-related comparisons using pediatric norms therefore provide few insights into the development of CRF in youth (23).

In pediatric exercise laboratories peak $\dot{V}O_2$ ($L \cdot min^{-1}$) is traditionally determined either running on a treadmill or pedaling on a cycle ergometer, and the interpretation of pediatric norms has been further confused by the common practice of increasing sample size by combining data regardless of sex, age, or maturity status. Ergometer differences are normally "corrected" by adding fixed arbitrary percentages (usually +5% or +7.5%) to cycle ergometer values but longitudinal data have demonstrated that the ~11% to ~14% differences within individuals in treadmill and cycle ergometer data vary over time with sex, age, and maturity status (14). The use of data pooled from different ergometers to describe the development of CRF, to establish "normal" values, or to promote the use of health-related CRF cut points (e.g., [12,24,25]) has therefore contributed to clouding the interpretation of CRF in youth, and misguiding recommendations designed to promote the cardiovascular health of children and adolescents.

Figure 1 describes 994 longitudinal measures of peak $\dot{V}O_2$ (13) The data show boys' peak $\dot{V}O_2$ to increase from age 10 to 16 yr and girls' values to rise from age 10 to ~13/14 yr before tapering off to age 16 yr. Overall, mean values show boys' peak $\dot{V}O_2$ to increase by ~82% and girls' peak $\dot{V}O_2$ to increase by ~44% from 10 to 16 yr with the mean sex difference in peak $\dot{V}O_2$ increasing from ~9% at age 10 yr to ~39% at 16 yr. However, the wide individual variations in peak $\dot{V}O_2$ in both sexes at all ages expose the limitations of age-related comparisons.

Peak Oxygen Uptake and Body Mass

As illustrated in Figure 1, boys' peak $\dot{V}O_2$ increases with BM from 10 to 16 yr of age but girls' values tend to level-off from ~60 kg. BM, however, includes both fat mass which is largely metabolically inert (27) and fat-free mass (FFM) which reflects active muscle mass and therefore better relates to peak $\dot{V}O_2$ (4,6,20). Figure 1 shows the strong association between peak $\dot{V}O_2$ and FFM in both sexes, from 10 through 16 yr of age.

Cross-sectional studies. To account for growth and maturation, the pioneers of pediatric exercise science focused on controlling for total BM and initiated a methodology which has confused understanding of youth CRF for over 80 yr. Robinson (17) initially reported his $\dot{V}O_2$ data in liters per minute and then without explanation "referred them to body weight" ([17], p. 280) by dividing peak $\dot{V}O_2$ ($mL \cdot min^{-1}$) by BM (kg) and expressing it as a ratio in milliliters per kilogram per minute ($mL \cdot kg^{-1} \cdot min^{-1}$). Åstrand (21) also discussed group comparisons of peak $\dot{V}O_2$ in both liters per minute and milliliters per kilogram per minute but insightfully commented that group comparisons of peak $\dot{V}O_2$ should ideally be interpreted in relation to active muscle mass. He did not, however, pursue empirically this line of enquiry. Morse et al. (18), without presenting either an underpinning rationale or a statistical justification, only reported their data in ratio with body mass. Subsequent pediatric exercise studies have consistently retained a focus on ratio scaling peak $\dot{V}O_2$ with total BM and disregarded concurrent sex-specific changes with age and maturity status in other morphological covariates. This practice has confounded the interpretation of CRF during growth and maturation.

In a classic paper, Tanner (28) elegantly demonstrated that expressing physiological variables in ratio with BM was "theoretically fallacious" (p. 14) and showed that correlations using ratio-scaled data resulted in specious conclusions that were physiologically untenable. In this journal almost 50 yr ago, Katch (29) comprehensively demonstrated with worked examples that ratio-scaled peak $\dot{V}O_2$ did not control for BM but remained correlated with BM and when used in subsequent correlational analyses produced spurious correlations (see also [30,31]).

In 1992, we showed empirically that ratio scaling of peak $\dot{V}O_2$ to control for BM misinterprets boys' CRF in relation to age and maturity status (32). In a series of subsequent tutorial papers (e.g., [33–35]), we argued that with cross-sectional data, allometric scaling with multiple covariates based in log–linear regression is the method of choice in the exploration of changes in peak $\dot{V}O_2$ with BM, body composition, age, and maturity status. The tutorial papers demonstrated both theoretically from first principles and empirically using large data sets that ratio scaling of peak $\dot{V}O_2$ does not control for BM in children or adolescents and misinterprets the development of

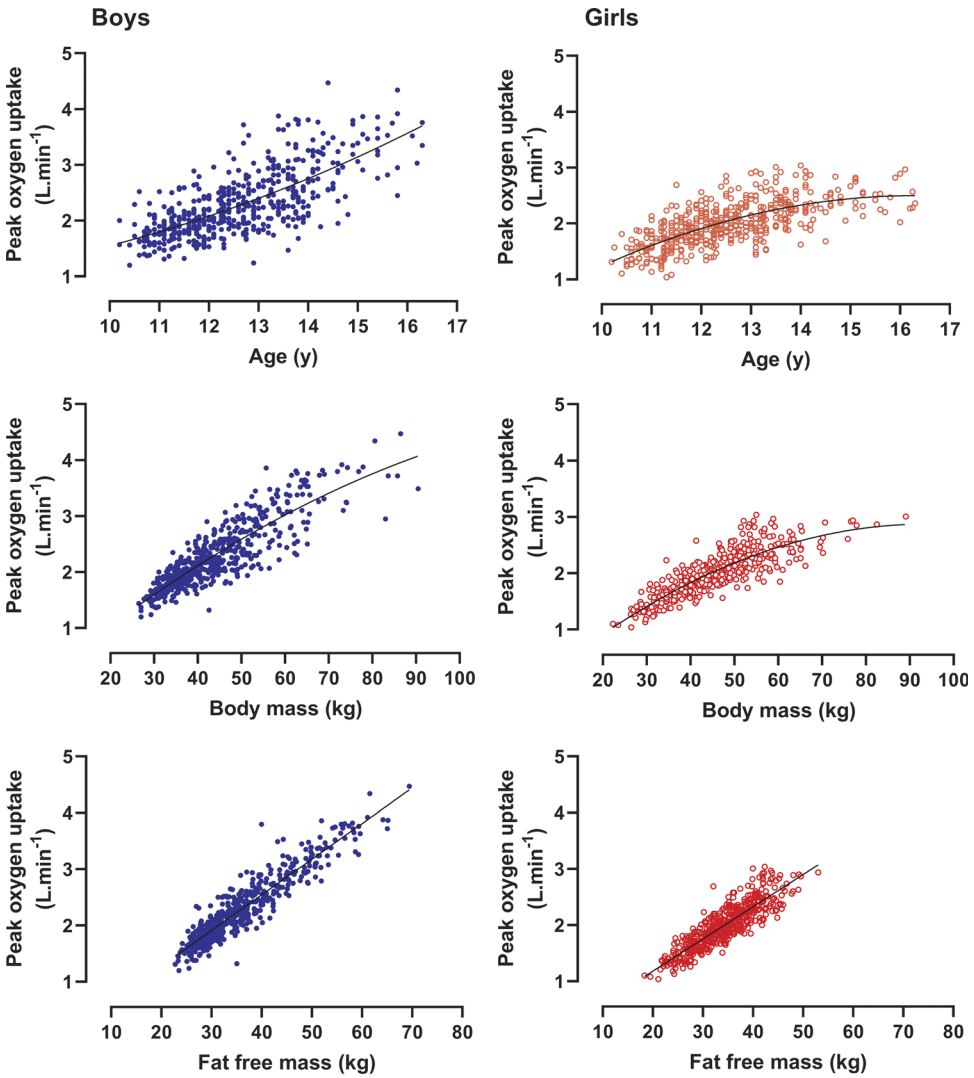

FIGURE 1—Peak oxygen uptake in relation to age, body mass, and FFM in 10- to 16-yr-olds. Figures founded on 994 determinations of peak $\dot{V}O_2$, girls (n = 475), boys (n = 519). FFM estimated from Slaughter et al. (26). Data from Armstrong and Welsman (13).

CRF. We recently confirmed this using cross-sectional data from 128 girls and 125 boys, 11 yr old, from the study described in Figure 1. In brief, if ratio scaling effectively controls for BM, then the product–moment correlation coefficient between peak $\dot{V}O_2$ ($mL \cdot kg^{-1} \cdot min^{-1}$) and BM (kg) would not be significantly different from zero. Significant ($P < 0.001$) correlations of $r = -0.52$ and $r = -0.54$ for boys and girls, respectively, were recorded clearly demonstrating that ratio scaling did not effectively control for BM. Analyzing the same data using allometric (log–linear) scaling revealed a BM exponent of 0.68 and subsequent correlations between allometrically scaled peak $\dot{V}O_2$ ($mL \cdot kg^{-0.68} \cdot min^{-1}$) and BM (kg) of $r = -0.13$ and $r = 0.07$ for boys and girls, respectively, which were not significantly different from zero. Allometric scaling therefore effectively controlled peak $\dot{V}O_2$ for BM ([6], pp. 168–9).

Ratio scaling peak $\dot{V}O_2$ with BM favors lighter youth (e.g., clinically underweight or delayed maturing) and penalizes heavier youth (e.g., overweight or advanced maturing). Therefore, studies that control for BM using ratio scaling consistently report boys' peak $\dot{V}O_2$ to remain stable from 10 to 16 yr, girls' values to decline from ~13 to 16 yr of age, and CRF to be unrelated to maturity status (for reviews, see [6,12,19]). By contrast, with BM appropriately controlled for using allometry, boys' peak $\dot{V}O_2$ has been showed to increase from 10 to 16 yr and girls' peak $\dot{V}O_2$ to level off rather than decline from 13 to 22 yr (36). Moreover, allometric scaling has demonstrated that maturity status exerts significant, positive, sex-specific effects on peak $\dot{V}O_2$ in addition to and independent of age and BM (37).

Associations of ratio-scaled peak $\dot{V}O_2$ with cardiovascular risk factors in overweight or obese youth are likely to be spurious and reflect overweight (or over fatness) to a greater extent than CRF (i.e., peak $\dot{V}O_2$ in liters per minute) (4). A recent systematic review emphasized how most publications investigating youth CRF in relation to cardiovascular health used ratio-scaled peak $\dot{V}O_2$ as their criterion measure of CRF and did not account for confounding factors such as BM or adiposity. The systematic review noted the presence

of spurious correlations so that, for example, a high peak $\dot{V}O_2$ appeared to be associated with lower body fatness and a lower ratio of total cholesterol to high-density lipoprotein cholesterol, but the spurious associations were only present when peak $\dot{V}O_2$ was expressed in ratio with BM (38). Moreover, the growing popularity (e.g., [25,39,40]) of specific values of ratio-scaled peak $\dot{V}O_2$ acting as "cut-points to classify children and youth with poor cardiometabolic health" and "to inform teachers and health authorities who to target with physical activity initiatives" ([39], p. 248) have further confounded understanding of any relationships between CRF and other health-related variables.

Longitudinal Studies

Over 30 yr ago, influential longitudinal studies readily acknowledged that ratio scaling peak $\dot{V}O_2$ is "open to criticism because of the spurious correlations that go with these indices" ([41], p. 119) and that "considerable difficulties in interpretation can arise" ([42], pp. 19–20). But with few innovative exceptions (e.g., [43,44]), longitudinal studies have persisted in interpreting mean age-related peak $\dot{V}O_2$ data in ratio with BM. Longitudinal studies have therefore generally reflected the trends reported in cross-sectional studies with the same limitations, and even further confounded interpretation of the development of CRF during growth and maturation.

Performance Tests

The interpretation of youth CRF has been misled by the 20-fold increase in the last decade of publications predicting peak $\dot{V}O_2$ from performance on 20-m shuttle run tests (20mSRT) (45). 20mSRT performance is not a physiological measure of CRF but a function of the willingness and capability of individuals to transport their BM between two lines 20 m apart while keeping pace with audio signals, which require running speed to increase each minute. We have extensively documented (3,46,47) the specious logic underpinning 20mSRT performance as a surrogate of youth CRF. Empirically, we have demonstrated that limits of agreement between 11- and 14-yr-old boys' directly determined peak $\dot{V}O_2$ and 20mSRT-predicted peak $\dot{V}O_2$ are only within ~40% (48). Similarly, a recent review investigated the validity of the 20mSRT in predicting peak $\dot{V}O_2$ through an analysis of 10 studies and concluded that for 9- to 17-yr-olds, "the 95% likely range for a true peak $\dot{V}O_2$ value estimated from the 20mSRT is ~10 $mL \cdot kg^{-1} \cdot min^{-1}$ or ~24%" ([49], p. 154). Moreover, a meta-analysis reported that over 50% of reported correlation coefficients between children's 20mSRT-predicted peak $\dot{V}O_2$ and laboratory-determined peak $\dot{V}O_2$ explained less than half the shared variance. It concluded that "the performance score of the 20mSRT is simply an estimation and not a direct measure of cardiorespiratory fitness" ([50], p. 536).

Peak $\dot{V}O_2$ predicted from 20mSRT performance is expressed in milliliters per kilogram per minute and, therefore, subject also to all the criticisms associated with traditional ratio scaling. In addition, during a 20mSRT, fat mass is carried as "deadweight," and the more fat carried, the greater the work done in each shuttle. This negatively affects 20mSRT performance and lowers the prediction of peak $\dot{V}O_2$ without influencing true peak $\dot{V}O_2$. Overweight and obese youth are doubly penalized by not only having to carry their metabolically inert fat mass during a 20mSRT but also having their performance score expressed as peak $\dot{V}O_2$ divided by BM (including fat mass).

A recent study (48) demonstrated empirically the absence of a significant relationship between surrogates of youth CRF predicted from 20mSRT performance and rigorously determined peak $\dot{V}O_2$. Figure 2A, using data from 11- to 14-yr-old boys, illustrates a nonsignificant correlation ($r = -0.25$, $P > 0.05$) between 20mSRT predictions of ratio-scaled peak $\dot{V}O_2$ and rigorously determined peak $\dot{V}O_2$ with BM appropriately controlled for using allometry. Figure 2B, using the same data set, illustrates a similar nonsignificant correlation ($r = -0.19$, $P > 0.05$) between 20mSRT predictions of ratio-scaled peak $\dot{V}O_2$ and rigorously determined peak $\dot{V}O_2$ appropriately controlled for both BM and age (48). Responding to criticisms of the use of ratio scaling of peak $\dot{V}O_2$ to express 20mSRT performance, advocates of the 20mSRT have suggested that running speed at the last completed shuttle may provide the best estimate of CRF (49). However, Figure 2 shows nonsignificant ($P > 0.05$) correlations when rigorously determined peak $\dot{V}O_2$ is compared with maximum 20mSRT speed. With BM appropriately controlled for (Fig. 2C) and both age and BM appropriately controlled for (Fig. 2D), the correlations were $r = -0.07$ and $r = 0.06$, respectively (48).

A typical example of how promotion of 20mSRT performance as a valid measure of youth CRF has seriously and adversely affected pediatric exercise and health science is the emergence and increasing popularity of "cardiometabolic risk factor cut-points" (25) or "clinical red flags" (40), founded on collations of cross-sectional data sets of 20mSRT performance. Predicted ratio-scaled peak $\dot{V}O_2$ values for 8- to 18-yr-olds below 42 and 35 $mL \cdot kg^{-1} \cdot min^{-1}$ for boys and girls, respectively, are suggested to raise a "clinical red flag" to identify "children and adolescents who may benefit from primary and secondary cardiovascular prevention programming" ([40], p. 1451). However, as will become evident later in this paper, CRF develops in accordance with concurrent changes in age, maturity status, morphological covariates, and maximum cardiovascular covariates. The timing and tempo of these changes are specific to individuals, and to classify 8- to 18-yr-old prepubertal, pubertal, and postpubertal youth based on a single predicted ratio-scaled value of peak $\dot{V}O_2$ is meaningless and potentially not in the best interests of some children and adolescents if acted upon.

NEW PERSPECTIVES

Nevill et al. (51) applied multilevel allometric modeling to developmental exercise physiology with a reanalysis of previously published data from elite youth athletes. Armstrong et al. (52) simultaneously applied the technique to a longitudinal study of 11- to 13-yr-olds' CRF. Collectively, these authors demonstrated that multilevel allometric modeling enabled the effects of sex, age, maturity status, and morphological

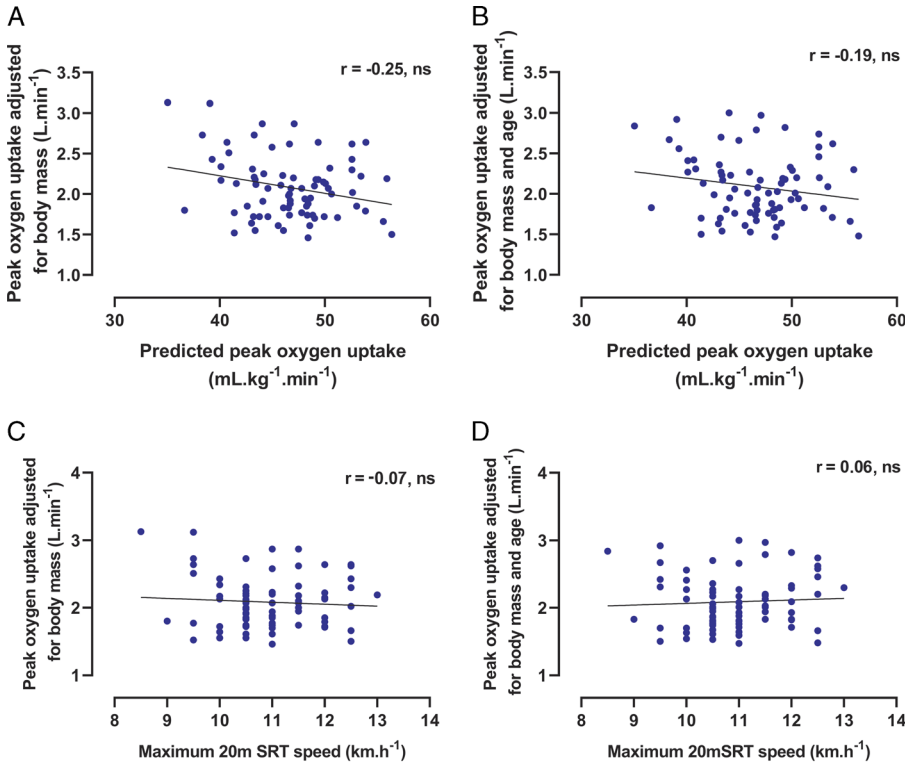

FIGURE 2—Peak oxygen uptake allometrically adjusted for body mass and age in relation to 20-m shuttle run test performance. A and B, founded on 76 determinations of 11- to 14-yr-old boys' peak $\dot{V}O_2$ allometrically adjusted for body mass (A) and body mass and age (B) in relation to peak $\dot{V}O_2$ in ratio with body mass as predicted from 20mSRT performance. C and D, founded on the same data with peak $\dot{V}O_2$ allometrically adjusted for body mass (C) and body mass and age (D) in relation to maximum speed attained in a 20mSRT. Data from Welsman and Armstrong (48).

covariates on peak $\dot{V}O_2$ to be partitioned concurrently within an allometric framework to provide a sensitive analysis of trained and untrained youth CRF. Further applications of multilevel allometric modeling of longitudinal data to developmental exercise physiology have, however, been remarkably sparse and recently focused on the development of short-term power output (53,54), although a regularly updated computer program and a comprehensive user's guide (55) are freely available.

Armstrong and Welsman (13–16) used multilevel modeling to investigate the longitudinal development of the peak $\dot{V}O_2$ of the 10- to 16-yr-olds described in Figure 1. The analyses were founded on the following baseline model where y is peak $\dot{V}O_2$,

$$y = \text{mass}^{k} \times \exp\left(a_j + b \times \text{age} + c \times \text{age}^2\right) \varepsilon_{ij} \qquad [1]$$

log transformation linearized the model to form the starting point for analysis,

$$\log_e y = k \times \log_e \text{mass} + a_j + b \times \text{age} + \text{c} \times \text{age}^2 + \log_e\left(\varepsilon_{ij}\right) \quad [2]$$

All parameters were fixed except the constant (a), which was allowed to vary randomly at level 2 (between individuals), and the multiplicative error ratio (ε), which also varied randomly at level 1 (within individual) as denoted by the subscripts i (level 1 variation) and j (level 2 variation). Age was centered on the group mean. In some models (e.g., models 1.2, 1.4, and 1.6), BM was replaced by FFM. In the initial modeling of the combined groups, sex differences were investigated using the indicator boys = 0, girls = 1, plus the age by sex interaction term. In subsequent analyses, effects due to maturation were explored through entry of the pubic hair (PH) stages described by Tanner (56), with stage 1 set as the baseline from which other stages varied.

In the following sections, multilevel allometric models of peak $\dot{V}O_2$ are used to illustrate new perspectives on the development of youth CRF. The recruitment of participants, the experimental methodology, and the detailed analyses of modeling longitudinal data are described in the original papers (13–16), with statistical significances set at $P < 0.05$.

Morphological covariates. Table 1 is founded on the 994 determinations of peak $\dot{V}O_2$ illustrated in Fig. 1. The baseline model (model 1.1) shows the concurrent effects of age and body mass on the CRF of 10- to 16-yr-olds and confirms the fallacy of traditional ratio scaling of peak $\dot{V}O_2$ in the interpretation of age-related CRF. The allometric exponent for BM is 0.68 with an exponent of 1.0 as required in ratio scaling, falling outside the 95% confidence limits. In conflict with the traditional ratio-scaled interpretation, the positive age term shows that with BM controlled, peak $\dot{V}O_2$ increases with age in both sexes, the negative age by sex term shows the age effect to be smaller in girls, and the negative age^2 term signifies that the size of the age effect reduces as the rate of change in growth decreases. The negative sex term indicates that with BM and age controlled, there remains a ~15% sex difference in peak $\dot{V}O_2$. In model 1.2, BM is replaced with estimated FFM, and

TABLE 1. Multilevel allometric models of the peak oxygen uptake of 10- to 16-yr-olds.

	Model 1.1	Model 1.2	Model 1.3 (Boys)	Model 1.4 (Boys)	Model 1.5 (Girls)	Model 1.6 (Girls)
Response	Log_e Peak $\dot{V}O_2$					
Fixed part						
Constant	−1.766 (0.085)	−2.284 (0.088)	−1.704 (0.125)	−2.358 (0.114)	−1.715 (0.126)	−2.282 (0.143)
Log_e body mass	0.684 (0.022)	–	0.656 (0.034)	–	0.623 (0.034)	–
Age	0.087 (0.007)	0.035 (0.008)	0.034 (0.006)	0.024 (0.005)	0.025 (0.006)	0.021 (0.005)
Age^2	−0.006 (0.001)	−0.005 (0.001)	ns	ns	−0.007 (0.002)	−0.005 (0.002)
Sex	−0.147 (0.009)	−0.092 (0.008)	–	–	–	–
Age by sex	−0.027 (0.004)	−0.007 (0.004)	–	–	–	–
PH stage 2	–	–	0.027 (0.011)	ns	0.035 (0.012)	ns
PH stage 3	–	–	0.059 (0.014)	ns	0.044 (0.015)	ns
PH stage 4	–	–	0.088 (0.017)	ns	0.046 (0.018)	ns
PH stage 5	–	–	0.092 (0.023)	ns	0.049 (0.022)	ns
Log_e FFM	–	0.868 (0.025)	–	0.887 (0.032)	–	0.841 (0.040)
Random part						
Level 2						
Variance (cons)	0.006 (0.001)	0.005 (0.000)	0.006 (0.001)	0.004 (0.001)	0.005 (0.001)	0.005 (0.001)
Level 1						
Variance (cons)	0.004 (0.000)	0.004 (0.000)	0.004 (0.000)	0.004 (0.000)	0.003 (0.000)	0.004 (0.000)
Units: level 2	420	420	210	213	206	207
Units: level 1	994	994	479	519	430	475
−2*log-likelihood	−2032.945	−2143.878	−957.949	−1105.574	−925.742	−1029.497

Values are presented as model estimates (SE). Data from Armstrong and Welsman (13).
FFM, fat-free mass estimated from Slaughter et al. (26); $\dot{V}O_2$, oxygen uptake; PH stages, pubic hair stages described by Tanner (56); –, not entered in models; ns, not significant ($P > 0.05$).

the model is a significantly better statistical fit for the data than model 1.1. With FFM controlled, the age, age^2, and age by sex terms remained significant but were reduced in magnitude. With FFM and age controlled, the sex difference in peak $\dot{V}O_2$ fell to ~9%.

The estimation of FFM from body mass and skinfold thicknesses rather than its direct measurement using more sophisticated technology can be criticized, but this methodology is well established in pediatric exercise physiology (26,43,44). Moreover, direct measures of the body fat of 12- to 14-yr-olds have recently been showed to vary widely across laboratory techniques (57). In our original publications (13,14), we demonstrated that simply using BM and the sum of triceps and subscapular skinfold thicknesses as a surrogate for FFM often produced multilevel allometric models with a better statistical fit than those founded on FFM estimated from youth-specific equations. However, on all occasions, the interpretation of FFM in relation to the development of CRF was independent of whether BM and sum of skinfold thicknesses were used as a surrogate of FFM or FFM was estimated from youth-specific equations. For brevity, only the models including estimated FFM (26) are presented herein. However, it should be noted that FFM includes tissues not involved in exercise. Ideally, active contracting muscle mass would be directly determined on each test occasion and introduced to the multilevel allometric models, but this is not currently feasible in pediatric exercise studies.

Sex-specific models 1.3 (boys) and 1.5 (girls) show that with age and BM controlled, the introduction of stage of PH as an indicator of maturity status had additional, significant, and incremental positive effects on peak $\dot{V}O_2$. In girls, the age effect decreased as they got older. This is in direct conflict with the traditional ratio-scaled interpretation where maturity status has been observed to have no additional effect on the development of peak $\dot{V}O_2$ in either boys or girls (58). In models 1.4 (boys) and 1.6 (girls), BM was replaced with FFM. The age effect remained positive, but with age and FFM controlled, additional effects of maturity status were not significant in either sex.

In both sexes, the strongest statistical models were those where FFM replaced BM (i.e., models 1.4 and 1.6). These findings evidence the dominant effect of maturation-driven changes in FFM on the development of peak $\dot{V}O_2$ in both sexes. Peak $\dot{V}O_2$ is increased by enhanced oxygen delivery to or oxygen utilization by the active muscles, or both. Growth and development of active muscle mass, reflected herein by gains in FFM, not only enhance muscle oxygen utilization during exercise but also, through the peripheral muscle pump, augment venous return, increase maximum cardiac output ($\dot{Q}_{max}$), and therefore improve oxygen delivery to the muscles (6,59). From 10 to 16 yr, boys' FFM increases by ~100% and girls' FFM by ~50%. The influence of maturity status on FFM is demonstrated in boys by an ~83% gain over the period 2 yr prepeak height velocity (PHV) to 2 yr post-PHV. The greatest gains in girls' FFM (~31%) occur over a shorter 2-yr period around PHV before plateauing in accord with the development of their peak $\dot{V}O_2$ (60,61).

Cardiovascular covariates. The models presented in Table 1 show the powerful influence of changes in FFM on peak $\dot{V}O_2$, but the development of CRF in youth is also influenced by growth- and maturation-driven changes in maximum cardiopulmonary covariates. It is consistently reported that maximum pulmonary ventilation does not limit the peak $\dot{V}O_2$ of healthy youth (62) so the predominant cardiopulmonary influences on CRF are likely to be maximum cardiovascular covariates of peak $\dot{V}O_2$. The Fick equation describes peak $\dot{V}O_2$ as the product of $\dot{Q}_{max}$ and maximum arteriovenous oxygen difference (a-vO_2diff_{max}), where $\dot{Q}_{max}$ is a function of maximum stroke volume (SV_{max}) and maximum heart rate (HR_{max}).

$\dot{Q}_{max}$ and SV_{max} are traditionally scaled in ratio with body surface area (BSA) as the cardiac or stroke index, but Tanner (28) specifically demonstrated that ratio scaling cardiac data with BSA is fallacious, and it has been compellingly argued

that the most appropriate method of scaling cardiac data is with a curvilinear allometric model (59). $\dot{Q}_{max}$ and SV_{max} are more closely related to metabolic demand than body size and should be considered in relation to active muscle mass rather than BSA. Given the experimental challenges of measuring young people's active muscle mass, FFM has emerged as an appropriate surrogate of active muscle mass with which to scale cardiac data in pediatric exercise science (59).

At the onset of the study, 51 of the children described in Figure 1 agreed to have their peak $\dot{V}O_2$, HR_{max}, and SV_{max} measured on three annual occasions (15,16). In agreement with the extant literature (6,12,59), HR_{max} was not significantly related to age, BM, FFM, or peak $\dot{V}O_2$ and was not explored further. Multilevel allometric models of maximum cardiovascular variables (not presented here but available in references [15] and [16]) showed that with FFM controlled, the introduction of age, maturity status, or blood hemoglobin concentration into the models was not significant. Moreover, with FFM controlled, there were no sex differences in SV_{max}, $\dot{Q}_{max}$, or a-vO_2diff_{max} (15,16).

Figure 3 illustrates the association of peak $\dot{V}O_2$ with HR_{max}, SV_{max}, $\dot{Q}_{max}$, and a-vO_2diff_{max}, respectively. Table 2 presents multilevel allometric models of 10- to 13-yr-olds' peak $\dot{V}O_2$. In model 2.1 with FFM controlled, there was a sex difference in peak $\dot{V}O_2$ of ~5%, which is less than the ~9% sex difference in the 10- to 16-yr-olds (model 1.2). This reflects the increasing sex difference in active muscle mass from 13 to 16 yr (60,61). In models 2.2, 2.3, and 2.4 with

TABLE 2. Multilevel allometric models of the peak oxygen uptake of 11- to 13-yr-olds (cardiovascular covariates).

	Model 2.1	Model 2.2	Model 2.3	Model 2.4
Response	**Log_e Peak $\dot{V}O_2$**			
Fixed part				
Constant	−2.708 (0.133)	−3.052 (0.302)	−1.716 (0.197)	−1.130 (0.302)
Log_e FFM	0.980 (0.037)	–	–	–
Sex	−0.046 (0.019)	ns	ns	ns
Log_e SV_{max}	–	0.862 (0.069)	–	–
Log_e $\dot{Q}_{max}$	–	–	0.876 (0.070)	–
Log_e a-vO_2diff_{max}	–	–	–	0.738 (0.119)
Random part				
Level 2				
Variance (constant)	0.003 (0.001)	0.006 (0.002)	0.006 (0.002)	0.020 (0.005)
Level 1				
Variance (constant)	0.003 (0.001)	0.011 (0.002)	0.011 (0.002)	0.014 (0.003)
Units: level 2	51	51	51	51
Units: level 1	110	110	110	110
−2*log-likelihood	−272.000	−144.898	−143.711	−84.733

Values are presented as model estimates (SE). The introduction of age, maturity status, or blood hemoglobin concentration into any model was not significant ($P > 0.05$). Data collated from Armstrong and Welsman (15,16).
$\dot{V}O_2$, oxygen uptake; SV, stroke volume; $\dot{Q}$, cardiac output; a-vO_2diff, arteriovenous oxygen difference; –, not entered in models; ns, not significant ($P > 0.05$).

SV_{max}, $\dot{Q}_{max}$, or a-vO_2diff_{max} controlled for and replacing FFM, there were no significant sex differences in peak$\dot{V}O_2$, but none of the models founded on a maximum cardiovascular covariate provided as good a statistical fit to the data as model 2.1 founded on FFM (15,16).

In Table 3, model 2.1 is replicated as model 3.1 for comparative purposes only. Models 3.2, 3.3, and 3.4 show that with FFM controlled, SV_{max}, $\dot{Q}_{max}$, or a-vO_2diff_{max}, respectively, are each additional, independent, and significant covariates

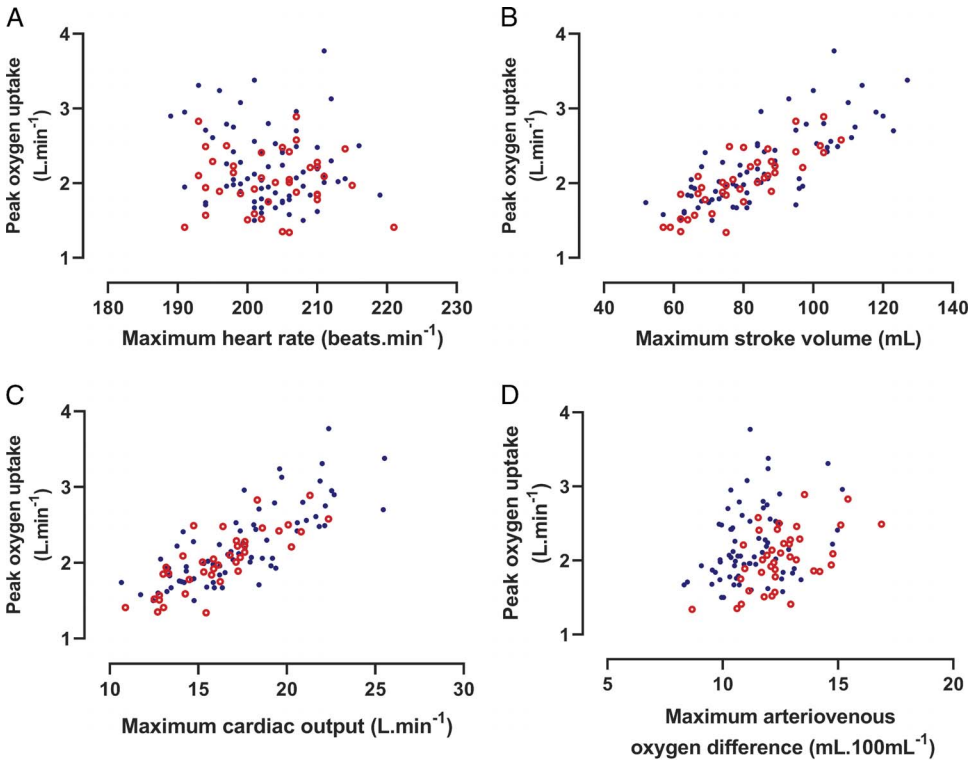

FIGURE 3—Peak oxygen uptake in relation to maximum heart rate, maximum stroke volume, maximum cardiac output, and maximum arteriovenous oxygen difference in 10- to 13-yr-olds. Figure founded on 110 determinations of peak $\dot{V}O_2$, ○ girls (n = 40), ● boys (n = 70). FFM estimated from Slaughter et al. (26). Data collated from Armstrong and Welsman (15,16).

of peak $\dot{V}O_2$. Each of the models including FFM and a maximum cardiovascular covariate provided a significantly better statistical fit to the data than model 3.1. This demonstrates the need to accommodate concurrent changes in multiple covariates to best describe changes in peak $\dot{V}O_2$ in youth. Models 3.2, 3.3, and 3.4 show a small reduction in the unexplained sex difference in peak $\dot{V}O_2$ over the 10- to 13-yr age range; the introduction of age, maturity status, or blood hemoglobin concentration into models was not significant (15,16).

Sex differences. In the data set described in Figure 1, mean sex differences in peak $\dot{V}O_2$ ($L \cdot min^{-1}$) range from ~9% at age 10 yr to ~39% at age 16 yr, but this type of analysis provides few insights into physiological differences between boys and girls growing and maturing at their own individual rate. Similarly, comparing mean ratio-scaled peak $\dot{V}O_2$ data ($mL \cdot kg^{-1} \cdot min^{-1}$) at specific ages confounds true comparisons. With both age and BM appropriately controlled, there is a ~15% sex difference in peak $\dot{V}O_2$ (model 1.1), which reduces to ~9% (model 1.2) with age and FFM controlled. Moreover, with FFM controlled, maturity status does not make a significant additional contribution to explaining changes in peak $\dot{V}O_2$ (models 1.4 and 1.6). It is clear from the models in Table 3 that with FFM controlled, the introduction of SV_{max}, $\dot{Q}_{max}$, or a-vO_2diff_{max} produces models of peak $\dot{V}O_2$ with a better statistical fit but an unexplained sex difference of ~4% in peak $\dot{V}O_2$ remains. Unfortunately, longitudinal data are currently only available for maximum cardiovascular covariates over the age range 10 to 13 yr, and a longer period of study is required to accommodate fully concurrent changes in morphological and maximum cardiovascular covariates due to growth and maturation.

Oxygen-carrying capacity in the blood is dependent on blood hemoglobin concentration, but in the 10- to 13-yr-olds' data set, there were no significant sex differences in blood hemoglobin concentration. Moreover, in all Table 3 models, it was not a significant independent covariate of peak $\dot{V}O_2$. Sex differences in blood hemoglobin concentration become more apparent in the mid-to-late teens (58), but it remains to be proven whether it is a contributory factor to sexual dimorphism in peak $\dot{V}O_2$ in older youth (6). A study using near infrared spectroscopy to estimate microcirculatory changes in deoxygenated hemoglobin and myoglobin reported a poorer matching of muscle oxygen delivery to oxygen utilization in 9- and 10-yr-old girls than in similar age boys, but this remains to be confirmed during adolescence (63).

TABLE 3. Multilevel allometric models of the peak oxygen uptake of 11- to 13-yr-olds (morphological and cardiovascular covariates).

	Model 3.1	Model 3.2	Model 3.3	Model 3.4
Response	Log_e Peak $\dot{V}O_2$			
Fixed part				
Constant	–2.708 (0.133)	–3.034 (0.168)	–2.770 (0.130)	–3.022 (0.146)
Log_e FFM	0.980 (0.037)	0.865 (0.053)	0.869 (0.053)	0.916 (0.038)
Sex	–0.046 (0.019)	–0.041 (0.018)	–0.041 (0.018)	–0.044 (0.019)
Log_e SV_{max}	–	0.164 (0.056)	–	–
Log_e $\dot{Q}_{max}$	–	–	0.161 (0.057)	–
Log_e a-vO_2diff_{max}	–	–	–	0.212 (0.052)
Random Part				
Level 2				
Variance (constant)	0.003 (0.001)	0.002 (0.001)	0.002 (0.001)	0.003 (0.001)
Level 1				
Variance (constant)	0.003 (0.001)	0.003 (0.001)	0.003 (0.001)	0.002 (0.000)
Units: level 2	51	51	51	51
Units: level 1	110	110	110	110
–2*log-likelihood	–272.000	–279.640	–279.275	–287.110

Values are presented as model estimates (SE). The introduction of age, maturity status, or blood hemoglobin concentration into any model was not significant ($P > 0.05$). Data collated from Armstrong and Welsman (15,16).

FFM, fat-free mass estimated from Slaughter et al. (26); $\dot{V}O_2$, oxygen uptake; SV, stroke volume; $\dot{Q}$, cardiac output; a-vO_2diff_{max}, arteriovenous oxygen difference; –, not entered in models; ns, not significant ($P > 0.05$).

It has been suggested that the sex difference might be partially due to gender behavior as boys are generally more physically active than girls (6,64,65). The challenges in estimating and interpreting habitual physical activity (PA) are extensively documented (65–68), but a plethora of cross-sectional studies have consistently found no compelling evidence to indicate a meaningful relationship between objectively monitored habitual PA and rigorously determined peak $\dot{V}O_2$ (for reviews, see [7,64,65,69]). The, very weak, at best, relationship reported between current levels of habitual PA and peak $\dot{V}O_2$ is not surprising as children and adolescents seldom experience the intensity, frequency, and duration of PA necessary to increase their CRF. However, to explore a potential relationship between habitual PA and CRF using longitudinal data, 104 boys and 98 girls, 11 yr old at study onset, had their habitual PA estimated using HR monitoring, over three 12-h periods, on three annual occasions (70). Pilot work walking and running on a horizontal treadmill at a range of speeds established that with 10- to 13-yr-olds, brisk walking (defined as moderate PA) and jogging (defined as vigorous PA) generated HR values of ~140 and ~160 bpm, respectively. With BM and maturity status allometrically controlled, percentage time spent with HR values above 140 and 160 bpm significantly ($P < 0.05$) decreased with age in both sexes. When peak $\dot{V}O_2$ was introduced to the multilevel allometric models, a nonsignificant ($P > 0.05$) parameter estimate was obtained. Collectively, with the cross-sectional data, these data imply that sex differences in the CRF of untrained, healthy youth are unlikely to be due to differences in habitual PA. However, more studies of appropriately analyzed longitudinal data, collected over a longer period, and using more refined measures of PA are required to confirm these findings.

CONCLUSIONS

It is remarkable that ratio scaling of physiological variables has persisted in the pediatric exercise literature despite widespread knowledge that its valid application is founded on a set of specific statistical assumptions, which have been demonstrated empirically to be seldom (if ever) met in children and adolescents (31). In his seminal 1983 book, Oded Bar-Or ([71], p. 4) noted that ratio scaling was "not the method of choice" but still focused on a discussion of ratio-scaled data in relation to age, "as it is the most common way of expressing maximal O_2 uptake for comparative purposes." Almost four decades later, peak $\dot{V}O_2$ ratio-scaled with BM is still being

reported and advocated as reference values for youth health-related CRF (e.g., [72]), cut points for cardiometabolic health (e.g., [39]), and raising clinical red flags (e.g., [40]) in internationally respected journals. For over 80 yr, publications appear to have evaded rigorous scrutiny by peer reviewers and journal editors because ratio scaling is "convenient and traditionally accepted" ([12], p. 7). We know of no other scientific discipline where an assumed relationship with "overwhelming scientific evidence of its many drawbacks" ([73], p. 254) has been widely accepted as an alternative to the application of scientific and statistical rigor. Ratio scaling of peak $\dot{V}O_2$ has been allowed to cloud understanding of developmental exercise physiology and to misguide recommendations designed to promote youth health.

Multilevel allometric modeling of longitudinal data has confirmed the importance of considering sex-specific, concurrent changes in age- and maturation-driven covariates when analyzing the development of peak $\dot{V}O_2$ in youth. In direct conflict with ratio-scaled data, multilevel allometric modeling has demonstrated that in both sexes with BM controlled, peak $\dot{V}O_2$ increases with age but the effect is smaller in girls than in boys. In both girls and boys maturity status exerts an independent, positive effect on peak $\dot{V}O_2$, in addition to those due to changes in age and BM. Age- and maturity status-driven FFM (as a surrogate for active muscle mass) is the most powerful influence on the development of peak $\dot{V}O_2$ in both boys and girls, but the best fit statistical model includes both FFM and a maximum cardiovascular variable. Even with age, maturity status, morphological, and maximum cardiovascular covariates controlled, there remains a residual, unexplained sex difference in peak $\dot{V}O_2$.

FUTURE DIRECTIONS

To elucidate fully the development of CRF requires further experimental exploration of age- and maturation-driven changes in intramuscular factors such as active contracting muscle mass, fiber types, blood volume, blood distribution, mitochondrial density, aerobic enzyme activity, and fiber activation (20,60,74,75). This awaits further development and ethical application of noninvasive technology to developmental exercise physiology. Longitudinal studies incorporating cardiovascular and intramuscular factors are necessary to rigorously monitor developmental changes in peak $\dot{V}O_2$, and multilevel allometric modeling provides an appropriate technique for analyzing these data as they become available.

(ORCID 0000-0002-3086-629X).

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

REFERENCES

1. Ross R, Blair SN, Arena R, et al. Importance of assessing cardiorespiratory fitness in clinical practice: a case for fitness as a clinical vital sign. A scientific statement from the American Heart Association. *Circulation*. 2016;134:e653–99.
2. Armstrong N., and Mechelen Wvan, editors. *Oxford Textbook of Children's Sport and Exercise Medicine*. 3rd ed. Oxford: Oxford University Press; 2017. pp. 225–409.
3. Armstrong N, Welsman J. Youth cardiorespiratory fitness: evidence, myths and misconceptions. *Bull World Health Organ*. 2019;97:777–82.
4. Loftin M, Sothern M, Abe T, Bonis M. Expression of VO_{2peak} in children and youth with special reference to allometric scaling. *Sports Med*. 2016;46:1451–60.
5. Welsman J, Armstrong N. Interpreting cardiorespiratory fitness in young clinical populations: folklore and fallacy. *JAMA Pediatr*. 2019;1485:E1–2.
6. Armstrong N, McManus AM. Aerobic fitness. In Armstrong N, Mechelen Wvan, editors. *Oxford Textbook of Children's Sport and Exercise Medicine*. 3rd ed. Oxford: Oxford University Press; 2017. pp. 161–80.
7. Armstrong N, Tomkinson GR, Ekelund U. Aerobic fitness and its relationship to sport, exercise training and habitual physical activity during youth. *Brit J Sport Med*. 2011;45:849–58.
8. Armstrong N, Welsman J. Aerobic fitness: what are we measuring? *Med Sport Sci*. 2007;50:5–25.
9. American Thoracic Society/American College of Chest Physicians. ATS/ACCP statement on cardiopulmonary exercise testing. *Am J Respir Crit Care Med*. 2003;167:211–77.
10. Mountjoy M, Andersen LB, Armstrong N, et al. International Olympic Committee consensus statement on the health and fitness of young people through physical activity and sport. *Brit J Sport Med*. 2011;45:839–48.
11. Falk B, Klentrou P, Armstrong N, Rowland T, Kemper HCG. A brief history of pediatric exercise physiology. *Pediatr Exerc Sci*. 2018;30:1–10.
12. Bar-Or O, Rowland TW. *Pediatric Exercise Medicine*. Champaign (IL): Human Kinetics; 2004. pp. 3–59.
13. Armstrong N, Welsman J. Sex-specific longitudinal modeling of youth peak oxygen uptake. *Pediatr Exerc Sci*. 2019;31:204–12.
14. Armstrong N, Welsman J. Development of peak oxygen uptake from 11–16 years determined using both treadmill and cycle ergometry. *Eur J Appl Physiol*. 2019;119:801–12.
15. Armstrong N, Welsman J. Multilevel allometric modelling of maximal stroke volume and peak oxygen uptake in 11–13-year-olds. *Eur J Appl Physiol*. 2019;119:2629–39.
16. Armstrong N, Welsman J. Multilevel allometric modelling of maximum cardiac output, maximum arteriovenous oxygen difference, and peak oxygen uptake in 11–13-year olds. *Eur J Appl Physiol*. 2020;120:527–37.
17. Robinson S. Experimental studies of physical fitness in relation to age. *Arbeitsphysiologie*. 1938;10:251–323.
18. Morse M, Schlutz FW, Cassels DE. Relation of age to physiological responses of the older boy to exercise. *J Appl Physiol*. 1949;1:683–709.
19. Armstrong N, Welsman J. Assessment and interpretation of aerobic fitness in children and adolescents. *Exerc Sport Sci Rev*. 1994;22:435–76.
20. Malina RM, Bouchard C, Bar-Or O. *Growth, Maturation, and Physical Activity*. 2nd ed. Champaign (IL): Human Kinetics; 2004. pp. 235–50.
21. Åstrand PO. *Experimental Studies of Physical Working Capacity in Relation to Sex and Age*. Copenhagen: Munksgaard; 1952. pp. 103–23.
22. Bongers BC, van Brussel BC, Hulzebos EHJ, Takken T. *Pediatric Norms for Cardiopulmonary Exercise Testing in Relation to Sex and Age*. 2nd ed. Uitgeverij: Box Press; 2014. pp. 1–198.
23. Falk B, Dotan R. Measurement and interpretation of maximal aerobic power in children. *Pediatr Exerc Sci*. 2019;31:144–51.
24. Krahenbuhl S, Skinner JS, Kohrt WM. Developmental aspects of maximal aerobic power in children. *Exerc Sports Sci Rev*. 1985;13:503–38.

25. Stavnsbo M, Resaland GK, Anderssen SA, et al. Reference values for cardiometabolic risk scores in children and adolescents: suggesting a common standard. *Atherosclerosis*. 2018;278:299–306.
26. Slaughter MH, Lohman TG, Boileau RA, et al. Skinfold equations for estimation of body fatness in children and youth. *Hum Biol*. 1988;60:709–23.
27. Goran M, Fields DA, Hunter GR, Herd SL, Weinster RL. Total body fat does not influence maximal aerobic capacity. *Int J Obes (Lond)*. 2000;24:841–8.
28. Tanner JM. Fallacy of per-weight and per-surface area standards and their relation to spurious correlation. *J Appl Physiol*. 1949;2:1–15.
29. Katch VL. Use of the oxygen/body weight ratio in correlational analyses: spurious correlations and statistical considerations. *Med Sport Sci*. 1973;5:253–7.
30. Katch VL, Katch FI. Use of weight-adjusted oxygen uptake scores that avoid spurious correlations. *Res Quart*. 1974;45:447–51.
31. Welsman J, Armstrong N. Interpreting aerobic fitness in youth: the fallacy of ratio scaling. *Pediatr Exerc Sci*. 2019;32:184–90.
32. Williams JR, Armstrong N, Winter EM, Crichton N. Changes in peak oxygen uptake with age and sexual maturation in boys: physiological fact or statistical anomaly? In Coudert J, Praagh Evan, editors. *Children and Exercise XVI*. Paris: Masson; 1992. pp. 35–7.
33. Welsman J, Armstrong N. Statistical techniques for interpreting body size-related exercise performance during growth. *Pediatr Exerc Sci*. 2000;12:112–27.
34. Welsman JR, Armstrong N. Interpreting performance in relation to body size. In: Armstrong N, editor. *Paediatric Exercise Physiology*. Edinburgh: Elsevier; 2007. pp. 27–46.
35. Welsman JR, Armstrong N. Scaling for size: relevance to understanding the effects of growth on performance. In: Hebestreit H, Bar-Or O, editors. *The Young Athlete*. Oxford: Blackwell Scientific; 2008. pp. 50–62.
36. Welsman JR, Armstrong N, Kirby BJ, Nevill AM, Winter EM. Scaling peak $\dot{V}O_2$ for differences in body size. *Med Sci Sports Exerc*. 1996;28:259–65.
37. Armstrong N, Welsman JR, Kirby BJ. Peak oxygen uptake and maturation in 12-year-olds. *Med Sci Sports Exerc*. 1998;30:165–9.
38. Mintjens S, Menting MD, Daams JG, van Poppel MNM, Roseboom TJ, Gemke RJBJ. Cardiorespiratory fitness in childhood and adolescence affects future cardiovascular risk factors: a systematic review of longitudinal studies. *Sports Med*. 2018;48:2577–605.
39. Aadland E, Anderssen SA, Andersen LB, Resaland GK, Kolle E, Steene-Johannessen J. Aerobic thresholds to define poor metabolic health in children. *Scand J Med Sci Sports*. 2019;29:240–50.
40. Ruiz JR, Cavero-Redondo I, Ortega FB, Welk GJ, Andersen LB, Martinez-Vizcaino V. Cardiorespiratory fitness cut points to avoid cardiovascular disease risk in children and adolescents: what level of fitness should raise a red flag? A systematic review and meta-analysis. *Brit J Sports Med*. 2016;50:1451–8.
41. Kemper HCG, Verschuur R. Maximal aerobic power. *Med Sport Sci*. 1985;20:107–26.
42. Mirwald RL, Bailey DA. *Maximal Aerobic Power*. London (Ontario): Sports Dynamics; 1986. pp. 1–80.
43. Janz KF, Burns TL, Witt JD, Mahoney LT. Longitudinal analysis of scaling $\dot{V}O_2$ for differences in body size during puberty: the Muscatine study. *Med Sci Sports Exerc*. 1998;30:1436–44.
44. Rowland T, Vanderburgh P, Cunningham L. Body size and the growth of maximal aerobic power in children: a longitudinal analysis. *Pediatr Exerc Sci*. 1997;9:262–74.
45. Welsman J, Armstrong N. Children's fitness and health: an epic scandal of poor methodology, inappropriate statistics, questionable editorial practices and a generation of misinformation. *BMJ Evidence-Based Med*. 2019; bmjebm-2019-111232.
46. Armstrong N, Welsman J. Editorial: twenty-metre shuttle run: (mis) representation, (mis)interpretation and (mis)use. *Br J Sports Med*. 2019;53:1199–200.
47. Armstrong N, Welsman J. Myths, misconceptions and the 20 metre shuttle run. *Phys Educ Matters*. 2019;14:36–7.
48. Welsman J, Armstrong N. The 20-metre shuttle run is not a valid test of cardiorespiratory fitness in boys age 11–14 years. *BMJ Open Sports Exerc Med*. 2019;5:e000627.
49. Tomkinson GR, Lang JJ, Blanchard J, Leger L, Tremblay MS. The 20-m shuttle run: assessment and interpretation of data in relation to youth aerobic fitness and health. *Pediatr Exerc Sci*. 2019; 31:152–63.
50. Mayorga-Vega D, Aguila-Solo P, Viciana J. Criterion-related validity of the 20-m shuttle run test for estimating cardiorespiratory fitness: a meta-analysis. *J Sports Sci Med*. 2015;14:536–47.
51. Nevill AM, Holder RL, Baxter-Jones A, Round JM, Jones DA. Modelling developmental changes in strength and aerobic power in children. *J Appl Physiol*. 1998;84:963–70.
52. 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.
53. Armstrong N, Welsman J. Sex-specific longitudinal modelling of short-term power in 11–18-year-olds. *Med Sci Sport Exerc*. 2019; 51:1055–63.
54. Armstrong N, Welsman J, Bloxham SR. Development of 11- to 16-year-olds' short-term power output determined using both treadmill running and cycle ergometry. *Eur J Appl Physiol*. 2019;119: 1565–80.
55. Rasbash J, Steel F, Browne WJ, Goldstein H.*A User's Guide to MLwiN Version 3.02*. University of Bristol Centre for Multilevel Modelling, Bristol: 2018.
56. Tanner JM. *Growth at Adolescence*. 2nd ed. Oxford: Blackwell; 1962. pp. 28–39.
57. Ferri-Morales A, Nascimento-Ferreira MV, Vlachopoulos D, et al. Agreement between standard body composition methods to estimate percentage of body fat in young male athletes. *Pediatr Exerc Sci*. 2018;30:402–10.
58. Armstrong N, Williams JR, Balding J, Gentle P, Kirby BJ. The peak oxygen uptake of British children with reference to age, sex and sexual maturity. *Eur J Appl Physiol*. 1991;62:369–75.
59. Rowland TW. Cardiovascular function. In Armstrong N, Mechelen Wvan, editors. *Oxford Textbook of Children's Sport and Exercise Medicine*. 3rd ed. Oxford: Oxford University Press, 2017. pp. 147–59.
60. Armstrong N. *Development of the Youth Athlete*. Oxford: Routledge; 2019. pp. 1–250.
61. Baxter-Jones ADG, Mirwald RL, McKay HA, Bailey DA. A longitudinal analysis of sex differences in bone mineral accrual in healthy 8–19-year-old boys and girls. *Ann Hum Biol*. 2003;30:160–75.
62. McManus AM, Armstrong N. Pulmonary function. In Armstrong N, Mechelen Wvan, editors. *Oxford Textbook of Children's Sport and Exercise Medicine*. 3rd ed. Oxford: Oxford University Press, 2017. pp. 133–46.
63. McNarry MA, Farr C, Middlebrooke A, et al. Aerobic function and muscle deoxygenation dynamics during ramp exercise in children. *Med Sci Sports Exerc*. 2015;47:1877–84.
64. Rowland TW. *Children's Exercise Physiology*. 2nd ed. Champaign (IL): Human Kinetics; 2005. pp. 1–295.
65. Armstrong N. Paediatric physical activity and aerobic fitness. In: Draper N, Stratton G, editors. *Physical Activity: A Multidisciplinary Approach*. Oxford: Routledge; 2019. pp. 186–204.
66. Armstrong N, Welsman JR. The physical activity patterns of European youth with reference to methods of assessment. *Sports Med*. 2006;36:1067–86.
67. Ekelund U, Tomkinson GR, Armstrong N. What proportion of youth are physically active? Measurement issues, levels and recent time trends. *Brit J Sports Med*. 2011;45:859–66.
68. Hildebrand M, Ekelund U. Assessment of physical activity. In: Armstrong N, van Mechelen W, editors. *Oxford Textbook of Children's Sport and Exercise Medicine*. 3rd ed. Oxford: Oxford University Press; 2017. pp. 303–14.

69. Armstrong N. Top 10 research questions related to youth aerobic fitness. *Res Quart Exerc Sport*. 2017;88:130–48.
70. Armstrong N, Welsman JR, Kirby BJ. Longitudinal changes in 11–13-year-olds' physical activity. *Acta Paediatr*. 2000;89:775–80.
71. Bar-Or O. *Pediatric Sports Medicine for the Practitioner*. New York: Springer-Verlag; 1983. p. 4.
72. Johansson L, Brissman M, Morinder G, Westersthahl M, Marcus C. Reference values and secular trends for cardiorespiratory fitness in children and adolescents with obesity. *Acta Paediatr*. 2020;00:1–7.
73. Blais S, Blanchard J, Dallaire F. An alternative to ratio scaling for the interpretation of aerobic fitness in children. *Pediatr Exerc Sci*. 2019; 31:254–5.
74. Armstrong N, Barker AR, McManus AM. Muscle metabolism during exercise. In Armstrong N, Mechelen Wvan, editors. *Oxford Textbook of Children's Sport and Exercise Medicine*. 3rd ed. Oxford: Oxford University Press, 2017. pp 69–87.
75. Dotan R, Mitchell C, Cohen R, Klentrou P, Gabriel D, Falk B. Child–adult differences in muscle activation—a review. *Pediatr Exerc Sci*. 2012;24:2–21.