The 20 m shuttle run is not a valid test of cardiorespiratory fitness in boys aged 11–14 years

Jo Welsman, Neil Armstrong

ABSTRACT

Objectives The 20 m shuttle run test (20mSRT) is used to estimate cardiorespiratory fitness (CRF) through the prediction of peak oxygen uptake (VO\(_2\)), but its validity as a measure of CRF during childhood and adolescence is questionable. This study examined the validity of the 20mSRT to predict peak VO\(_2\).

Methods Peak VO\(_2\) was measured during treadmill running. Log-linear regression was used to correct peak VO\(_2\) for body mass and sum of skinfolds plus age. Boys completed the 20mSRT under standardised conditions. Maximum speed (km/h) was used with age to predict peak VO\(_2\) using the equation developed by Léger et al. Validity was examined from linear regression methods and limits of agreement (LoA). Relationships between 20mSRT performance and allometrically adjusted peak VO\(_2\), and predicted per cent fat were examined.

Results The sample comprised 76 boys aged 11–14 years. Predicted and measured mass-related peak VO\(_2\) (mL/kg/min) shared common variance of 32%. LoA revealed that measured peak VO\(_2\) ranged from 15% below to 25% above predicted peak VO\(_2\). There were no significant relationships (p>0.05) between predicted peak VO\(_2\) and measured peak VO\(_2\) adjusted for mass, age and skinfold thicknesses. Adjusted for body mass and age, peak VO\(_2\) was not significantly related (p>0.05) to 20mSRT final speed but a weak, statistically significant (r=0.24, p<0.05) relationship was found with peak VO\(_2\) adjusted for mass and fatness. Predicted per cent fat was negatively correlated with 20mSRT speed (r=-0.61, p<0.001).

Conclusions The 20mSRT reflects fatness rather than CRF and has poor validity grounded in its flawed estimation and interpretation of peak VO\(_2\), in mL/kg/min.

INTRODUCTION

Cardiorespiratory fitness (CRF) reflects the body’s integrated ability to deliver oxygen from the atmosphere to the skeletal muscles and to consume it to provide energy to support muscular activity during exercise. The health-related benefits of CRF are widely recognised, and cardiorespiratory exercise tests are an established component of paediatric exercise medicine and the clinical assessment of peak VO\(_2\) is a routine element of disease evaluation and intervention monitoring. According to a Scientific Statement from the American Heart Association, CRF can be considered "a reflection of total body health" (Ross, p. e654). However, the use of CRF in clinical practice and health-related recommendations with children and adolescents must be founded on its rigorous assessment and interpretation.

Laboratory determined peak oxygen uptake (VO\(_2\)) is the ‘gold standard’ measure of young people’s CRF and its assessment, interpretation and development are extensively documented. However, rigorous determination of the peak VO\(_2\) of large samples of young people requires technical expertise, interpersonal skills, sophisticated apparatus and significant laboratory resources. These extensive requirements appear to have stimulated a resurgence of interest in estimating predicting young people’s peak VO\(_2\) from performance on the 20 m shuttle run test (20mSRT).

Over 30 years ago, we investigated the 20mSRT as a potential means for assessing
CRF. On finding a common variance (r²) of 29% between 20mSRT performance and laboratory-determined peak VO₂ in pre-adolescent and early adolescent boys, we concluded that the 20mSRT was not a valid surrogate for rigorously measured peak VO₂. We assumed that with the development of new technologies the use of performance tests in scientific and medical research would gradually disappear. On the contrary, judging by the plethora of publications in the last few years the 20mSRT appears to have become the method of choice in determining youth CRF in scientific, medical and health-related research.

Prominent advocates of the 20mSRT recommend the equation developed by Léger et al for those aged 8–19 years as the basis for predicting CRF. Using this equation, Léger et al reported predicted peak VO₂ to have a correlation of r=0.71 with peak VO₂ predicted from retro-extrapolation to VO₂ measured at the end of the test. Reported SE of the estimate (SEE) was 5.9mL/kg/min or 12.1%. Thus, although the equation developed by Léger et al was not originally tested against directly measured peak VO₂, it has been used to predict peak VO₂ from 20mSRT performance scores which have been used to construct international CRF ‘norms’, examine intercountry comparisons and to generate temporal trends in CRF. The protocol has even been modified to develop ‘reference standards’ for children as young as 2 years of age. Moreover, 20mSRT performance has been recommended to evaluate physical activity interventions, to survey and monitor international health and fitness, to determine metabolic and cardiovascular risk and to identify individuals who warrant intervention to improve their current and future health.

Proponents of the 20mSRT claim that there has been, ‘a substantial decline in CRF since 1981, which is suggestive of a meaningful decline in population health’ (Tomkinson, p. 4). As well-documented and resolved in the IOC Consensus Statement on health and fitness of young people, there is no compelling evidence to suggest that youth CRF has declined over this period. In explanation of the alleged decline in CRF, 20mSRT advocates have asserted that, ‘direct analysis of the causal fitness-fatness connection indicates that increases in fatness explain 35%–70% of the declines in CRF’ (Tomkinson, p. 5). But fat is largely metabolically inert and there is no ‘causal fitness-fatness connection’. Being fat is not the same as being unfit. These contrasting findings do, however, clearly illustrate the difference between true CRF and the willingness and capability of individuals to carry their body mass between two lines 20 m apart and maintain the required running speed. Fat mass does not influence CRF but carrying fat as ‘dead-weight’ does increase the work performed in each 20 m shuttle and adversely affect 20mSRT performance.

This misinterpretation of data is compounded further by 20mSRT equations predicting peak VO₂ expressed in simple ratio with body mass, that is, as mL/kg/min. This is a fundamental methodological flaw which was elegantly explained by Tanner 70 years ago and confirmed by ourselves in children in 1992. Given these concerns, the present paper aimed to i) empirically investigate 20mSRT predicted peak VO₂ as a valid estimate of laboratory-determined peak VO₂ expressed as a simple per body mass ratio and appropriately adjusted for age, body size and fatness and ii) examine the influence of body fatness on shuttle run performance expressed as maximal speed.

METHODS
We have re-visited our raw data, calculated predictions of peak VO₂ from 20mSRT performance using currently recommended equations and methodology and investigated 20mSRT estimations of peak VO₂ in relation to treadmill-determined peak VO₂ and a range of appropriately scaled morphological covariates.

Participants
Seventy-six boys aged 11–14 years from a local state school volunteered to participate in a project investigating CRF and cardiovascular health. All boys and their guardians provided written informed consent. Participants were habituated to the laboratory environment, personnel and experimental procedures prior to data collection.

Experimental procedures
Anthropometry
Maturity status was visually assessed by a medical doctor and recorded by a nurse based on indices for pubic hair development. Anthropometric measures were taken as described by the International Biological Programme. Stature was measured with a Harpenden Stadiometer and body mass determined using Avery balance scales (Avery, Birmingham, UK). Skinfold thicknesses at triceps and subcapular sites were measured by the same experienced investigator using Holtain skinfold callipers (Holtain, Crosswell, UK). Per cent fat was predicted from maturity status, body mass and skinfold thicknesses using sex-specific equations.

Peak oxygen uptake
Peak VO₂ was determined on a motorised treadmill (Woodway, Cranlea Medical, Birmingham, UK) using a discontinuous, incremental protocol. Tests began with a 5 min warm-up at a belt speed of 8 km.h⁻¹ (2.22 m.s⁻¹). The belt speed was increased to 10 km.h⁻¹ (2.78 m.s⁻¹) and the treadmill gradient was raised by 2.0% every 3 min interspersed with 1 min rest. Boys continued running until voluntary exhaustion. Heart rate and expired respiratory gases were monitored continuously using an electrocardiograph (Cardionics, Stockholm, Sweden) and an online respiratory gas-analysis system (Covox, Exeter, UK), which was calibrated against reference gases and an appropriate range of flow rates using a Hans Rudolph calibration syringe (Cranlea, Birmingham, UK) before each test. The highest 30 s VO₂ was accepted as a maximal value if signs of intense exertion (eg, hyperpnea, facial flushing unsteady gait, profuse sweating) were demonstrated and supported...
by a heart rate which was levelling-off over the final stages of the test at a value within 5% of previously measured mean maximal heart rates of boys aged 11–14 years.

20 m shuttle run
The 20mSRT was performed in a school gymnasium. The boys ran in small groups back and forth between two lines at a measured distance of 20 m apart. Following a brief warm-up, the test started at a speed of 8.5 km.h⁻¹ and increased by 0.5 km.h⁻¹ every minute in accordance with audio signals emitted at set intervals from a prerecorded tape. Each boy continued until he could no longer maintain the pace. The final speed was noted and, with age, converted into predicted peak VO₂ using the equation developed by Léger et al.¹⁶:

\[ Y = 31.025 + 3.238X_1 - 3.248X_2 + 0.1536X_1X_2 \]

where \( Y \) is peak VO₂ (mL/kg/min); \( X_1 \) is maximal shuttle running speed (km.h⁻¹) and \( X_2 \) is age (years).

Data analyses
Data were analysed using SPSS V.24 software (IBM, Armonk, New York, USA). Descriptive data (means and SD) were computed for the boys’ physical characteristics, maximal performance in the 20mSRT and for absolute (L.min⁻¹) and mass-related (mL/kg/min) treadmill-determined peak VO₂. Significance was accepted as \( p\leq0.05 \).

Validity of 20mSRT using recommended per body mass comparisons
The strength of the linear relationship between predicted and measured peak VO₂, expressed in mL/kg/min, was investigated using Pearson’s correlation. Subsequently, Bland and Altman’s limits of agreement (LoA) between predicted peak VO₂ and treadmill determined peak VO₂, expressed in mL/kg/min, were computed using GraphPad Prism (GraphPad Software, San Diego, California, USA).

Allometric relationships between treadmill peak VO₂ and anthropometric measures were investigated using log-linear regression. Thirty-one. Initially, body mass was included as a sole covariate. In subsequent analyses, age was added with body mass and, finally, sum of two skinfold thicknesses. The log-linear regression equations obtained were used to calculate log peak VO₂ adjusted for these covariates as appropriate. The antilog of these values provided adjusted peak VO₂ expressed in L.min⁻¹. Pearson’s correlation coefficients were computed between 20mSRT performance (predicted peak VO₂ (mL/kg/min) and maximum speed attained (km.h⁻¹)) versus peak VO₂ allometrically adjusted for body mass, age and sum of skinfolds. The relationship between predicted fat percentage and 20mSRT performance (maximal speed) was examined using Pearson’s correlation.

RESULTS
Descriptive data for the boys’ physical characteristics are presented in table 1. Table 2 summarises the 20mSRT results expressed in maximum speed attained (km.h⁻¹) and predicted mass-related peak VO₂ (mL/kg/min) and contains a comparative value for treadmill determined mass-related peak VO₂.

Figure 1A illustrates the linear relationship between 20mSRT predicted and treadmill-determined values of mass-related peak VO₂. The Pearson’s correlation analysis demonstrated a shared variance (\( r^2 \)) of 32%. Initial graphical analysis (figure 2A) of the differences between predicted and measured peak VO₂ versus their average yielded a bias of 1.4 mL/kg/min with 95% LoA of -9.1 to +11.9 mL/kg/min. The differences were normally distributed (Shapiro-Wilk test, \( p>0.05 \)), but the size of the difference was significantly related to the average of the two peak VO₂ measurements (Spearman’s rank correlation \( r=0.34, p<0.01 \)). Therefore, the data were expressed as the ratio between the two scores versus the average score¹⁰ (figure 2B). This analysis revealed minimal bias (mean ratio 1.03) but measured peak VO₂ ranged from 15% below (ratio 0.85) to 25% above (ratio 1.25) 20mSRT predicted peak VO₂.

<table>
<thead>
<tr>
<th>Table 1 Physical characteristics</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>12.7</td>
<td>1.0</td>
</tr>
<tr>
<td>Stature (m)</td>
<td>1.54</td>
<td>0.09</td>
</tr>
<tr>
<td>Body mass (kg)</td>
<td>43.4</td>
<td>9.1</td>
</tr>
<tr>
<td>Sum of triceps and subscapular skinfolds (mm)</td>
<td>18.5</td>
<td>7.3</td>
</tr>
<tr>
<td>Predicted fat* (%)</td>
<td>16.1</td>
<td>6.1</td>
</tr>
</tbody>
</table>

*Predicted from the equations of Slaughter et al.²⁹

<table>
<thead>
<tr>
<th>Table 2 20 m shuttle run and treadmill-determined exercise data</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>20 m shuttle run data</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum 20 m shuttle run speed (km.h⁻¹)</td>
<td>10.9</td>
<td>0.9</td>
</tr>
<tr>
<td>Predicted peak oxygen uptake* (mL/kg/min)</td>
<td>47</td>
<td>5</td>
</tr>
</tbody>
</table>

**Treadmill-determined data**

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peak oxygen uptake (L.min⁻¹)</td>
<td>2.10</td>
<td>0.47</td>
</tr>
<tr>
<td>Peak oxygen uptake (mL/kg/min)</td>
<td>48</td>
<td>6</td>
</tr>
<tr>
<td>Peak oxygen uptake adjusted† for body mass (L.min⁻¹)</td>
<td>2.08</td>
<td>0.39</td>
</tr>
<tr>
<td>Peak oxygen uptake adjusted† for body mass and age (L.min⁻¹)</td>
<td>2.09</td>
<td>0.40</td>
</tr>
<tr>
<td>Peak oxygen uptake adjusted† for body mass and skinfolds (L.min⁻¹)</td>
<td>2.08</td>
<td>0.43</td>
</tr>
<tr>
<td>Peak heart rate (beats.min⁻¹)</td>
<td>201</td>
<td>8</td>
</tr>
</tbody>
</table>

†Values adjusted for the specified covariates in log-linear regression analyses.

*Predicted using the equation developed by Léger et al.¹⁶
Investigating the log-linear relationships between peak VO$_2$ and body mass, age and sum of skinfolds yielded the following results: with log$_e$ body mass as the sole covariate: constant: -2.61; mass exponent: 0.89 (SE 0.08); for log$_e$ body mass and age: constant: -2.74; mass exponent: 0.74 (SE, 0.09), coefficient for age: 0.05 (SE, 0.02); with log$_e$ sum of skinfolds added as a covariate, the age term became redundant (p>0.05): constant: -2.55; mass exponent: 1.10 (SE 0.08); skinfolds exponent: -0.298 (SE, 0.05). Values for peak VO$_2$ adjusted according to these three equations and expressed in L.min$^{-1}$ (ie, the antilog of the predicted values) are presented in table 2. Relationships between allometrically adjusted peak VO$_2$ and 20mSRT predicted peak VO$_2$ are shown in figure 1B–D. These graphs show that once CRF was appropriately adjusted there was no significant (p>0.05) residual correlation between actual and 20mSRT predicted peak VO$_2$.

Relationships between mass-related and allometrically adjusted peak VO$_2$ and 20mSRT maximum speed are illustrated graphically in figure 3. Peak VO$_2$ in mL/kg/min was significantly (p<0.01) related to maximum 20mSRT speed with a common variance of 34%, but once scaled for body mass or body mass plus age, this relationship was non-significant (p>0.05). With peak VO$_2$ adjusted for body mass and skinfolds, a weak, significant (p<0.05) relationship with 20mSRT speed was observed. Further investigation of this revealed an SEE of 0.42L.min$^{-1}$ (95% CI 0.36 to 0.50 (L.min$^{-1}$) representing a coefficient of variation of 21.5%).

Finally, the statistically significant, negative (p<0.001) relationship between predicted per cent fat and 20mSRT performance (maximum running speed) is shown in figure 4.

**DISCUSSION**

**Predicted versus measured peak VO$_2$: correlation and linear regression**

This study examined the validity of the 20mSRT to predict laboratory-determined peak VO$_2$ in boys aged 11–14 years. When, for comparative purposes only, data were analysed as recommended by proponents of the test, that is, as predicted versus laboratory measured peak VO$_2$.
expressed in ratio with body mass, the moderate correlation coefficient obtained indicated a common variance of 32%. More appropriate statistics investigating agreement between the two scores, yielded LoA demonstrating that 95% of the time, directly measured peak VO$_2$ ranged from 15% below to 25% above 20mSRT predicted peak VO$_2$.

Most studies assess validity based on Pearson’s correlation. This is well-recognised as a poor indicator of agreement between two measurement methods being based on linear association and sensitive to sample heterogeneity. Data from the present study are, however, in line with a meta-analysis of 20mSRT validity, which noted that over half of published validity studies with children report a common variance of $<50\%$ between predicted and measured peak VO$_2$. Some studies with children and adolescents have reported validity based on SEE of the linear regression between predicted and measured peak VO$_2$. These studies were recently summarised by Tomkinson et al. into an average SEE of 4.9 mL/kg/min. These authors also reported the ‘$95\%$ likely range for a true peak VO$_2$ predicted from the 20mSRT to be around $10$ mL/kg/min or $24\%$’ (Tomkinson, p. 154). However, these analyses are similarly based on linear association and so are subject to the same limitations as for correlational analyses.

**Predicted versus measured peak VO$_2$: limits of agreement**

Only one previous study appears to have examined agreement between measured and predicted peak VO$_2$ using LoA: Matsuzuka et al. compared measurements from a sample of participants aged 8–17 years. Based on the equation developed by Léger et al., LoA of $-9.8$ to $+6.4$ mL/kg/min were reported, a range representing $\sim 32\%$ of measured peak VO$_2$. In the present sample of boys, our LoA are slightly larger, approximating 40% of measured peak VO$_2$. Neither study presents convincing evidence for the validity of the 20mSRT to either reflect young peoples’ peak VO$_2$ or assign individuals into high, medium or low CRF groups. This is particularly important given current trends towards classifying individuals by single 20mSRT-predicted peak VO$_2$ ‘cut-off’ points as warranting targeted interventions for cardiovascular prevention. Moreover, the use of a single value peak VO$_2$ ‘cut-off’ point, typically 42 mL/kg/min for boys aged 8–18 years, has no scientific basis as CRF increases with age-driven and maturity status-driven concurrent changes in anthropometric and physiological variables with the timing and tempo of these changes specific to individuals.

**Validity of peak VO$_2$, expressed as per body mass ratios**

Concerns over the validity of the 20mSRT to predict CRF should also focus on its overriding weakness: its reliance on equations that only predict peak VO$_2$ expressed in ratio with body mass. There is now unequivocal evidence to refute this contention including the data from the present study which, once differences in age were controlled for, reported a mass exponent of 0.74 (SE 0.09)—a value significantly different from the exponent of 1.0 assumed by ratio scaling.

We have recently reanalysed treadmill peak VO$_2$ from 20 published cross-sectional data sets collected, using rigorous and consistent methodology, in our laboratory over a period of 30 years. In none of these cross-sectional data sets, based on males and females aged 9–18 years, was simple ratio scaling adequate to normalise data. Similarly, longitudinal data based on 1000 individual treadmill-determined peak VO$_2$ tests confirmed that i) if peak VO$_2$ is allometrically scaled with body mass as the sole covariate, the value of the mass exponent is significantly lower than 1.0, ii) that ratio scaling with body mass is too simplistic to describe peak VO$_2$ from prepuberty through to postpuberty and iii) that age-driven and maturity status-driven changes in fat-free mass (reflected here by body mass and skinfold thicknesses) underpin adolescent growth in peak VO$_2$. Thus, based on our own recently analysed data comprising over 2000 laboratory determinations of peak VO$_2$ plus data from systematic reviews, we believe that there is overwhelming evidence to refute the validity of any CRF test based on ratio scaling. Moreover, specific verification of this in the present data can be seen in figure 1B–1D, which graphically and statistically confirm that no significant relationship existed between measured and predicted 20mSRT CRF once the laboratory data were appropriately scaled for age and anthropometric covariates.

**Relationships between allometrically scaled peak VO$_2$ and 20mSRT maximum speed**

Although the vast majority of published studies of 20mSRT performance compute and report mass-related peak VO$_2$, good practice recommendations from advocates of the test include reporting the number of stages or maximal speed attained, with the latter suggested as the only ‘unequivocal metric’. Having shown 20mSRT predicted peak VO$_2$ to lack validity as well as being grounded in inappropriate normalisation for body size, we explored relationships between allometrically scaled peak VO$_2$ and 20mSRT maximum speed. For illustrative purposes, figure 3A reveals a significant relationship between mass-related peak VO$_2$ and maximum speed: a spurious correlation reflecting inappropriate per body mass scaling. However, once adjusted for body mass using allometric methods (figure 3B) or mass and age (figure 3C), this relationship became non-significant. Interestingly, when adjusted for body mass and skinfolds, as demonstrated to best reflect adolescent changes in CRF, a weak but significant correlation was observed. However, a common variance of 6% and a coefficient of variation of 21.5% does not represent convincing support for the use of 20mSRT maximal speed to reflect CRF in boys aged 11–14 years.
Influence of overweight and 20mSRT performance

Where CRF is expressed in mL/kg/min, it is unsurprising that children and adolescents who are overweight or have obesity score poorly on the 20mSRT as they are doubly penalised: first through increased individual workload from carrying excess fat mass during shuttle running and second via the use of this low score to predict mass-related CRF. Here, they are penalised a second time as the denominator includes both fat-free mass (which contributes to CRF) and fat mass (which does not). As we noted earlier, there is no causal relationship between fatness and CRF. Indeed, in our recent longitudinal analysis of CRF in boys aged 10–18 years, we demonstrated empirically that once body mass and skinfold thicknesses were appropriately accounted for in a log-linear multilevel model, the difference in CRF between boys considered overweight and boys in the normal weight range for their age and stature was removed.

Data from the present study were used to provide explicit insights into the effect of excess body fat on 20mSRT performance: despite this being a sample of normal, healthy teenage boys with only five (6%) classified as overweight according to international cutoff points, figure 4 clearly illustrates the decline in 20mSRT maximal speed with increased fatness—a relationship which is likely to be strengthened in groups of individuals with higher levels of adiposity. Using 20mSRT performance as a measure of CRF in overweight young people may therefore lead to spurious relationships with indicators of health where associations with, for example, cardiovascular risk factors are more likely to reflect overweight or obese status than true CRF.

CONCLUSIONS

In this study, we have presented empirical evidence to show that the 20mSRT is not a valid predictor of CRF in boys aged 11–14 years. International recommendations for its use are therefore unjustified, misleading and potentially undermining a generation of research into current and future CRF and health. We encourage researchers and medical practitioners involved with health promotion in young people to abandon its use, and to turn to more rigorous, scientific health promotion in young people to abandon its use, and to turn to more rigorous, scientific health promotion in young people. Researchers and medical practitioners involved with health promotion in young people to abandon its use, and to turn to more rigorous, scientific health promotion in young people. Researchers and medical practitioners involved with health promotion in young people to abandon its use, researchers and medical practitioners involved with health promotion in young people.

Twitter Jo Welsman @Jo Welsman

Acknowledgements The authors would like to acknowledge the commitment of the participants and the technical assistance of the Children’s Health and Exercise Research Centre team.

Contributors JW and NA jointly conceived and designed the research, led the research team and analysed the data. Both authors contributed to the drafting of the manuscript, both authors reviewed and approved the final version.

Funding This study was funded by Northcott Devon Medical Foundation and the Darlington Trust.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval Ethical approval for this study was received from the Exeter and District Health Authority Ethics Committee.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. Contact: n.armstrong@exeter.ac.uk

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iD

Jo Welsman http://orcid.org/0000-0002-8877-3926

REFERENCES

21 Ruiz JR, Cavero-Redondo I, Ortega FB, et al. Cardiorespiratory fitness cut points to avoid cardiovascular disease risk in children and adolescents: what level of fitness should raise a red flag?


