3-2-2017

# Review of criterion-referenced standards for cardiorespiratory fitness: what percentage of 1 142 026 international children and youth are apparently healthy?

Justin J. Lang

Mark S. Tremblay

Francisco B. Ortega

Jonatan R. Ruiz

Grant Tomkinson
*University of North Dakota*, grant.tomkinson@und.edu

# Review of criterion-referenced standards for cardiorespiratory fitness: what percentage of 1 142 026 international children and youth are apparently healthy?

Justin J Lang, Mark S Tremblay, Francisco B Ortega, Jonatan R Ruiz, Grant R Tomkinson

## Abstract

**Purpose** To identify criterion-referenced standards for cardiorespiratory fitness (CRF); to estimate the percentage of children and youth that met each standard; and to discuss strategies to help improve the utility of criterion-referenced standards for population health research.

**Methods** A search of four databases was undertaken to identify papers that reported criterion-referenced CRF standards for children and youth generated using the receiver operating characteristic curve technique. A pseudo-dataset representing the 20-m shuttle run test performance of 1 142 026 children and youth aged 9–17 years from 50 countries was generated using Monte Carlo simulation. Pseudo-data were used to estimate the international percentage of children and youth that met published criterion-referenced standards for CRF.

**Results** Ten studies reported criterion-referenced standards for healthy CRF in children and youth. The mean percentage (±95% CI) of children and youth that met the standards varied substantially across age groups from 36%±13% to 95%±4% among girls, and from 51%±7% to 96%±16% among boys. There was an age gradient across all criterion-referenced standards where younger children were more likely to meet the standards compared with older children, regardless of sex. Within age groups, mean percentages were more precise (smaller CI) for younger girls and older boys.

**Conclusion** There are several CRF criterion-referenced standards for children and youth producing widely varying results. This study encourages using the interim international criterion-referenced standards of 35 and 42 mL/kg/min for girls and boys, respectively, to identify children and youth at risk of poor health—raising a clinical red flag.

## Introduction

Cardiorespiratory fitness (CRF) is often measured as peak oxygen uptake ($\dot{V}O_{2peak}$) in children and youth, which is described as the highest rate of consumed oxygen during a bout of exercise.[1] CRF in children and youth is a good determinant of sporting success[1] and is meaningfully associated with health.[2] In children and youth, CRF is moderately to strongly associated with cardiometabolic,[3 4] mental[5] and physical[6] health markers. Being a trait, CRF levels in adolescence track moderately well into adulthood.[7] Furthermore, recent cohort studies of over 700 000 Swedish adolescents with a median follow-up of 29–34 years have identified that individuals with low CRF in late adolescence have an increased risk of myocardial infarction[8] and early mortality[9] in adulthood. Together, these findings suggest that the measurement and surveillance of CRF in childhood and adolescence can help describe the current and future health of populations.

The 20-m shuttle run test (20mSRT), originally published in 1984 by Léger and colleagues,[10] is the most widely reported field-based assessment of CRF with published normative data available on over 1.1 million children from 50 countries.[11 12] This test displays moderate-to-high criterion related validity[13] and high reliability,[14] suggesting that the 20mSRT is a good field-based measure of $\dot{V}O_{2peak}$. It is also considered an excellent population health surveillance tool because of its simplicity, low cost, flexibility (eg, indoor/outdoor), minimal equipment and ability to test large groups of children and youth simultaneously.[14] As a result, the 20mSRT is included as part of several national and international physical fitness test batteries, such as the Eurofit,[15] FITNESSGRAM,[16] Canadian Assessment of Physical Literacy,[17] Australian Council for Health, Physical Education and Recreation,[18] Assessing Levels of Physical Activity and Fitness,[19] and Assessing FITness in PREschoolers.[20–22]

There are two distinct approaches to interpret 20mSRT performance results. Individual results can be compared with a normative-referenced standard—described as a relative comparison with the performance of peers, and expressed as a percentile rank. Normative-referenced standards for the 20mSRT exist for international[23] and various national populations.[24–29] Alternatively, results can be converted to predicted $\dot{V}O_{2peak}$ using one of several equations[23] to compare with a criterion-referenced standard—described as an absolute measure of the minimum performance required to ‘achieve’ the criterion-referenced (eg, healthy metabolic or cardiovascular profile) standard. These data are often expressed as the percentage (prevalence) of the sample meeting the criterion-referenced standard. Several of these standards have recently emerged applying a receiver operating characteristic (ROC) curve analysis—a diagnostic evaluation that identifies the optimal standards to distinguish a health outcome in a population.[30] The recent emergence of multiple criterion-referenced standards may potentially create confusion regarding the healthy CRF status of children and youth. Furthermore, international percentages of children and youth meeting these criterion-referenced standards do not exist. These referenced percentages would provide utility for international surveillance, screening and CRF comparisons across populations.

Therefore, the purpose of this study was threefold: (a) to identify all published ROC curve-generated criterion-referenced standards for CRF in children and youth; (b) to estimate the percentage of children and youth that met each criterion-referenced standard using an international sample of 20mSRT performance scores from 1 142 026 children and youth aged 9–17 years; and (c) to discuss strategies to improve the utility of CRF criterion-referenced standards for population health research.

## Methods

### Search strategy

A literature search was conducted to identify all studies that reported health-related criterion-referenced standards for CRF in children and youth. A single author (JL) searched MEDLINE, PsycINFO, EMBASE and SPORTDiscus (1980–2016) using a combination of three separate search strategies: (a) child* OR adoles* OR paediatric* OR schoolchild; (b) cardiorespiratory* OR cardiovascular* OR aerobic* OR $\dot{V}O_2$* OR criterion*; and (c) cardiovascular OR metabolic OR insulin OR blood pressure OR cardiometabolic OR psychosocial OR well-being OR self-efficacy. All search strategies were combined using the Boolean operator ‘AND.’ No restrictions were placed on language. Furthermore, the reference lists of all relevant studies were reviewed for additional studies not included in the original search.

### Inclusion criteria

One author (JL) assessed titles and abstracts for those likely to report criterion-referenced standards for children and youth. Studies were included if they reported: (a) CRF standards ($\dot{V}O_{2peak/max}$, mL/kg/min) associated with cardiovascular, metabolic and/or psychosocial health markers; (b) standards generated using ROC curve statistics; and (c) standards generated for children and youth. No restrictions were placed on study design (eg, cohort, cross-sectional).

### Data extraction

One author (JL) independently reviewed all the included studies and extracted the following information: authors, year of publication, country, health criterion, CRF assessment, $\dot{V}O_{2peak}$ prediction equation, sex, age (years), sample size, criterion-referenced standards (mL/kg/min), area under the ROC curve and 95% CI. A second author (GT) reviewed the data for transcription errors.

**International sample and percentage analysis**

To achieve the second objective of this study, international normative values for the 20mSRT performance were obtained from Tomkinson and colleagues (table 1).[23] In brief, the international norms were developed using a systematic review protocol to identify 177 studies that reported 20mSRT performance scores for apparently healthy (eg, free from known disease) children and youth aged between 9 and 17 years. A total of 1 142 026 20mSRT performances were extracted and standardised to running speed (km/hour) at the last completed stage using the data treatment procedure reported by Tomkinson et al.[23] Sample-weighted means and SD at the sex by age by country level were corrected for oversampling and/or undersampling using a poststratification population-weighted procedure.[31] The corrected means and SD at the sex by age level were imported into LMSchartmaker[32] to generate the international 20mSRT normative-referenced values using the Lambda Mu Sigma (LMS) method.

Table 1
International 20mSRT norms for predicted $\dot{V}O_{2peak}$ (mL/kg/min) obtained by converting Tomkinson *et al's* [23] normative centile values using the Léger equation[34]

| | Girls | | | Boys | | |
|---|---|---|---|---|---|---|
| Age (years) | n | 50th percentile | SD | n | 50th percentile | SD |
| 9 | 98 166 | 46.7 | 3.2 | 101 532 | 48.1 | 4.0 |
| 10 | 81 415 | 45.1 | 3.7 | 81 538 | 46.9 | 4.3 |
| 11 | 78 029 | 43.5 | 4.1 | 79 660 | 45.8 | 5.0 |
| 12 | 69 497 | 42.0 | 4.4 | 70 266 | 45.2 | 5.7 |
| 13 | 62 491 | 40.4 | 4.7 | 72 913 | 45.0 | 6.2 |
| 14 | 55 035 | 38.8 | 5.0 | 65 317 | 44.6 | 6.8 |
| 15 | 43 730 | 37.2 | 5.3 | 53 287 | 44.0 | 7.3 |
| 16 | 34 978 | 35.6 | 5.4 | 35 490 | 43.2 | 7.7 |
| 17 | 29 559 | 34.1 | 5.6 | 29 123 | 42.6 | 8.2 |
| Total | 552 900 | | | 589 126 | | |

- 20mSRT, 20-m shuttle run test.

International normative running speed values were used to randomly generate pseudo-data points for each sex by age group using Monte Carlo simulation, a technique described in detail elsewhere.[12 33] This technique assumes that data are normally distributed, which is true for 20mSRT running speed at the last completed stage. Pseudo-data were repeatedly generated until the calculated mean differed from the reported mean by less than 0.5%, and the calculated SD differed from the reported SD by less than 2.5%. These international running speed pseudo-data were then converted to $\dot{V}O_{2peak}$ using the Léger prediction equation:

$$\dot{V}O_{2peak}\ (\text{mL/kg/min}) = 31.025 + 3.238 \bullet X_1 - 3.248 \bullet X_2 + 0.1536 \bullet X_1 \bullet X_2$$

where $X_1$ is the running speed at the last completed stage (km/hour) and $X_2$ is the age at the last birthday.[34]

The Léger equation was applied to help compare outcomes across different criterion-referenced standards. This equation was developed using the 20mSRT, a wide range of age groups and two inputs that were readily available and similar to the pseudo-data used in this study.[34] In addition, the Léger equation is the most widely used equation to predict $\dot{V}O_{2peak}$ in children and youth.

The international 20mSRT pseudo-dataset was then applied to each reported CRF criterion-referenced standard:

$$\text{HealthyCRF}(\%) = \frac{\text{number of participants that attained the criterion} - \text{referenced standard}}{\text{totalnumberofparticipants}}$$

In addition, these sex by age by study percentages were used to calculate the overall smoothed population-weighted mean percentages and 95% CIs using the poststratification population-weighted procedure described by Levy and Lemeshow.[31] These percentages represent the gross estimate of the international percentage of children and youth with apparently healthy CRF.

## Results

### Criterion-referenced standards

Ten studies reported CRF criterion-referenced standards for children and youth between 8 and 19 years of age. These studies were published over a 10-year period from 2006 to 2016 (table 2). The criterion-referenced standards for girls and boys ranged from 32.6 to 43.9 mL/kg/min and from 40.0 to 47.0 mL/kg/min, respectively. Standards were developed using European, North American or South American samples of children and youth. The health criteria were largely based on summative markers of metabolic and/or cardiovascular health (eg, triglycerides, cholesterol, glucose, body composition, etc). In general, each study estimated $\dot{V}O_{2peak}$ using a combination of various CRF assessment protocols, exercise modes (eg, walk/run field tests, graded treadmill test, cycle ergometry) and $\dot{V}O_{2peak}$ prediction equations. Only one study included a direct measure of $\dot{V}O_{2peak}$ for a portion of the study participants.[35] Five studies assessed CRF using the 20mSRT, representing half of the included studies.

Table 2
Descriptive characteristics of included studies

| Study | Year | Country | Criterion | CRF test | Sex | Age (years) | Sample size (n) | Healthy CRF (mL/kg/min) | AUC (95% CI) |
|---|---|---|---|---|---|---|---|---|---|
| Mesa *et al*[39] | 2006 | Spain | TC, HDLc, LDLc, Lp(a) | 20-m shuttle run test with estimated $VO_{2peak}$ | Boys | 13–18.5 | 248 | 45–47 | 0.72 (0.65 to 0.78) |
| Ruiz *et al*[41] | 2007 | Estonia and Sweden | Insulin, glucose, HDLc, TG, sum of five skinfolds, BP | Maximal cycle ergometer test with estimated $VO_{2peak}$ | Girls<br>Boys | 9–10<br>9–10 | 444<br>429 | 37<br>42 | 0.68 (0.62 to 0.73)<br>0.67 (0.61 to 0.73) |
| Lobelo *et al* [53] | 2009 | USA | SBP, triceps and subscapular skinfolds, HOMA, TG, TC/HDLc | Submaximal treadmill walking test with estimated $VO_{2peak}$ | Girls<br>Boys | 12–15<br>16–19<br>12–15<br>16–19 | 315<br>255<br>308<br>369 | 36<br>36<br>44<br>40 | 0.54 (0.42 to 0.66)<br>0.54 (0.38 to 0.70)<br>0.77 (0.68 to 0.86)<br>0.72 (0.64 to 0.79) |
| Adegboye *et al* [35] | 2011 | Denmark, Estonia, Portugal and Norway | SBP, TG, TC/HDLc, HOMA, sum of four skinfolds | Maximal cycle ergometer test with direct measurement of $VO_{2peak}$ | Girls<br>Boys | 8–11<br>14–17<br>8–11<br>14–17 | 1 181<br>1 051<br>1 219<br>1 049 | 37.4<br>33<br>43.6<br>46 | 0.67<br>0.69<br>0.68<br>0.55 |
| Moreira *et al* [50] | 2011 | Portugal | TG, HDLc, LDLc, waist to hip ratio, TC, glucose, BP | 20-m shuttle run test with estimated $VO_{2peak}$ | Girls<br>Boys | 10–18<br>10–18 | 255<br>195 | 39.5<br>41.8 | 0.65 (0.59 to 0.71)<br>0.65 (0.58 to 0.72) |
| Welk *et al*[36] | 2011 | USA | Waist circumference, SBP/DBP, HDLc, TG, glucose | Submaximal treadmill walking test with estimated $VO_{2peak}$ | Girls<br>Boys | 12–18<br>12–18 | 591<br>649 | 38–40<br>40–44 | 0.77<br>0.83 |
| Boddy *et al* [54] | 2012 | England | BMI, waist circumference, %BF | 20-m shuttle run test with estimated $VO_{2peak}$ | Girls<br>Boys | 9–10<br>9–10 | 8 237<br>8 382 | 41.9<br>46.6 | 0.58 to 0.77<br>0.74 to 0.80 |
| Silva *et al*[55] | 2012 | Portugal | Waist circumference, HDLc, TG, glucose, MAP | 20-m shuttle run test with estimated $VO_{2peak}$ | Girls<br>Boys | 10–18<br>10–18 | 2 789<br>2 770 | 32.6–43.9<br>43.1–46.3 | NA |
| Ruiz *et al*[47] | 2014 | Greece, Germany, Belgium, France, Hungary, Italy, Sweden, Austria and Spain | Smoking, BMI, physical activity, diet, TC, BP, glucose | 20-m shuttle run test with estimated $VO_{2peak}$ | Girls<br>Boys | 12.5–17.5<br>12.5–17.5 | 259<br>251 | 34.6<br>43.8 | 0.69 (0.61 to 0.78)<br>0.81 (0.73 to 0.88) |
| Silva *et al*[40] | 2016 | Brazil | BP | Modified Canadian Aerobic Fitness Test with estimated $VO_{2max}$ | Girls<br>Boys | 14–19<br>14–19 | 464<br>411 | 32<br>40 | 0.70 (0.65 to 0.74)<br>0.63 (0.58 to 0.67) |

- AUC, area under the curve; BMI, body mass index; BP, blood pressure; CRF, cardiorespiratory fitness; DBP, diastolic blood pressure; HDLc, high-density lipoprotein cholesterol; HOMA, homeostasis model assessment; Lp(a), Lipoprotein(a); LCLc, low-density lipoprotein cholesterol; MAP, mean arterial pressure; NA, not applicable; SBP, systolic blood pressure; TC, total cholesterol; TG, triglycerides; %BF, per cent body fat.

Figure 1 illustrates the variability of reported criterion-referenced standards for each sex by age by study group across the international 20mSRT normative values reported by Tomkinson and colleagues.[23] For both girls and boys, this figure follows a similar pattern where the criterion-referenced standards consistently fall below the international 20mSRT 50th percentile for younger children, indicating that the standards are attained at lower 20mSRT performance percentile ranks. However, for both sexes, there are age gradients where older children must attain a higher 20mSRT performance percentile rank to attain the criterion-referenced standards when compared with younger children. For girls, the variability across age-matched criterion-referenced standards spanned ~25 to 50 percentile points using Tomkinson and colleagues' normative-referenced standards.[23] The variability across age-matched standards was slightly smaller for boys, spanning ~20 to 45 percentile points.

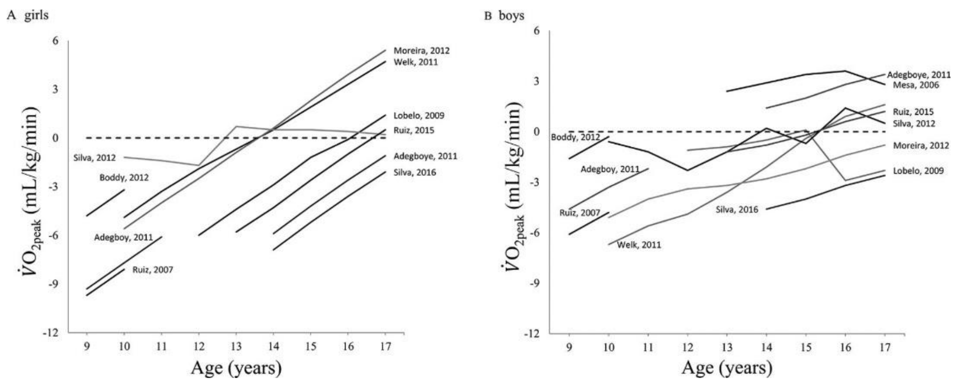


Figure 1
Study of criterion-referenced standards minus international normative 20-m shuttle run test (20mSRT) performance as a function of age for girls (A) and boys (B). The dash line, centred at 0, represents the international 20mSRT 50th percentile (predicted O2peak) for 9- to 17-year-old children and youth.23 Lines below the dash line indicate that the criterion-referenced standards are below the international 50th percentile of 20mSRT performance, and lines above represent standards above the 50th percentile for each age group.

## Percentage estimates for 'healthy' CRF

Table 3 displays 105 sex by age by study percentages (%) of children and youth that achieved each reported criterion-referenced standard. These represent the study-specific international percentage estimates of apparently 'healthy' children and youth calculated using the international 20mSRT norms. In general, younger children were more likely to meet the standards than older children, and mean percentages were more precise (smaller CI) for younger girls and older boys. More specifically, the percentage pattern across all studies was such that girls aged 9–11 years were apparently healthier than age-matched boys, girls and boys aged between 12 and 15 years were similarly healthy, and older boys aged between 16 and 17 years were apparently healthier than girls of the same age. This pattern was true for all studies, except for the Welk (FITNESSGRAM) criterion-referenced standards,[36] where boys were considered to be healthier than girls at every age group.

In addition, the smoothed population-weighted mean percentage (±95% CI) of girls and boys meeting the criterion-referenced CRF standards ranged from 36%±13% to 95%±4% and from 51%±7% to 96%±16%, respectively (figure 2). These smoothed population-weighted mean percentages represent our best estimate of the international mean percentage of children and youth aged 9–17 years old with apparently healthy CRF.

Table 3
Percentage of the 1 142 026 international 20mSRT performance pseudo-datasets that met each CRF criterion-referenced standard

| Age (years) | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
|---|---|---|---|---|---|---|---|---|---|
| Girls<br>n=552 900 | | | | | | | | | |
| Ruiz, 200741 | 99.9 | 98.6 | | | | | | | |
| Lobelo, 200953 | | | | 90.8 | 82.9 | 71.7 | 59.7 | 49.3 | 40.0 |
| Welk, 201136 | | 91.3 | 77.0 | 67.8 | 55.6 | 47.1 | 37.8 | 24.7 | 19.8 |
| Adegboye, 2011 35 | 99.9 | 93.5 | 93.5 | | | 87.7 | 78.8 | 68.9 | 57.3 |
| Moreira, 2012 | | 93.3 | 82.5 | 72.0 | 56.6 | 46.6 | 34.5 | 21.8 | 16.6 |
| Silva, 201255 | | 64.5 | 61.0 | 65.8 | 45.1 | 47.1 | 47.9 | 45.6 | 48.5 |
| Boddy, 201254 | 93.7 | 81.3 | | | | | | | |
| Ruiz, 2015 | | | | | 89.2 | 80.7 | 70.1 | 57.8 | 46.6 |
| Silva, 201640 | | | | | | 92.3 | 85.2 | 74.8 | 64.5 |
| Boys<br>n=589 126 | | | | | | | | | |
| Mesa, 200639 | | | | | 36.7 | 34.3 | 32.2 | 32.5 | 36.9 |
| Ruiz, 200741 | 92.9 | 87.0 | | | | | | | |
| Lobelo, 200953 | | | | 59.4 | 58.8 | 52.4 | 51.9 | 65.2 | 59.2 |
| Welk, 201136 | | 94.8 | 87.6 | 81.5 | 73.7 | 60.4 | 53.9 | 45.8 | 41.6 |
| Adegboye, 201135 | 88.3 | 79.0 | 67.4 | | | 41.4 | 40.8 | 35.5 | 32.9 |
| Moreira, 2012 | | 88.8 | 79.7 | 73.0 | 71.2 | 64.0 | 63.1 | 57.2 | 53.3 |
| Silva, 201255 | | 56.8 | 60.6 | 66.9 | 59.8 | 48.6 | 55.5 | 44.0 | 47.6 |
| Boddy, 201254 | 66.1 | 53.4 | | | | | | | |
| Ruiz, 2015 | | | | | 59.8 | 54.3 | 53.0 | 47.3 | 43.7 |
| Silva, 201640 | | | | | | 73.8 | 71.8 | 67.0 | 60.2 |

- 20mSRT, 20-m shuttle run test; CRF, cardiorespiratory fitness.

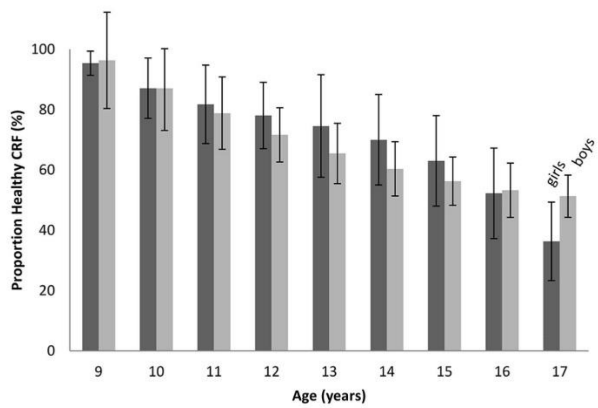


Figure 2
Overall smoothed population-weighted mean percentages and 95% CIs for healthy CRF in girls and boys. Calculated across age-specific percentage estimates for each criterion-referenced standard ($O_{2peak}$) available in table 3. CRF, cardiorespiratory fitness.

## Discussion

The main objectives of this paper were to identify all criterion-referenced standards for CRF generated using ROC curve technique; to estimate the international percentage of children and youth that achieved each CRF standard; and to discuss strategies to enhance the utility of CRF criterion-referenced standards for population health research.

### Criterion-referenced standards

Our paper identified 10 studies that reported CRF criterion-referenced standards using samples of children and youth from Europe, North America or South America. The reported criterion-referenced standards for CRF were developed using a variety of methods and protocols, and as a result, the standards vary considerably across sex by age groups. Recent research suggests that temporal trends exist in CRF[33] and body composition in children,[37] and metabolic health markers in adults.[38] These trends suggest that the relationship between CRF and health is not static, which may also help describe the different sex by age specific standards across studies given our 10-year time period between the first[39] and last[40] reported standards. For these reasons, comparisons of percentages of healthy CRF while using different standards should be interpreted with caution.

### Percentage estimates for 'healthy' CRF

Our study also provides percentage estimates for each reported criterion-referenced standard using international data that represent 1 142 026 children and youth aged 9–17 years. These data demonstrate that a high percentage of young children (9–10 years old) are apparently healthy. In some cases (Ruiz[41] and Adegboye[35]), girls aged 9 and 10 years would only need to achieve a 20mSRT performance below the 5th percentile to achieve the criterion-referenced standard when using the Léger prediction equation, equivalent to completing the first stage (1 min) of the 20mSRT. Comparatively, boys aged 9–10 years would have to perform above the 5th percentile to achieve most criterion-referenced standards, suggesting that a relatively better performance is required from younger boys than younger girls. Furthermore, these data illustrate an age gradient where older children must achieve a higher 20mSRT percentile than younger children to achieve the criterion-referenced standards. This translates into a higher overall percentage of healthy CRF in younger children than older children. For example, data in table 3 indicate that young children (9–10 years old) are apparently healthy with 72% of all cases having a success rate over 80%, in comparison with only 7% of older children (14–17 years old).

The FITNESSGRAM criterion-referenced standards (Welk et al [36]) for CRF are probably the most widely used standards to date. National percentages of children and youth who met these standards exist for USA,[42] Chile[43] and Hungary.[44] In USA and Hungary, a sex gradient existed where boys met the standards more often than girls at every age group. In addition, an age gradient existed for these countries where younger children met the standards more often than older children, supporting the pattern observed in Welk's percentages in table 3. The US-based percentages[42] indicated that 26%–49% of girls and 38%–62% of boys had healthy CRF, which was lower than our international percentages (using the Welk standards). This result is expected given that US children and youth have low CRF relative to international norms.[11 12] Our international percentage estimates are more similar to the Chilean estimates[43] which reported healthy CRF percentages in grade 8 children (13- and 14-year-olds) for 45% and 74% of girls and boys, respectively. Furthermore, percentage estimates for Hungarian children and youth aged 11–18 years[44] are between 32% to 86% for girls and between 45% to 86% for boys. These percentage estimates fall above our international percentage estimates, being consistent with previous findings showing that Hungarian (central European) children have better CRF relative to international norms.[11 12]

In addition, the Welk standards for boys are among the easiest to achieve and resulted in a higher percentage of boys with healthy CRF than girls. In comparison, most other reported criterion-referenced standards identified higher percentages of girls with healthy CRF than boys (table 3). This highlights the fact that using standards other than the Welk standards would have most likely resulted in a higher percentage of girls than boys with healthy CRF in USA, Chile and Hungary. Thus, the criterion-referenced standard applied to the sample makes a difference in the overall interpretation of results.

### International percentages of healthy children and youth

Figure 2 displays the international percentages of children and youth with apparently healthy CRF. However, these percentages should be interpreted with caution because of the variability across all 10 criterion-referenced standards, resulting in large 95% CIs. In addition, this study applied the Léger prediction equation which may yield systematically different $\dot{V}O_{2peak}$ values in comparison with other prediction equations. This may impact the published criterion-referenced standards as well as the percentage estimates of healthy CRF across each sex by age group, indicating that the prediction equation used may artificially increase or decrease the percentage estimates in a systematic way.

**Strategies for population health research**

The availability of different criterion-referenced standards for CRF across the same sex by age groups may create confusion. A recent meta-analysis by Ruiz et al [45] combines seven published criterion-referenced standards[35 36 39 41 47 50 53] on 9280 children and youth aged 8–19 years from 14 countries and provides the first international criterion-referenced standards for this age group. Their meta-analysis determined that fitness levels below 35 and 42 mL/kg/min for girls and boys, respectively, should raise a red flag. Internationally, 80% (95% CI 66% to 94%) of girls and 68% (95% CI 60% to 76%) of boys meet the international criterion-referenced standards, with younger children meeting the standards more often than older children. See online supplement 1 that provides an unweighted sex by age percentage for the Ruiz standards using the Tomkinson international pseudo-dataset. The Ruiz standards could be used as the interim overarching international standards to help compare percentages of apparently healthy children and youth across studies, countries and continents, providing clear international comparisons in this area of research. International standards could also be applied in a clinical and/or school setting by clinicians and practitioners, similar to the international body mass index growth curves[48] for children and youth, providing an effective screening instrument to help identify children and youth who need to increase their physical activity levels to improve overall health.

Our study also reports international referenced percentages for CRF in children and youth aged 9–17 years. These data are the first of their kind and carry specific utility for national and international population health research. These international referenced percentages can help researchers and practitioners (eg, physical educators, public health workers, health promotion practitioners, etc) interpret their percentage estimates by providing a reference to indicate whether their sample percentage is above or below the international percentage for a given sex by age group. This can help provide researchers and practitioners with context to their respective results. These international referenced percentages can also provide the first step towards establishing CRF and the 20mSRT as a tool for population health screening and surveillance.

**Strengths and limitations**

Our study compared 10 reported CRF criterion-referenced standards for children and youth, all of which were developed using the ROC curve method. We also describe percentages of apparently healthy children and youth using international normative centile values developed from a sample of over 1.1 million children and youth from 500 different countries.[23] These normative values were developed using rigorous data standardisation, poststratification population weighting and LMS techniques.

A limitation of our study may be comparing criterion-referenced standards across different studies. For instance, study-specific standards were generated using different combinations of CRF field-based or laboratory-based assessment protocols, submaximal-effort or maximal-effort protocols, different exercise modes and $\dot{V}O_{2peak}$ prediction equations. This is of particular concern given that treadmill protocols tend to elicit a $\dot{V}O_{2peak}$ roughly 9% higher than cycle ergometer protocols,[49] and the 20mSRT uses prediction equations to estimate $\dot{V}O_{2peak}$ that may introduce unintended error causing varying results.[50] Thus, each method used to estimate $\dot{V}O_{2peak}$ may yield different results due to different underlying levels of construct validity. Furthermore, each criterion-referenced standard used a different combination of clustered cardiometabolic health markers to establish the criterion domain. As a result, it is plausible that an individual is healthy given one set of cardiometabolic criteria, but unhealthy given a different set of criteria. This could explain some of the variability in our population percentage estimates. The existing criterion-referenced standards also fail to identify psychosocial health outcomes in children and youth, which remains an essential area of research given the importance of mental health.[51] Furthermore, these standards do not incorporate a measure of maturation and the use of ratio scaling means that body size differences may not have been appropriately partitioned out, both of which may have an impact on identifying developmentally appropriate age-specific and sex-specific standards.[52]

## Conclusion

Our study provides a review of criterion-referenced standards for CRF and demonstrates the inconsistencies among different standards using pseudo-data derived from a large international sample of 20mSRT performances. This study also provides international referenced percentages of children and youth meeting each criterion-referenced standard. Although there are several limitations to ROC-generated criterion-referenced standards, these represent the best available data, and percentage estimates should be interpreted with caution. Nonetheless, these standards are the first step towards generating clinically relevant standards that are comparable with body mass index standards. Researchers, clinicians and the general public are encouraged to use the interim international criterion-referenced standards of 35 and 42 mL/kg/min for girls and boys, respectively, to identify children and youth at risk of poor health—raising a clinical red flag.

**Acknowledgments**

The authors would like to acknowledge Kevin Belanger and Joel Barnes for their help with formatting and designing the figures for this manuscript.

**Contributors** JJL ,GRT and MST developed the research questions and objectives. JJL led the data analysis and synthesis of results, and drafted the manuscript. All authors contributed to the interpretation of results, editing and reviewing of the final manuscript. All authors approved the final manuscript.

**Funding** FBO and JRR have received a grant from the Spanish Ministry of Science and Innovation (grant nos. RYC-2011–09011 and RYC-2010–05957, respectively), and FBO's current research activity is under the umbrella of the ActiveBrains project (Reference DEP2013-47540).

**Competing interests** None declared.

**Provenance and peer review** Not commissioned; externally peer reviewed.

## References

1 Armstrong N, Tomkinson G, Ekelund U. Aerobic fitness and its relationship to sport, exercise training and habitual physical activity during youth. *Br J Sports Med* 2011;45:849–58.

2 Ortega FB, Ruiz JR, Castillo MJ, *et al.* Physical fitness in childhood and adolescence: a powerful marker of health. *Int J Obes* 2008;32:1–11.

3 Jiménez-Pavón D, Castillo MJ, Moreno LA, *et al*; Helena Study Group. Fitness and fatness are independently associated with markers of insulin resistance in European adolescents; the HELENA study. *Int J Pediatr Obes* 2011;6:253–60.

4 Thomas NE, Baker JS, Graham MR, *et al.* C-reactive protein in schoolchildren and its relation to adiposity, physical activity, aerobic fitness and habitual diet. *Br J Sports Med* 2008;42:357–60.

5 Morales PF, Sánchez-López M, Moya-Martínez P, *et al.* Health-related quality of life, obesity, and fitness in schoolchildren: the Cuenca study. *Qual Life Res* 2013;22:1515–23.

6 Ortega FB, Labayen I, Ruiz JR, *et al*. Improvements in fitness reduce the risk of becoming overweight across puberty. *Med Sci Sports Exerc* 2011;43:1891–7.

7 Ruiz JR, Castro-Piñero J, Artero EG, *et al*. Predictive validity of health-related fitness in youth: a systematic review. *Br J Sports Med* 2009;43:909–23.

8 Högström G, Nordström A, Nordström P. High aerobic fitness in late adolescence is associated with a reduced risk of myocardial infarction later in life: a nationwide cohort study in men. *Eur Heart J* 2014;35:3133–40.

9 Högström G, Nordström A, Nordström P. Aerobic fitness in late adolescence and the risk of early death: a prospective cohort study of 1.3 million Swedish men. *Int J Epidemiol* 2015:1159–68.

10 Léger L, Lambert J, Goulet A, *et al*. Capacité aérobie des québécois de 6 à 17 ans - Test navette de 20 mètres avec paliers de 1 minute [Aerobic capacity of 6 to 17 year old Québécois – 20 metre shuttle run test with 1 minute stages]. *Can J Appl Sport Sci* 1984;9:64–9.

11 Olds T, Tomkinson G, Léger L, *et al.* Worldwide variation in the performance of children and adolescents: an analysis of 109 studies of the 20-m shuttle run test in 37 countries. *J Sports Sci* 2006;24:1025–38.

12 Lang JJ, Tremblay MS, Léger L, *et al.* International variability in 20 m shuttle run performance in children and youth: who are the fittest from a 50-country comparison? A systematic literature review with pooling of aggregate results. *Brit J Sports Med*. 2016.

13 Mayorga-Vega M, Aguilar-Soto 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.

14 Tomkinson GR, Olds TS. Field tests of fitness. In: Armstrong N, van Mechlen W, eds. *Paediatric Exercise Science and Medicine.* New York, NY: Oxford University Press, 2000:109–28.

15 Council of Europe. Eurofit: handbook for the EUROFIT tests of physical fitness. *Secretariat of the Committee for the Development of Sport within the Council of Europe.* Rome: Council of Europe, 1998:1–75.

16 In: Plowman SA, Meredith MD, eds. *Fitnessgram/Activitygram Reference Guide*. 4th Edn. Dallas, TX: The Cooper Institute, 2013:1–202.

17 Longmuir PE, Boyer C, Lloyd M, *et al*. The Canadian Assessment of physical Literacy: methods for children in grades 4 to 6 (8 to 12 years). *BMC Public Health* 2015;15:767.

18 Australian Council for Health, Physical Education and Recreation. *Australian Fitness Education Award: User's Manual and Curriculum Ideas. Physical Education and Recreation*. Australia, SA: Australian Council for Health, 1996:1–59.

19 Ruiz JR, Castro-Piñero J, España-Romero V, *et al.* Field-based fitness assessment in young people: the ALPHA health-related fitness test battery for children and adolescents. *Br J Sports Med* 2011;45:518–24.

20 Cadenas-Sánchez C, Alcántara-Moral F, Sánchez-Delgado G, *et al*. Assessment of cardiorespiratory fitness in preschool children: Adaptation of the 20 metres shuttle run test. *Nutr Hosp* 2014;30:1333–43.

21 Cadenas-Sánche C, Martíne-Tellez B, Sánchez-Delgado G, *et al.* Assessing physical fitness in preschool children: Feasibility, reliability and practical recommendations for the PREFIT battery. *J Sci Med Sport* 2016;19:910–5.

22 Ortega FB, Cadenas-Sánchez C, Sánchez-Delgado G, *et al.* Systematic review and proposal of a field-based physical fitness-test battery in preschool children: the PREFIT battery. *Sports Med* 2015;45:533–55.

23 Tomkinson GR, Lang JJ, Tremblay MS, *et al.* International normative 20 M shuttle run values from 1142026 children and youth representing 50 countries. *Brit J Sports Med* 2016;0:1–14.

24 Catley MJ, Tomkinson GR. Normative health-related fitness values for children: analysis of 85347 test results on 9-17-year-old Australians since 1985. *Br J Sports Med* 2013;47:98–108.

25 Sandercock G, Voss C, Cohen D, *et al.* Centile curves and normative values for the twenty metre shuttle-run test in English schoolchildren. *J Sports Sci* 2012;30:679–87.

26 Santos R, Mota J, Santos DA, *et al.* Physical fitness percentiles for Portuguese children and adolescents aged 10–18 years. *J Sports Sci* 2014;32:1510–8.

27 Sauka M, Priedite IS, Artjuhova L, *et al*. Physical fitness in northern European youth: reference values from the Latvian Physical Health in Youth Study. *Scand J Public Health* 2011;39:35–43.

28 Secchi JD, García GC, España-Romero V, *et al.* Physical fitness and future cardiovascular risk in argentine children and adolescents: an introduction to the ALPHA test battery. *Arch Argent Pediatr* 2014;112:132–40.

29 Tambalis KD, Panagiotakos DB, Psarra G, *et al*. Physical fitness normative values for 6-18-year-old Greek boys and girls, using the empirical distribution and the lambda, mu, and sigma statistical method. *Eur J Sports Sci* Sept 24th 2015.

30 McNeil BJ, Keller E, Adelstein SJ. Primer on certain elements of medical decision making. *N Engl J Med* 1975;293:211–5.

31 Levy PS, Lemeshow S. Stratification random sampling: further issues. In: Levy PS, Lemeshow S, eds. *Sampling of Populations: Methods and Application*. Hoboken, NJ: John Wiley & Sons, Inc, 2008:143–88.

32 Cole TJ, Green PJ. Smoothing reference centile curves: the LMS method and penalized likelihood. *Stat Med* 1992;11:1305–19.

33 Tomkinson GR, Léger LA, Olds TS, *et al.* Secular trends in the performance of children and adolescents (1980-2000): an analysis of 55 studies of the 20m shuttle run test in 11 countries. *Sports Med* 2003;33:285–300.

34 Léger LA, Mercier D, Gadoury C, *et al.* The multistage 20 metre shuttle run test for aerobic fitness. *J Sports Sci* 1988;6:93–101.

35 Adegboye AR, Anderssen SA, Froberg K, *et al*. Recommended aerobic fitness level for metabolic health in children and adolescents: a study of diagnostic accuracy. *Br J Sports Med* 2011;45:722–8.

36 Welk GJ, Laurson KR, Eisenmann JC, *et al.* Development of youth aerobiccapacity standards using receiver operating characteristic curves. *Am J Prev Med* 2011;41:S111–S116.

37 Tremblay MS, Shields M, Laviolette M, *et al.* Fitness of Canadian children and youth: results from the 2007–2009 Canadian health measures survey. *Health Rep* 2009;21:7–20.

38 Danaei G, Finucane MM, Lu Y, *et al.* National, regional, and global trends in fasting plasma glucose and diabetes prevalence since 1980: systematic analysis of health examination surveys and epidemiological studies with 370 country-years and 2·7 million participants. *The Lancet* 2011;378:31–40.

39 Mesa JL, Ruiz JR, Ortega FB, *et al.* Aerobic physical fitness in relation to blood lipids and fasting glycaemia in adolescents: Influence of weight status. *Nutr Metab Cardiovasc Dis* 2006;16:285–93.

40 Silva DA, Tremblay M, Pelegrini A, *et al*. Association between aerobic fitness and high Blood Pressure in adolescents in Brazil: evidence for criterion-referenced cut-points. *Pediatr Exerc Sci* 2016;28:312–20.

41 Ruiz JR, Ortega FB, Rizzo NS, *et al.* High cardiovascular fitness is associated with low metabolic risk score in children: the European Youth Heart Study. *Pediatr Res* 2007;61:350–5.

42 Bai Y, Saint-Maurice PF, Welk GJ, *et al*. Prevalence of youth fitness in the United States: baseline results from the NFL PLAY 60 FITNESSGRAM partnership project. *J Pediatr* 2015;167:662–8.

43 Garber MD, Sajuria M, Lobelo F. Geographical variation in health-related physical fitness and body composition among Chilean 8th graders: a nationally representative cross sectional study. *PLoS One* 2014;9:e108053.

44 Welk GJ, Saint-Maurice PF, Csányi T. Health-related physical fitness in Hungarian youth: age, sex, and regional profiles. *Res Q Exerc Sport* 2015;86(Suppl 1): S45–S57.

45 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? A systematic review and meta-analysis. *Brit J Sports Med*. 2015.

46 Ruiz JR, Silva G, Oliveira N, *et al*. Criterion-related validity of the 20-m shuttle run test in youths aged 13–19 years. *J Sports Sci* 2009;27:899–906.

47 Ruiz JR, Huybrechts I, Cuenca-García M, *et al.* Cardiorespiratory fitness and ideal cardiovascular health in European adolescents. *Heart* 2014;0:1–8.

48 Cole TJ, Bellizzi MC, Flegal KM, *et al.* Establishing a standard definition for child overweight and obesity worldwide: international survey. *BMJ* 2000;320:1240–3.

49 Armstrong N, Davies B. An ergometric analysis of age group swimmers. *Br J Sports Med* 1981;15:20–6.

50 Moreira C, Santos R, Ruiz JR, *et al*. Comparison of different VO(2max) equations in the ability to discriminate the metabolic risk in Portuguese adolescents. *J Sci Med Sport* 2011;14:79–84.

51 Tomkinson G. Aerobic fitness thresholds for cardio metabolic health in children and adolescents. *Br J Sports Med* 2011;45:686–7.

52 Armstrong N, Welsman JR. Aerobic fitness: What are we measuring? *Sports Med* 2003;33:5–25.

53 Lobelo F, Pate RR, Dowda M, *et al.* Validity of cardiorespiratory fitness criterion-referenced standards for adolescents. *Med Sci Sports Exerc* 2009;41:1222–9.

54 Boddy LM, Thomas NE, Fairclough SJ, *et al*. ROC generated thresholds for field-assessed aerobic fitness related to body size and cardiometabolic risk in schoolchildren. *PLoS One* 2012;7:e45755.

55 Silva G, Aires L, Mota J, *et al.* Normative and criterion-related standards for shuttle run performance in youth. *Pediatr Exerc Sci* 2012;24:157–69.