2019

The 20-m shuttle run: Assessment and interpretation of data in relation to youth aerobic fitness and health

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

Justin J. Lang

Joel Blanchard

Luc A. Léger

Mark S. Tremblay

Title

The 20-m shuttle run: Assessment and interpretation of data in relation to youth aerobic fitness and health

Authors

Grant R. TOMKINSON,[*1,2] Justin J. LANG,[3,4] Joel BLANCHARD,[4] Luc A. LÉGER,[5] and Mark S. TREMBLAY[4]

Institutional affiliations

[1]Department of Education, Health and Behavior Studies, University of North Dakota, Grand Forks, ND, USA

[2]Alliance for Research in Exercise, Nutrition and Activity (ARENA), School of Health Sciences & Sansom Institute for Health Research, University of South Australia, Adelaide, SA, Australia

[3]Public Health Agency of Canada, Ottawa, Ontario, Canada

[4]Healthy Active Living and Obesity (HALO) Research Group, CHEO Research Institute, Ottawa, ON, Canada

[5]Department of Kinesiology, University of Montréal, Montréal, QC, Canada

Running title

20-m shuttle run: Assessment and interpretation

Corresponding author

Dr. Grant R. TOMKINSON[*]

✉ Department of Education, Health and Behavior Studies
University of North Dakota
2751 2nd Avenue North, Stop 8235
Grand Forks, ND, 58202, USA

☎ +1 701-777-4041

grant.tomkinson@und.edu

Abstract

Cardiorespiratory fitness (CRF) is a good summative measure of the body's ability to perform continuous, rhythmic, dynamic, large-muscle group physical activity and exercise. In children, CRF is meaningfully associated with health, independent of physical activity levels, and it is an important determinant of sports and athletic performance. Although gas-analyzed peak oxygen uptake is the criterion physiological measure of children's CRF, it is not practical for population-based testing. Field testing offers a simple, cheap, practical alternative to gas analysis. The 20-meter shuttle run test (20mSRT) — a progressive aerobic exercise test involving continuous running between two lines 20 meters apart in time to audio signals — is probably the most widely used field test of CRF. This review aims to clarify the international utility of the 20mSRT by synthesizing the evidence describing measurement variability, validity, reliability, feasibility, and the interpretation of results, as well as to provide future directions for international surveillance. We show that the 20mSRT is an acceptable, feasible and scalable measure of CRF and functional/exercise capacity, and that it has moderate criterion validity and high-to-very high reliability. The assessment is pragmatic, easily interpreted, and results are transferable to meaningful and understandable situations. We recommend that CRF, assessed by the 20mSRT, be considered as an international population health surveillance measure to provide additional insight into pediatric population health.

Keywords: 20-m shuttle run; cardiorespiratory fitness; children; validity; reliability; criterion-referenced standards; normative data.

1 Introduction

Cardiorespiratory fitness (CRF) provides a measure of the body's capacity to deliver and utilize oxygen for energy transfer to support muscle activity during physical activity and exercise (4). The CRF of children and youth has long been recognized as important to measure and monitor. At the turn of the 20th century, CRF was widely seen as an important component of physical fitness for wartime preparedness and talent identification in sports (40,97,98). In the 1970s, the measurement of CRF progressed into an important indicator of health (39), providing a reflection of an individual's overall health status (18). The utility of CRF as an index of health and as an indicator of aerobic athletic performance potential remains valued today. Although indirect calorimetry using expired gas analysis is considered the method of choice to measure peak oxygen uptake (peak $\dot{V}O_2$; the highest rate at which oxygen can be consumed during a progressive exercise test to volitional fatigue), this type of testing is neither popular nor feasible in school and public health settings. Historically, field tests of CRF have been the most common measure, beginning with distance/timed running tests (61), and later moving into progressive exercise shuttle run tests (68). More recently, the 20-meter shuttle run test (20mSRT), also called the "beep" test or the PACER (Progressive Aerobic Cardiovascular Endurance Run) test, has been identified as probably the most widely used field test of CRF among children and youth (64,128).

The 20mSRT involves continuous running back and forth between two parallel lines 20 meters apart in time to audio signals. It comprises a number of stages (also called levels), each lasting about 1 minute, with each stage comprising a number of 20-meter laps (also called shuttles). At each stage, the required running speed increases, until the child can no longer run the 20-meter distance in time with the audio signal (on two consecutive occasions) or when the child stops due to volitional fatigue. The test has moderate criterion validity against gas-analyzed peak $\dot{V}O_2$ (mL/kg/min) (see Section 3.1), high-to-very high test-retest reliability (see Section 3.2), and it is meaningfully associated with a variety of health indicators among children and youth (62). The 20mSRT has been supported by European experts from the Assessing Levels of Physical Activity (ALPHA) project (107), North American experts from the Institute of Medicine (IOM) (58), and UK experts from the British Association of Sport and Exercise Science (13), all of which provide authoritative recommendations about fitness testing in children and youth.

Given the international recognition of the 20mSRT, there has been a recent push to promote the international surveillance of pediatric CRF as a way to not only monitor current population health, but also to help anticipate future health and to help guide public health resource allocation (62). There is also potential merit in using the 20mSRT as a feasible and objective measure to help evaluate the impact of physical activity interventions (16,45,62,66,67,101,112,135,150). Consistent with these views, this review aims to further clarify the international applicability of the 20mSRT by synthesizing the evidence describing measurement variability, validity, reliability, feasibility and interpretation. We also aim to provide future direction for international 20mSRT surveillance.

2 What are we measuring?

CRF fitness reflects the overall capacity of the physiological systems (cardiovascular, respiratory, metabolic and neuromuscular) to perform continuous, rhythmic, dynamic, large-muscle group physical activity of moderate-to-high intensity for long periods. Field tests of CRF (e.g., the 20mSRT) provide a simple, single measure that assesses the integrated responses of the physiological systems required to perform prolonged exercise in a natural setting, whereas gas-analyzed tests of CRF (e.g., peak $\dot{V}O_2$ tests) directly measure the function or contribution of the underlying physiological systems to deliver and utilize oxygen (38). However, while laboratory-based peak $\dot{V}O_2$ is an excellent prognostic measure at the individual level, field tests of CRF are often preferred because they are low in cost and can be easily administered to a large groups of people, simultaneously. The 20mSRT provides a progressively intense aerobic physiological challenge that elicits a maximal effort in most children (143). At the individual level, the 20mSRT is a true indicator of peak $\dot{V}O_2$ (absolute or relative to body mass) because peak $\dot{V}O_2$ is achieved at the end of a maximal 20mSRT performance (111). It is also an excellent marker of functional/exercise capacity. It authentically imitates typical youth physical activities (e.g., jogging, running, sprinting, and agility through changes of speed/direction) and assesses the physiological responses of such activities at high and maximal intensities, which is important for children's physical activity levels and successful sports participation. In addition, run/walk tests induce the highest metabolic rate (peak $\dot{V}O_2$) compared to other exercise modes (e.g., cycling) in children (5,71,132) and adults (23,60,132).

While it is widely accepted that gas-analyzed peak $\dot{V}O_2$ is the criterion physiological measure of CRF, it is not the only indicator of CRF. In addition to peak $\dot{V}O_2$, maximal aerobic performance could also be affected by other physiological (e.g., running economy, fractional utilization of oxygen, $\dot{V}O_2$ kinetics, anaerobic capacity [15], agility [149]), physical (e.g., fat mass [42]) and psychosocial factors (e.g., motivation, self-efficacy [27]). Theoretically, a lower relative peak $\dot{V}O_2$ will impair 20mSRT performance because peak $\dot{V}O_2$ limits the rate at which oxygen can be delivered and utilized in energy transfer; a lower mechanical efficiency will change the running speed-$\dot{V}O_2$ relationship and increase the oxygen cost for any given running speed; a lower fractional utilization of oxygen will reduce the length of time sustained for any given running speed; poorer $\dot{V}O_2$ kinetics will reduce the rate at which $\dot{V}O_2$ rises at the start of exercise and to meet progressive increases in running speed; a lower anaerobic capacity will reduce the amount of energy supplied (via anaerobic metabolism) for maximal short-duration exercise, which will be especially important for children when trying to complete the final laps of the 20mSRT or for young and/or poorly conditioned children who can only complete a few 20mSRT stages; and, poorer agility will impair the ability to rapidly change direction/speed and increase the oxygen cost of running.

Fat mass could also affect 20mSRT performance. Because fat mass is partly metabolically inactive and constitutes an additional load to carry, increased fat mass will decrease mass-specific peak $\dot{V}O_2$ (mL/kg/min) approximately on a pro rata basis (42,50). It is also possible that increased fat mass will increase the oxygen cost of running at any given speed relative to total body mass (94). This is because peripherally located fat mass (on the arms and legs) increases both rotational and translational kinetic energy (31), and fat mass results in additional metabolic maintenance costs (e.g., breathing, thermoregulation) (94). However, it is probable that major increases in fat mass are required to alter the running mechanics enough to change energy requirements (42).

Psychosocial factors, both cognitive (e.g., motivation, the ability to tolerate discomfort, the ability to judge pace and effort) and affective (e.g., self-efficacy), also appear to affect 20mSRT performance. Self-efficacy is a significant positive correlate of 20mSRT, with children who have higher levels of generalized self-efficacy toward physical activity (i.e., those who perceive themselves to be more physically capable and who prefer physical activities over sedentary activities) tending to perform well on the 20mSRT (27). Fitness, fatness, physical activity and self-efficacy can interact through many possible causal pathways (27). For example, fatter children tend to have lower self-efficacy, which may make them less motivated to produce a maximal aerobic effort (27). It is also possible that children with low CRF are less motivated to exercise resulting in increased fatness, or those who are more physically active feel competent because they are fitter and have lower fatness.

3 Validity, reliability and feasibility

3.1 Validity

Validity is the degree to which a test measures what it is supposed to measure. Validity has important implications for: (a) predicting an individual's criterion value, (b) deciding whether a test has sufficient validity for tracking changes in an individual's criterion value, (c) comparing the validity of different tests, and (d) sample size estimation in descriptive or cross-sectional studies (57).

The criterion validity of the 20mSRT as a measure of CRF is *moderate*. In a 2015 meta-analysis of 57 studies and 78 validity correlations, Mayorga-Vega et al. (84) estimated the population criterion validity of the 20mSRT in children and adults. For children, they reported the criterion validity as moderate for estimating gas analyzed peak $\dot{V}O_2$ in mL/kg/min (corrected mean *r* at the population level [95%CI]: r_p=0.78 [0.72, 0.85]). Mayorga-Vega et al. (84) showed that mean criterion validity was considerably higher:

(1) for Léger's protocol (68,70) than other protocol variants (e.g., the Eurofit and Queen's University of Belfast protocols — see Section 4 for protocol descriptions);

(2) when other variables (e.g., sex, age or body mass) were added to the prediction model; and

(3) for adults than for children.

Sex and peak $\dot{V}O_2$ levels did not significantly affect the criterion validity of the 20mSRT.

Léger et al. (70) suggested that the lower criterion validity in children relative to adults may be the result of larger inter-individual variability in biological age, since chronological age significantly predicts peak $\dot{V}O_2$ in children but not adults. Despite the child-adult differences, the 20mSRT has moderate criterion validity in apparently healthy children and youth and can be used as a good alternative for estimating CRF when gas analyzed measurement of peak $\dot{V}O_2$ is not appropriate, practical, or feasible. However, the 20mSRT may not be a good measure of CRF for all children, especially young children or those with physical impairment (see Section 4). Further examination of the data from Mayorga-Vega et al. (84) revealed that the criterion validity of the 20mSRT increases throughout childhood and adolescence, from moderate criterion validity in childhood to higher validity in adolescence (Figure 1). The slopes of the regression lines relating age and criterion validity were similar for boys and girls ($\beta_1 - \beta_2$ [95%CI]: 0.02 [–0.12, 0.17]), indicating that sex did not significantly influence the age-related changes in validity (Figure 1). The residual variability is probably explained by other physiological, physical, and psychosocial factors (see Section 2), which may be more pronounced in the 20mSRT performance of children than youth.

****insert Figure 1 about here****

It is important to note that the validity correlation is sensitive to sample heterogeneity, which helps explain the considerable variability within age groups in Figure 1; in contrast, the standard error of estimate (SEE) (i.e., the average random error in prediction) is not (57). While SEEs have been less frequently reported for the 20mSRT than validity correlations, an analysis of 10 studies (3,14,35,53,70,78,83,87,111,121) estimates the average SEE to be 4.9 mL/kg/min (range: 1.3 to 7.1) or ~12% (range 3–18%, assuming a peak $\dot{V}O_2$ of 40 mL/kg/min, which is typical of children aged 9–17 years using international norms [128]). The 95% likely range for a true peak $\dot{V}O_2$ value estimated from the 20mSRT is ~10 mL/kg/min or ~24%.

3.2 Reliability

Test-retest reliability (henceforth termed "reliability") is the degree to which a test produces consistent results. Reliability has important decision-making implications for: (a) assessing an individual with a single measurement or repeated measurements, (b) estimating the extent of individual responses to a treatment in an experiment, (c) comparing the reliability of different tests or measurers, and (d) sample size estimation in experimental or longitudinal studies (57).

The reliability of the 20mSRT is *high-to-very high*. In a systematic review of 32 reliability studies, five of which assessed 20mSRT, Artero et al. (6) reported reliability coefficients (intra-class correlation coefficients) for the 20mSRT ranging from 0.78 to 0.93 in children aged 8–18 years. Test-retest differences in means are negligible in children and youth (25,70,72

96,100); meaning, 20mSRT performance is independent of test familiarity and prior practice. The 95% limits of agreement are approximately ±2.5 stages/minutes (96).

3.3 Feasibility

The feasibility of the 20mSRT is *high* according to the IOM's evaluation criteria for administrative feasibility of a fitness test (58). It can be conducted in a timely and efficient manner; it imposes acceptable preparation burden on both participants and testers; it can be administered with acceptable privacy, minimal equipment and space; and performance is independent of test familiarity and prior practice (see Section 3.2). Relative to other field tests of CRF (e.g., distance/timed running tests), the external pacing helps to mitigate pacing variability and the physiological burden on the child is somewhat reduced because only the final stage is maximal. The 20mSRT also has strong utility and scalability for population health surveillance (63). Domone et al. (47) described the scalability of field-based fitness measures as the ability of a measure to attain six criteria:

(1) *delivery* (can the test be delivered in schools, in a timely manner, and by non-technical staff, and is the test suitable for longitudinal research?);

(2) *evidence of operating at scale* (is the test appropriate for population-based testing, and are schools likely to accept the test?);

(3) *effectiveness* (does the test demonstrate acceptable validity and reliability, and is there a high level of participation and completion?);

(4) *cost* (is the test cost-effective?);

(5) *resource requirements* (are the equipment, space, skills, competence, and workforce requirements minimal?); and

(6) *practical implications* (can the test be implemented and scored with ease?).

In their systematic review of field-based tests of CRF, Domone et al. (47) identified the 20mSRT as the most scalable for school-based fitness testing. Furthermore, the 20mSRT can help identify children at increased risk of developing future cardiometabolic disease (65), as well as help participants better understand CRF through comprehensive fitness education (49). Unfortunately, the 20mSRT is vulnerable to interpretation misuse if physical educators, exercise professionals, and those interpreting and communicating 20mSRT results are not fully familiar with the factors affecting measurement variability, validity, reliability, and the meaning of norm-referenced and criterion-referenced standards (see Section 5).

4 Assessment

While there has never been a formal cost-benefit analysis of CRF tests, the 20mSRT offers several advantages over other field tests of CRF (e.g., distance/timed running tests such as the 1.6-kilometer and the 12-minute runs). First, the 20mSRT running course is shorter and requires less space. Second, the 20mSRT can be conducted indoors, where the environmental conditions can be more easily controlled. Third, because the 20mSRT is externally paced, cognitive aspects of maximal aerobic performance are less likely to be important. Fourth, testing personnel can more closely monitor participants. Fifth, for many children, because the intensity is progressively increased, the initial stages of the 20mSRT serve as a warm-up. Sixth, the 20mSRT, and not distance/timed running tests, is favorably related to health risk, specifically indicators of adiposity and cardiometabolic risk (58). Finally, while the 20mSRT and distance/run tests are similarly reliable, data from two large meta-analyses (84,85) indicate that in children and youth, the 20mSRT has similar criterion validity to the 2400-meter (1.5-mile) and 12-minute run tests and higher validity than other distance/timed run tests.

However, the 20mSRT is not without its disadvantages. For some children (e.g., young and/or poorly conditioned children), the initial stages of the most widely used 20mSRT protocols (see next paragraph) may be too fast, leading to premature onset of fatigue and a performance unrepresentative of a true maximal aerobic effort. Ideally, the 20mSRT, like other progressive aerobic exercise tests, should last at least five minutes in order to obtain a valid estimate of peak $\dot{V}O_2$ (89). Unfortunately, international norms indicate that most children and youth fail to run at least five minutes (128). This is why several authors have used adapted versions of the 20mSRT for children and youth with substantially reduced starting speeds (e.g., 4 km/h in obese youth [102] and 6.5 km/h in preschool children [26]). In addition, the 20mSRT can only be performed by able-bodied, ambulatory children, and may not be suitable for those with physical impairment. However, adapted versions of the 20mSRT have been developed for children with physical impairment, including the 10-meter shuttle run test for children with cerebral palsy who can run/walk (141) and the 10-meter shuttle ride test for wheelchair using children with cerebral palsy (142).

Despite its wide usage, there are numerous methodological issues associated with the 20mSRT, including the existence of several protocol variants, different audio versions used to conduct the test, and different performance metrics. While some of these issues are briefly described below, the reader is referred to Tomkinson et al. (128,129) for a thorough review of these methodological issues. Tomkinson et al. (129) describe the three major protocol variants of the 1-minute stage test as:

(1) Léger's original 1-minute protocol (68,70), which starts at a speed of 8.5 km/h and increases in speed by 0.5 km/h each minute;

(2) The protocol used by the Eurofit (41), the FitnessGram® (125), the Australian Sports Commission (10), and the British National Coaching Foundation (21), among others. In this protocol, participants start at a speed of 8.0 km/h, the second stage is at 9.0 km/h, and thereafter increases in speed by 0.5 km/h each minute; and

(3) The Queen's University of Belfast protocol (104), which starts at 8.0 km/h and increases in speed by 0.5 km/h each minute.

It is also important to realize that some researchers have been unaware of these protocol variants. For example, some have inappropriately cited Léger and Lambert (69) in their methods sections — Léger and Lambert (69) actually describe a 20mSRT with 2-minute stages designed for use in adults. Other 1-minute protocol variants have also been described. The 15-meter shuttle run test (15mSRT) was developed for testers with smaller-sized facilities, with a conversion chart required to adjust 15mSRT results to 20mSRT results (125). An adapted 20-meter shuttle run/walk test, starting at a speed of 4.0 km/h and increasing in speed by 0.5 km/h each minute, has been used in a longitudinal interventional study of obese youth admitted to a rehabilitation center for a 9-month obesity management program (102). The FITness testing in PREschool children [PREFIT] 20mSRT (26), designed for use in pre-school children aged 3–5 years, is also conducted over the 20-meter distance. However, it starts at a speed of 6.5 km/h and increases in speed by 0.5 km/h each minute, and it is recommended that two testers concurrently run with a small group of children in order to assist with pacing. Given the lower criterion validity associated with the 20mSRT in children compared with youth (see Section 3.1), protocol variants such as the PREFIT are necessary in order to obtain an acceptably valid assessment of CRF in preschool children (90).

There are also several different audio versions (digital audio files, compact discs or compact cassettes), often produced in-house, that have been used for the same protocol. Methodological variations on these versions (e.g., calling the stage number at the start of the stage versus the end of each stage; using only full minutes versus both full minutes and half minutes to indicate completed stages) means that identical 20mSRT performances can be reported in different ways. In addition to variation in protocols, there has been variation in how results have been reported. Individual results have been reported as the running speed (km/h) at the last completed stage, the number of completed stages, the number of minutes the test lasted, the distance run, the number of completed laps (or stages plus laps), or as predicted peak $\dot{V}O_2$ (mL/kg/min) based on regression equations (128,129). Furthermore, 20mSRT performances are also affected by factors such as environmental conditions, clothing, field surfaces, footwear, motivation, pre-test instructions, diurnal variation, and the purpose and context of testing.

We make the following recommendations to minimize methodological variability:

(1) accurately report the 20mSRT protocol (e.g., based on criterion validity analyses, Léger's protocol (68,70) is recommended in apparently healthy children and youth [see Section 3.1]);

(2) take care to minimize and report factors that affect 20mSRT performance (e.g., children should be given the opportunity to practice the test to minimize the potential influence of affective factors such as motivation; factors such as testing/environmental conditions [ground, weather, temperature] should be accurately reported; and testing should be postponed if the conditions are too adverse); and

(3) best practice should include 20mSRT results being reported in multiple metrics (e.g., as the number of completed laps, half stages, and running speed [km/h] at the last completed stage), and considering the various metrics, speed is the only unequivocal metric.

At the completion of the 20mSRT, children should complete a cool-down period involving light-to-moderate intensity aerobic activity lasting ~5–10 minutes (2). Children at risk for cardiovascular complications during or immediately after maximal aerobic exercise should seek medical advice before participating in the 20mSRT (2). Any adverse events (or lack thereof) associated with the 20mSRT should also be reported (75). While data on adverse events resulting from children performing maximal aerobic exercise are scant (62,75), no adverse events have been recorded in the two largest UK-based fitness studies (Liverpool SportsLinx and East of England Healthy Hearts Study) in which 20mSRT assessments were made on ~80,000 children aged 9–16 years (13).

5 Interpretation

There are two main approaches used to help interpret 20mSRT results: criterion-referenced and normative-referenced standards. Health-related criterion-referenced standards are used as a screening tool to identify children and youth at increased risk of future diseases. This approach is largely driven by a public health or medical lens by providing standards to help practitioners identify children who may need intervention (65). Similarly, performance-related fitness screening is used in the armed forces, police force and fire brigade to identify recruits who can effectively carry out operational tasks, as well as for athlete identification (9). In contrast, normative-referenced standards allow for comparison to a reference population in order to determine how well a child compares to his/her peers. This approach is common in physical education and has historically been used to compare and track sports/athletic performances against centile bands to identify expected, better than expected, or worse than expected developmental changes. Taken together, criterion-referenced and normative-referenced standards provide a detailed interpretation of a child's 20mSRT performance (65). Below, we review both approaches while highlighting important next steps.

5.1 Health-related criterion-referenced standards

In 2017, a review (65) highlighted 10 sets of criterion-referenced standards for children and youth aged 8–18 years (1,19,73,88,91,109,110,118,120,144). These studies were largely published between 2006 and 2016 on North American, European, or South American children and youth, and used a variety of health outcomes to derive the standards. Since then there have been three newly published studies, one in British children (22), one in Canadian children (119), and another in Macedonian children (99). These papers help build a growing body of evidence with large population and public health implications (62). Figure 2 shows the available criterion-referenced standards in mL/kg/min values for boys and girls aged 8–18 years. Indeed, the variability across standards is problematic and creates difficulties in selecting an appropriate standard for research, public health, or clinical practice. Previously, we have proposed (65) the development of universal health-related criterion-referenced standards for CRF in children and youth to help create international consensus, similar to global physical activity guidelines (147). To date, the best available international standards are those from the Ruiz et al. (108) meta-analysis, which indicate that values below 42 and 35 mL/kg/min should raise concern among boys and girls, respectively. The main criticism of these standards is the potential lack of face validity through the absence of age-specific standards. More research is needed to investigate this concern.

****insert Figure 2 about here****

Similar to the meta-analysis approach used by Ruiz et al. (108), future research should determine whether universal age- and sex-specific criterion-referenced standards can be identified by pooling standards that use different health-related criterion indicators. It may also be possible to identify convergence or divergence across all available standards to help inform a universal standard. The sample weighted age- and sex-specific mean standards shown in Figure 2 may help inform this type of conversation. Furthermore, research should evaluate whether universal standards would be valid for children and youth across ethnic or cultural groups, and whether standards should be developed using measured 20mSRT performance values (i.e., laps, stages, or running speed) instead of predicted peak $\dot{V}O_2$ values, similar to the approach used by Buchan et al. (22).

5.2 Normative-referenced standards

In 2017, a large systematic review and analysis developed the first international normative-referenced standards for 20mSRT performance in 9- to 17-year-old boys and girls by pooling data from 1,142,026 children and youth from 50 countries (128). These norms include data from 16 low- and middle-income countries from Africa, Asia, and South America. Table 1 summarizes additional normative-referenced standards available from around the world for apparently healthy children and youth aged 6–18 years. In addition, Cadenas-Sánchez et al. (26) provide the first norms for pre-school children as part of the PREFIT test battery. These types of normative-referenced standards could be used to standardize test scores (i.e., develop *z*-scores) to facilitate comparisons between countries or jurisdictions, similar to what has been done elsewhere (64). This approach is similar to the international growth curves used in studies to standardize BMI scores among children and youth (36). We recommend reporting standardized scores using the international normative-referenced standards to facilitate international comparisons. When comparing within countries, it is recommended to standardize test scores using regional or national norms to supplement scores standardized using the international norms.

****insert Table 1 here****

Future research should further evaluate whether normative-referenced standards should be corrected for testing conditions. For example, the Colombian norms presented by Ramirez-Velez et al. (102) adjusted 20mSRT performances by a factor of 1.11 to account for the 2,625 m elevation of Bogotá — the city in which the majority of data were collected. There is also a need to facilitate the rapid update of national, regional, and international normative-referenced values to help accommodate temporal trends in CRF (127). We have previously recommended web-surveillance to allow researchers from across the world to share their 20mSRT results to help update norms and potentially identify target populations at increased risk (64). This is an ongoing effort that could assist with future research and surveillance.

5.3 Application to policy

Among children and youth, CRF levels are strongly associated with current health status (62), predictive of future health status in adulthood (106), and highly correlated with income inequality (64,127) — an important indicator of population health status (145). The inclusion of CRF in both national and international health surveillance systems could help complement physical activity and obesity measures to provide a better interpretation of pediatric health. Recently, there has been increased discussion about including CRF measures to help evaluate the impact of health promoting policy (62,66,113). This approach may hold promise because measures of CRF (e.g., indirect calorimetry [45,93,122,150], the 20mSRT [67,93,122], distance/timed running/walking [16,67,122,150], and heart rate recovery methods [16,45,67,122]) appear sensitive among children and youth to changes in physical activity levels resulting from school-based interventions (45,93,122) and other intervention settings (16,67,112,150), whereas other measures (e.g., BMI, blood pressure, blood cholesterol) appear less sensitive to these types of interventions (45,122). Thus, monitoring temporal trends (127) or geographic variability (64) in CRF through standardized surveillance efforts could help provide additional insight into pediatric population health to help better inform practitioners and policy makers about not only the impact of implemented health interventions, but also the potential changes in health that may have implications for future resource allocation.

6 Future directions — where to from here?

Field testing of CRF has been occurring around the world for many decades. The motivation for such testing varies from country to country and has evolved over time, but generally has clustered around military and occupational preparedness, fitness assessment and monitoring, obesity prevention, and health promotion and surveillance. Widespread concern over decreasing physical activity and fitness, and increasing overweight and obesity among children and youth worldwide (17,52,58,74,91,127,130,148) has provoked several domestic and global directives, strategies, and action plans, including the Toronto Charter for Physical Activity (24), the United Nation's Political Declaration of the High-level Meeting of the General Assembly on the Prevention and Control of Non-communicable Diseases (133), the Bangkok Declaration on Physical Activity for Global Health and Sustainable Development (59), and most recently the World Health Organization Global Action Plan on Physical Activity (146). It could be argued the time is right for a fitness renaissance and legitimization of surveillance of fitness as an important indicator of personal and population health (58,62,91,105). Supporting this, the Active Healthy Kids Global Alliance recently added "physical fitness" as an indicator in the harmonized process used to develop country report cards on the physical activity of children and youth (8).

In this paper we built the case for the 20mSRT to be widely used as a measure of CRF in a resurgence of fitness monitoring and surveillance among children and youth, worldwide. We have shown that the 20mSRT is an acceptable, feasible and scalable measure of CRF and functional/exercise capacity in children and youth, and that it has moderate criterion validity and high-to-very high reliability. The assessment is pragmatic, easily interpreted and results are transferable to meaningful and understandable situations. It is likely for these reasons that the 20mSRT has been used globally for many years (128). It is used in interventions for baseline fitness testing and to monitor progress in response to exercise training or physical activity programs, often preferred over laboratory measured peak $\dot{V}O_2$ (either maximal or submaximal) because of its low cost and ease of administration. Importantly, the 20mSRT is a predictor of future health (106). The 20mSRT provides a simple, single measure that assesses the integrated responses of the physiological systems' ability to perform continuous, rhythmic, sustained, large-muscle group exercise. The 20mSRT does not provide specific information on the function or contribution of specific systems that can be obtained from a peak $\dot{V}O_2$ test (38). However, because the 20mSRT is reflective of functional/exercise capacity and peak $\dot{V}O_2$, it can be used in apparently healthy populations as an index of one and/or the other.

Sustainable Development Goals adopted by the United Nations in 2015 (134) have charted the course for the global community until 2030. The Bangkok Declaration (59) shows how intricately physical activity is woven into many of these goals. The World Health Organization's Global Action Plan for Physical Activity provides a beacon for global efforts to increase the physical activity of people around the world. Monitoring progress towards increasing physical activity is very challenging (52). Measures of fitness, including the 20mSRT, provide a robust, standardized alternative to assess population health at a country level. To move forward the adoption of the 20mSRT for future surveillance efforts there is a need to:

- seek consensus on recording and presentation of testing protocols, scoring and interpretation of results;

- orchestrate international efforts to establish criterion-based CRF standards (especially for the 20mSRT) for a variety of health indicators, across all ages;
- make and calibrate necessary adjustments to testing and screening protocols across ages from the early years through to adolescence;
- develop a standard multilingual test package, including digital audio files and a standard operating manual;
- create a global repository to house CRF data and periodically compile global norms for use in monitoring and surveillance;
- establish a global effort to track and monitor changes in CRF using standardized protocols; and
- conduct further research to better understand the merit of using CRF to evaluate physical activity policy and interventions.

7 Conclusion

The 20mSRT is a good marker of functional/exercise capacity and CRF is favorably associated with health among children and youth. The test is feasible and evidence suggests that it could be applied in a variety of contexts. Although future research and international efforts are needed, the time seems right to engage in collaborations to further discuss the merits of including CRF (assessed by the 20mSRT) as an international surveillance measure to help better understand the health status of pediatric populations, and to help evaluate the impact of health promoting policies and interventions.

Disclaimer

The content and views expressed in this article are those of the authors and do not necessarily reflect those of the Government of Canada.

Conflicts of interest

All authors declare no conflicts of interest and received no funding for this project.

Contributors

GRT and JJL developed the aims of the paper. All authors contributed to the writing of the paper, editing and critical reviewing of the final paper, and approved the final paper.

References

1 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(9):722–728.

2 American College of Sports Medicine. ACSM's guidelines for exercise testing and prescription. 10th ed. Baltimore (MA): Lippincott Williams & Wilkins; 2018.

3 Anderson GS. The 1600-m run and multistage 20-m shuttle run as predictive tests of aerobic capacity in children. Pediatr Exerc Sci. 1992;4(4):312–318.

4 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(11):849–858.

5 Armstrong N, Williams J, Balding J, Gentle P, Kirby B. The peak oxygen uptake of British children with reference to age, sex and sexual maturity. Eur J Appl Physiol Occup Physiol 1991;62(5):369–375.

6 Artero EG, España-Romero V, Castro-Piñero J, Ortega FB, Suni J, Castillo-Garzon MJ, Ruiz JR. Reliability of field-based fitness tests in youth. Int J Sports Med. 2011;32(3):159–169.

7 Aslan E, Müniroğlu S, Alemdaroğlu U, Karakoç B. Investigation of the performance responses of yo-yo and shuttle run tests with the treadmill run test in young soccer players. Pamukkale J Sport Sci. 2012;3(3):104–112.

8 Aubert S, Barnes JD, Adbeta C, et al. Global Matrix 3.0 of physical activity report card grades for children and youth: Results and analysis from 49 countries. J Phys Act Health. 2018;15(S2):S251–273.

9 Australian Sports Commission. Australia's winning edge 2012–2022: Our game plan for moving from world class to world best. Belconnen (ACT): Australian Sports Commission; 2012.

10 Australian Sports Commission. 20 m shuttle run test: A progressive shuttle run test for measuring aerobic fitness. Belconnen (ACT): Australian Coaching Council; 1999.

11 Aziz AR, Mukherjee S, Chia MY, Teh KC. Relationship between measured maximal oxygen uptake and aerobic endurance performance with running repeated sprint ability in young elite soccer players. J Sports Med Phys Fitness. 2007;47(4):401–407.

12 Aziz AR, Tan FH, Teh KC. A pilot study comparing two field tests with the treadmill run test in soccer players. J Sports Sci Med. 2005;4(2):105–112.

13 Barker A, Williams C, Tolfrey K, Fawkner S, Sandercock G. The BASES expert statement on measurement and interpretation of aerobic fitness in young people. Leeds: British Association of Sport and Exercise Sciences; 2013.

14 Barnett A, Chan LYS, Bruce IC. A preliminary study of the 20-m multistage shuttle run as a predictor of peak $\dot{V}O_2$ in Hong Kong Chinese students. Pediatr Exerc Sci. 1993;5(1):42–50.

15 Bassett DR Jr, Howley ET. Limiting factors for maximum oxygen uptake and determinants of endurance performance. Med Sci Sports Exerc. 2000;32(1):70–84.

16 Beets MW, Beighle A, Erwin HE, Huberty JL. After-school program impact on physical activity and fitness: A meta-analysis. Am J Prev Med. 2009;36(6):527–537.

17 Blair SN. Physical inactivity: The biggest public health problem of the 21st century. Br J Sports Med. 2009;43(1):1–2.

18 Blair SN, Kohl HW 3rd, Paffenbarger RS Jr, Clark DG, Cooper KH, Gibbons LW. Physical fitness and all-cause mortality: A prospective study of healthy men and women. JAMA. 1989;262(17):2395–2401.

19 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(9):e45755.

20 Boreham CA, Paliczka VJ, Nichols AK. A comparison of the PWC170 and 20-MST tests of aerobic fitness in adolescent schoolchildren. J Sports Med Phys Fitness. 1990;30(1):19–23.

21 Brewer J, Ramsbottom R, Williams C. Multistage fitness test: A progressive shuttle-run test for the prediction of maximum oxygen uptake. Leeds: National Coaching Foundation; 1988.

22 Buchan DS, Knox G, Jones AM, Tomkinson GR, Baker JS. Utility of international normative 20 m shuttle run values for identifying youth at increased cardiometabolic risk. J Sport Sci. 2019;37(5):507–514.

23 Buchfuhrer MJ, Hansen JE, Robinson TE, Sue DY, Wasserman K, Whipp BJ. Optimizing the exercise protocol for cardiopulmonary assessment. J Appl Physiol Respir Environ Exerc Physiol. 1983;55(5):1558–1564.

24 Bull FC, Gauvin L, Bauman A, Shilton T, Kohl HW 3rd, Salmon A. The Toronto charter for physical activity: A global call for action. J Phys Act Health. 2010;7(4):421–422.

25 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(6):1333–1343.

26 Cadenas-Sanchez C, Intemann T, Labayen I, et al. Physical fitness reference standards for preschool children: The PREFIT project. J Sci Med Sport. Published Online First: 20 September 2018. DOI: 10.1016/j.jsams.2018.09.227.

27 Cairney J, Hay JA, Faught BE, Léger L, Mathers B. Generalized self-efficacy and performance on the 20-metre shuttle run in children. Am J Hum Biol. 2008;20(2):132–138.

28 Carrel AL, Bowser J, White D, et al. Standardized childhood fitness percentiles derived from school-based testing. J Pediatr. 2012;161(1):120–124.

29 Castro-Piñeiro J, Ortega FB, Keating XD, González-Montesinos JL, Sjöstrom M, Ruiz JR. Percentile values for aerobic performance running/walking field tests in children aged 6 to 17 years: Influence of weight status. Nutr Hosp. 2011;26(3):572–578.

30 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(2):98–108.

31 Cavanagh PR, Kram R. Stride length in distance running: Velocity, body dimensions, and added mass effects. Med Sci Sports Exerc. 1989;21(4):467–479.

32 Chatterjee P, Banerjee AK, Das P. Applicability of an indirect method to predict maximum oxygen uptake in young badminton players of Nepal. Int J Sports Sci Eng. 2010;4(4):209–214.

33 Chatterjee P, Banerjee AK, Das P. A prediction equation to estimate the maximum oxygen uptake of school-age girls from Kolkata, India. Malays J Med Sci. 2011;18(1):25–29.

34 Chatterjee P, Banerjee AK, Das P, Debnath P, Chatterjee P. Regression equations to predict $\dot{V}O_2$ max in untrained boys and junior sprinters of Kolkata. J Exerc Sci Physio. 2008;4(2):104.

35 Chia M, Aziz AR, Tan F, Teh KC. Examination of the performance of youth soccer players in a 20-metre shuttle run test and a treadmill run test. Adv Exerc Sports Physiol. 2005;11(3):95–101.

36 Cole TJ, Lobstein T. Extended international (IOTF) body mass index cut-offs for thinness, overweight and obesity. Pediatr Obes. 2012;7(4):284–294.

37 Cooley D, McNaughton L. Aerobic fitness of Tasmanian secondary school children using the 20-m shuttle run test. Percept Mot Skills. 1999;88(1):188–198.

38 Coombes J, Schaumberg M. Exercise capacity. In: Coombes J, Skinner T, editors. ESSA's student manual for health, exercise and sport assessment. Sydney (NSW): Elsevier; 2014. pp. 351–367.

39 Cooper KH, Pollock ML, Martin RP, White SR, Linnerud AC, Jackson A. Physical fitness levels vs selected coronary risk factors: A cross-sectional study. JAMA. 1976;236(2):166–169.

40 Corbin CB. C. H. McCloy lecture: Fifty years of advancements in fitness and activity research. Res Q Exerc Sport. 2012;83(1):1–11.

41 Council of Europe. Eurofit: Handbook for the Eurofit tests of physical fitness. Rome: Council of Europe; 1988.

42 Cureton KJ, Boileau RA, Lohman TG, Misner JE. Determinants of distance running performances in children: Analysis of a path model. Res Q. 1977;48(2):270–279.

43 De Miguel-Etayo P, Gracia-Marco L, Ortega FB, et al. Physical fitness reference standards in European children: The IDEFICS study. Int J Obes (Lond). 2014;38(S2):S57–66.

44 Dickau L. Examination of aerobic and anaerobic contributions to yo-yo intermittent recovery level 1 test performance in female adolescent soccer players [thesis]. Victoria (BC): University of Victoria; 2011.

45 Dobbins M, Husson H, DeCorby K, LaRocca RL. School-based physical activity programs for promoting physical activity and fitness in children and adolescents aged 6 to 18. Cochrane Database Syst Rev. 2013;(2):CD007651.

46 Dobosz J, Mayorga-Vega D, Viciana J. Percentile values of physical fitness levels among Polish children aged 7 to 19 years—A population-based study. Cent Eur J Public Health. 2015;23(4):340–351.

47 Domone S, Mann S, Sandercock G, Wade M, Beedie C. A method by which to assess the scalability of field-based fitness tests of cardiorespiratory fitness among schoolchildren. Sports Med. 2016;46(12):1819–1831.

48 Falgairette G, Bedu M, Fellmann N, Spielvogel H, Van Praagh E, Obert P, Coudert J. Evaluation of physical fitness from field tests at high altitude in circumpubertal boys: Comparison with laboratory data. Eur J Appl Physiol Occup Physiol. 1994;69(1):36–43.

49 Freedson PS, Cureton KJ, Heath GW. Status of field-based fitness testing in children and youth. Prev Med. 2000;31(2):S77–85.

50 Goran M, Fields DA, Hunter GR, Herd SL, Weinsier RL. Total body fat does not influence maximal aerobic capacity. Int J Obes Relat Metab Disord. 2000;24(7):841–848.

51 Gulías-González R, Sánchez-López M, Olivas-Bravo Á, Solera-Martínez M, Martínez-Vizcaíno V. Physical fitness in Spanish schoolchildren aged 6–12 years: Reference values of the battery EUROFIT and associated cardiovascular risk. J Sch Health. 2014;84(10):625–635.

52 Hallal PC, Andersen LB, Bull FC, Guthold R, Haskell W, Ekelund U; Lancet Physical Activity Series Working Group. Global physical activity levels: Surveillance progress, pitfalls, and prospects. Lancet. 2012;380(9838):247–257.

53 Hamlin MJ, Fraser M, Lizamore CA, Draper N, Shearman JP, Kimber NE. Measurement of cardiorespiratory fitness in children from two commonly used field tests after accounting for body fatness and maturity. J Hum Kinet. 2014;40:83–92.

54 Haugen T, Høigaard R, Seiler S. Normative data of BMI and physical fitness in a Norwegian sample of early adolescents. Scand J Public Health. 2014;42(1):67–73.

55 Hemmings S, Nevill A, Nevill M. Validation of the 20-m multi-stage shuttle test as predictor of peak oxygen uptake in young elite sports performers. J Sports Sci. 2003;21(4):277.

56 Hobold E, Pires-Lopes V, Gómez-Campos R, de Arruda M, Andruske CL, Pacheco-Carrillo J, Cossio-Bolaños MA. Reference standards to assess physical fitness of children and adolescents of Brazil: An approach to the students of the Lake Itaipú region—Brazil. PeerJ. 2017;5:e4032.

57 Hopkins WG. Measures of reliability in sports medicine and science. Sports Med. 2000;30(1):1–15.

58 Institute of Medicine. Fitness measures and health outcomes in youth. Washington (DC): The National Academies Press, 2012.

59 International Society for Physical Activity and Health. The Bangkok Declaration on physical activity for global health and sustainable development. Br J Sports Med. 2017;51(19):1389–1391.

60 Kamon E, Pandolf KB. Maximal aerobic power during laddermill climbing, uphill running, and cycling. J Appl Physiol. 1972;32(4):467–473.

61 Kemper HCG, Van Mechelen W. Physical fitness testing of children: A European perspective. Pediatr Exerc Sci. 1996;8(3):201–214.

62 Lang JJ, Belanger K, Poitras V, Janssen I, Tomkinson GR, Tremblay MS. Systematic review of the relationship between 20 m shuttle run performance and health indicators among children and youth. J Sci Med Sport. 2018;21(4):383–397.

63 Lang JJ, Tomkinson GR, Janssen I, Ruiz JR, Ortega FB, Léger L, Tremblay MS. Making a case for cardiorespiratory fitness surveillance among children and youth. Exerc Sport Sci Rev. 2018;46(2):66–75.

64 Lang JJ, Tremblay MS, Léger L, Olds T, Tomkinson GR. 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. Br J Sports Med. 2018;52(4):276.

65 Lang JJ, Tremblay MS, Ortega FB, Ruiz JR, Tomkinson GR. Review of criterion-referenced standards for cardiorespiratory fitness: What percentage of 1 142 026 international children and youth are apparently healthy? Br J Sports Med. Published Online First: 02 March 2017. DOI: 10.1136/bjsports-2016-096955.

66 Lang JJ, Wolfe Phillips E, Orpana HM, et al. Field-based measurement of cardiorespiratory fitness to evaluate physical activity interventions. Bull World Health Organ. 2018;96(11):794–796.

67 Larouche R, Saunders TJ, Faulkner G, Colley R, Tremblay M. Associations between active school transport and physical activity, body composition, and cardiovascular fitness: A systematic review of 68 studies. J Phys Act Health. 2014;11(1):206–227.

68 Léger L, Lambert J, Goulet A, Rowan C, Dinelle Y. Aerobic capacity of 6 to 17-year-old Quebecois—20 meter shuttle run test with 1 minute stages [Article in French]. Can J Appl Sport Sci. 1984;9(2):64–69.

69 Léger LA, Lambert J. A maximal multistage 20m shuttle run test to predict $\dot{V}O_{2max}$. Eur J Appl Physiol. 1982;49(1):1–12.

70 Léger LA, Mercier D, Gadoury C, Lambert J. The multistage 20 meter shuttle run test for aerobic fitness. J Sports Sci. 1988;6(2):93–101.

71 LeMura LM, von Duvillard SP, Cohen SL, et al. Treadmill and cycle ergometry testing in 5- to 6-year-old children. Eur J Appl Physiol. 2001;85(5):472–478.

72 Liu NY, Plowman SA, Looney MA. The reliability and validity of the 20-meter shuttle test in American students 12 to 15 years old. Res Q Exerc Sport. 1992;63(4):360–365.

73 Lobelo F, Pate RR, Dowda M, Liese AD, Ruiz JR. Validity of cardiorespiratory fitness criterion-referenced standards for adolescents. Med Sci Sports Exerc. 2009;41(6):1222–1229.

74 Lobstein T, Baur L, Uauy R, IASO International Obesity TaskForce. Obesity in children and young people: A crisis in public health. Obes Rev. 2004;5(S1):S4–104.

75 Longmuir PE, Colley RC, Wherley VA, Tremblay MS. Canadian Society for Exercise Physiology position stand: Benefit and risk for promoting childhood physical activity. Appl Physiol Nutr Metab. 2014;39(11):1271–1279.

76 Mahar MT, Crotts DJ, Mccammon MR, Rowe DA. Validity of the PWC170 and PACER tests as measures of aerobic capacity in 12- to 14-year old girls. Med Sci Sports Exerc. 2002;34(5 S1): S294.

77 Mahar MT, Guerieri AM, Hanna MS, Kemble CD. Estimation of aerobic fitness from 20-m multistage shuttle run test performance. Am J Prev Med. 2011;41(4 S2):S117–123.

78 Mahar MT, Welk GJ, Rowe DA. Estimation of aerobic fitness from PACER performance with and without body mass index. Meas Phys Educ Exerc Sci. 2018;22(3):239–249.

79 Mahar MT, Welk GJ, Rowe DA, Crotts DJ, McIver KL. Development and validation of a regression model to estimate $\dot{V}O_2$peak from PACER 20-m shuttle run performance. J Phys Act Health. 2006;3(S2):S34–46.

80 Mahoney C. 20-MST and PWC170 validity in non-Caucasian children in the UK. Br J Sports Med. 1992;26(1):45–47.

81 Mahoney CA, Boreham CAG. Validity and reliability of fitness testing in primary school children. In: Williams T, Almond L, Sparkes A, editors. Sport and physical activity: Moving towards excellence. London: E & FN Spon; 1992. pp. 429–437.

82 Massicotte D. Partial curl-ups, push ups and multistage 20 meter shuttle run, national norms for 6 to 17 year-olds. Final report, Project No. 240-0010-88/89, submitted to Canadian Association for Health, Physical Education and Recreation (CAHPER) and Fitness and Amateur Sport Canada. Montréal: University of Quebec at Montréal; 1990.

83 Matsuzaka A, Takahashi Y, Yamazoe M, Kumakura N, Ikeda A, Wilk B, Bar-Or O. Validity of the multistage 20-m shuttle-run test for Japanese children, adolescents, and adults. Pediatr Exerc Sci. 2004;16(2):113–125.

84 Mayorga-Vega D, 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(3):536–547.

85 Mayorga-Vega D, Bocanegra-Parrilla R, Ornelas M, Viciana J. Criterion-related validity of the distance- and time-based walk/run field tests for estimating cardiorespiratory fitness: A systematic review and meta-analysis. PLoS One. 2016;11(3):e0151671.

86 McIver K, Pfeiffer KA, Mahar MT, Pate RR. Associations between peak $\dot{V}O_2$ and field tests of cardiorespiratory fitness in adolescent males. Med Sci Sports Exerc. 2004;36(5):S134.

87 McVeigh SK, Payne AC, Scott S. The reliability and validity of the 20-meter shuttle test as a predictor of peak oxygen uptake in Edinburgh school children, age 13 to 14 years. Pediatr Exerc Sci. 1995;7(1):69–79.

88 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(4):285–293.

89 Midgley AW, Bentley DJ, Luttikholt H, McNaughton LR, Millet GP. Challenging a dogma of exercise physiology: Does an incremental exercise test for valid $\dot{V}O_{2max}$ determination really need to last between 8 and 12 minutes? Sports Med. 2008;38(6):441–447.

90 Mora-Gonzalez J, Cadenas-Sanchez C, Martinez-Tellez B, Sanchez-Delgado G, Ruiz JR, Léger L, Ortega FB. Estimating $\dot{V}O_{2max}$ in children aged 5–6 years through the preschool-adapted 20-m shuttle-run test (PREFIT). Eur J Appl Physiol. 2017;117(11):2295–2307.

91 Moreira C, Santos R, Ruiz JR, Vale S, Soares-Miranda L, Marques AI, Mota J. Comparison of different $\dot{V}O_{2max}$ equations in the ability to discriminate the metabolic risk in Portuguese adolescents. J Sci Med Sport 2011;14(1):79–84.

92 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. Br J Sports Med. 2011;45(11):839–848.

93 Mura G, Rocha NB, Helmich I, et al. Physical activity interventions in schools for improving lifestyle in European countries. Clin Pract Epidemiol Ment Health. 2015;11(S1 M5):S77–101.

94 Olds TS, Ridley K, Tomkinson GR. Declines in aerobic fitness: Are they only due to increasing fatness? Med Sport Sci. 2007;50:226–240.

95 Ortega FB, Artero EG, Ruiz JR, et al. Physical fitness levels among European adolescents: The HELENA study. Br J Sports Med. 2011;45(1):20–29.

96 Ortega FB, Artero EG, Ruiz JR, et al. Reliability of health-related physical fitness tests in European adolescents. The HELENA Study. Int J Obes (Lond). 2008;32(S5):S49–57.

97 Park RJ. Measurement of physical fitness: A historical perspective. ODPHP monograph series. Washington (DC): U.S. Department of Health and Human Services; Public Health Service; 1988.

98 Pate RR. The case for large-scale physical fitness testing in American youth. Ped Exerc Sci. 1989;1(4):290–294.

99 Pireva A, Selimi M, Gontarev S, Georgi G. Association between aerobic fitness and high blood pressure in adolescents in Macedonia evidence for criterion-referenced cut-points. J Phys Educ Sport. 2018;18(2):853–861.

100 Pitetti KH, Fernhall B, Figoni S. Comparing two regression formulas that predict $\dot{V}O_2$ peak using the 20-m shuttle run for children and adolescents. Ped Exerc Sci. 2002;14(2):125–134.

101 Poitras VJ, Gray CE, Borghese MM, et al. Systematic review of the relationship between objectively measured physical activity and health indicators in school-aged children and youth. Appl Physiol Nutr Metab 2016;41(6 S3):S197–239.

102 Quinart S, Mougin F, Simon-Rigaud ML, Nicolet-Guénat M, Nègre V, Regnard J. Evaluation of cardiorespiratory fitness using three field tests in obese adolescents: Validity, sensitivity and prediction of peak $\dot{V}O_2$. J Sci Med Sport. 2014;17(5):521–525.

103 Ramírez-Vélez R, Palacios-López A, Prieto-Benavides DH, Correa-Bautista JE, Izquierdo M, Alonso-Martínez A, Lobelo F. Normative reference values for the 20 m shuttle-run test in a population-based sample of school-aged youth in Bogota, Colombia: The FUPRECOL study. Am J Hum Biol. 2017;29(1):e22902.

104 Riddoch CJ. The Northern Ireland health and fitness survey—1989: The fitness, physical activity, attitudes and lifestyles of Northern Ireland post-primary schoolchildren. Belfast: The Queen's University of Belfast; 1990.

105 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. Circulation. 2016;134(24):e653–699.

106 Ruiz JR, Castro-Piñero J, Artero EG, Ortega FB, Sjöström M, Suni J, Castillo MJ. Predictive validity of health-related fitness in youth: A systematic review. Br J Sports Med. 2009;43(12):909–923.

107 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(6):518–524.

108 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 metaanalysis. Br J Sports Med. 2016;50(23):1451–1458.

109 Ruiz JR, Huybrechts I, Cuenca-García M, et al. Cardiorespiratory fitness and ideal cardiovascular health in European adolescents. Heart. 2015;101(10):766–773.

110 Ruiz JR, Ortega FB, Rizzo NS, Villa I, Hurtig-Wennlöf A, Oja L, Sjöström M. High cardiovascular fitness is associated with low metabolic risk score in children: The European Youth Heart Study. Pediatr Res. 2007;61(3):350–355.

111 Ruiz JR, Silva G, Oliveira N, Ribeiro JC, Oliveira JF, Mota J. Criterion-related validity of the 20-m shuttle run test in youths aged 13–19 years. J Sports Sci. 2009;27(9):899–906.

112 Saavedra JM, Escalante Y, Garcia-Hermoso A. Improvement of aerobic fitness in obese children: A meta-analysis. Int J Pediatr Obes. 2011;6(3–4):169–177.

113 Sandercock G, Jones B. Is it time to give population health surveillance a late fitness test? Br J Sports Med. Published Online First: 02 February 2018. DOI: 10.1136/bjsports-2017-098659.

114 Sandercock G, Voss C, Cohen D, Taylor M, Stasinopoulos DM. Centile curves and normative values for the twenty metre shuttle-run test in English schoolchildren. J Sports Sci. 2012;30(7):679–687.

115 Santos R, Mota J, Santos DA, Silva AM, Baptista F, Sardinha LB. Physical fitness percentiles for Portuguese children and adolescents aged 10–18 years. J Sports Sci. 2014;32(16):1510–1518.

116 Sauka M, Priedite IS, Artjuhova L, Larins V, Selga G, Dahlström O, Timpka T. Physical fitness in northern European youth: Reference values from the Latvian Physical Health in Youth Study. Scand J Public Health. 2011;39(1):35–43.

117 Secchi JD, García GC, España-Romero V, Castro-Piñero J. Physical fitness and future cardiovascular risk in argentine children and adolescents: An introduction to the ALPHA test battery. Arch Argent Pediatr. 2014;112(2):132–140.

118 Silva DA, Tremblay M, Pelegrini A, Dos Santos Silva RJ, Cabral de Oliveira AC, Petroski EL. Association between aerobic fitness and high blood pressure in adolescents in Brazil: evidence for criterion-referenced cut-points. Pediatr Exerc Sci. 2016;28(2):312–320.

119 Silva DAS, Lang JJ, Barnes JD, Tomkinson GR, Tremblay MS. Cardiorespiratory fitness in children: Evidence for criterion-referenced cut-points. PLoS ONE. 2018;13(8):e0201048.

120 Silva G, Aires L, Mota J, Oliveira J, Ribeiro JC. Normative and criterion-related standards for shuttle run performance in youth. Pediatr Exerc Sci. 2012;24(2):157–169.

121 Suminski RR, Ryan ND, Poston CS, Jackson AS. Measuring aerobic fitness of Hispanic youth 10 to 12 years of age. Int J Sports Med. 2004;25(1):61–67.

122 Sun C, Pezic A, Tikellis G, et al. Effects of school-based interventions for direct delivery of physical activity on fitness and cardiometabolic markers in children and adolescents: A systematic review of randomized controlled trials. Obes Rev. 2013;14(10):818–838.

123 Tambalis KD, Panagiotakos DB, Arnaoutis G, Sidossis LS. Endurance, explosive power, and muscle strength in relation to body mass index and physical fitness in Greek children aged 7–10 years. Pediatr Exerc Sci. 2013;25(3):394–406.

124 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 Sport Sci. 2016;16(6):736–746.

125 The Cooper Institute. FitnessGram administration manual: The journey to MyHealthyZone. 5th ed. Champaign (IL): Human Kinetics; 2017.

126 Tomkinson GR, Carver KD, Atkinson F, et al. European normative values for physical fitness in children and adolescents aged 9–17 years: Results from 2 779 165 Eurofit performances representing 30 countries. Br J Sports Med. 2018;52(22):1445–1456.

127 Tomkinson GR, Lang JJ, Tremblay MS. Temporal trends in the cardiorespiratory fitness of children and adolescents representing 19 high- and upper-middle-income countries between 1981 and 2014. Br J Sports Med. Published Online First: 30 October 2017. DOI: 10.1136/bjsports-2017-097982.

128 Tomkinson GR, Lang JJ, Tremblay MS, et al. International normative 20 m shuttle run values from 1,142,026 children and youth representing 50 countries. Br J Sports Med. 2017;51(21):1545–1554.

129 Tomkinson GR, Léger LA, Olds TS, Cazorla G. Secular trends in the performance of children and adolescents (1980–2000): An analysis of 55 studies of the 20 m shuttle run in 11 countries. Sports Med. 2003;33(4):285–300.

130 Tremblay MS, Barnes JD, González SA, Katzmarzyk PT, Onywera VO, Reilly JJ, Tomkinson GR, Global Matrix 2.0 Research Team. Global Matrix 2.0: Report card grades

on the physical activity of children and youth comparing 38 countries. J Phys Act Health. 2016;13(11 S2):S343–366.

131 Tsiaras V, Zafeiridis A, Dipla K, Patras K, Georgoulis A, Kellis S. Prediction of peak oxygen uptake from a maximal treadmill test in 12-to 18-year-old active male adolescents. Pediatr Exerc Sci. 2010;22(4):624–637.

132 Turley KR, Wilmore JH. Cardiovascular responses to treadmill and cycle ergometer exercise in children and adults. J Appl Physiol. 1997;83(3):948–957.

133 United Nations. Political declaration of the high-level meeting of the general assembly on the prevention and control of non-communicable diseases. New York: United Nations, Department of Economic and Social Affairs; 2012.

134 United Nations. Transforming our world: The 2030 agenda for sustainable development. New York: United Nations, Department of Economic and Social Affairs; 2015.

135 U.S. Department of Health and Human Services, 2018 Physical Activity Guidelines Advisory Committee. 2018 Physical Activity Guidelines Advisory Committee Scientific Report. Washington, DC: U.S. Department of Health and Human Services, 2018.

136 Vanhelst J, Labreuche J, Béghin L, et al. Physical fitness reference standards in French youth: The BOUGE program. J Strength Cond Res. 2017;31(6):1709–1718.

137 van Mechelen W, Hlobil H, Kemper HC. Validation of two running tests as estimates of maximal aerobic power in children. Eur J Appl Physiol Occup Physiol. 1986;55(5):503–506.

138 Van Praagh E, Bedu M, Falgairette G, Fellmann N, Coudert J. Comparison of direct and indirect $\dot{V}O_2$ max in 7- and 12-year-old children: A field test validation study [Article in French]. Sci Sports. 1988;3(4):327–332.

139 Van Praagh E, Falgairette G, Bedu M, Fellmann N, Coudert J. Laboratory and field tests in 7-year-old boys. In: Oseid S, Carlsen K-H, editors. Children and exercise XIII. Champaign (IL): Human Kinetics; 1989.

140 Varness T, Carrel AL, Eickhoff JC, Allen DB. Reliable prediction of insulin resistance by a school-based fitness test in middle-school children. Int J Pediatr Endocrinol. 2009; 2009(1):487804.

141 Verschuren O, Takken T, Ketelaar M, Gorter JW, Helders PJ. Reliability and validity of data for 2 newly developed shuttle run tests in children with cerebral palsy. Phys Ther. 2006;86(8):1107–1117.

142 Verschuren O, Zwinkels M, Ketelaar M, Reijnders-van Son F, Takken T. Reproducibility and validity of the 10-meter shuttle ride test in wheelchair-using children and adolescents with cerebral palsy. Phys Ther. 2013;93(7):967–974.

143 Voss C, Sandercock G. Does the twenty meter shuttle-run test elicit maximal effort in 11- to 16-year-olds? Pediatr Exerc Sci. 2009;21(1):55–62.

144 Welk GJ, Laurson KR, Eisenmann JC, Cureton KJ. Development of youth aerobic capacity standards using receiver operating characteristic curves. Am J Prev Med. 2011;41(4 S2):S111–116.

145 Wilkinson RG, Pickett KE. Income inequality and population health: A review and explanation of the evidence. Soc Sci Med. 2006;62(7):1768–1784.

146 World Health Organization. Global action plan on physical activity 2018–2030: More active people for a healthier world. Geneva: World Health Organization; 2018.

147 World Health Organization. Global recommendations on physical activity for health. Geneva: World Health Organization; 2010.

148 World Health Organization. Noncommunicable diseases progress monitor, 2017. Geneva: World Health Organization; 2017.

149 Zamparo P, Zadro I, Lazzer S, Beato M, Sepulcri L. Energetics of shuttle runs: the effects of distance and change of direction. Int J Sports Physiol Perform. 2014;9(6):1033–1039.

150 Zeng N, Gao Z. Exergaming and obesity in youth: current perspectives. Int J Gen Med. 2016;9:275–284.

Tables

Table 1. Normative-referenced standards for the 20mSRT in apparently healthy children and youth.

Region/Country	Citation	Sample size	Year of testing	Age range (years)	Sex (male/female ratio)	20mSRT protocol
International	Tomkinson et al. (128)	1,142,026	1984–2015	9–17	48.4/51.6	Léger
Regional						
Europe	De Miguel-Etayo et al. (43)	7,862	2007–08	6–11	49.3/50.7	Léger
Europe	Ortega et al. (95)	3,428	2006–08	12–18	46.2/53.8	Léger
Europe	Tomkinson et al. (126)	445,092	1981–2016	9–17	53.2/46.8	Eurofit
National						
Argentina	Secchi et al. (117)	1,867	2012	6–19	48.2/51.8	Léger
Australia	Catley et al. (30)	18,075	1990–2009	9–17	50.6/49.4	Léger
Australia (Tasmania)	Cooley and McNaughton (37)	6,061	1998	11–16	49.2/50.8	Eurofit
Brazil	Hobold et al. (56)	5,962	2012	6–18	49.3/50.7	Léger
Canada	Massicotte (82)	6,644	1989–90	6–17	55.2/44.8	Léger
Colombia	Ramírez-Vélez et al. (102)	7,244	2014–15	9–18	44.3/55.7	Léger
England (East England)	Sandercock et al. (114)	7,366	2006–10	10–16	52.9/47.1	Léger
France	Vanhelst et al. (136)	11,186	2009–13	10–15	49.6/50.4	Léger
Greece	Tambalis et al. (123)	141,169	1997	7–10	51.1/48.9	Léger
Greece	Tambalis et al. (124)	424,328	2014	6–18	51.0/49.0	Léger
Latvia	Sauka et al. (116)	10,464	2004–09	6–17	52.6/47.4	Léger
Norway	Haugen et al. (54)	1,059	2004–06	13–15	50.0/50.0	Léger
Poland	Dobosz et al. (46)	47,682	2009–10	7–19	52.3/47.7	Eurofit
Portugal	Santos et al. (115)	22,048	2008	10–18	49.4/51.6	Léger
Portugal	Silva et al. (120)	5,559	2008–09	10–18	50.2/49.8	PACER
Spain	Cadenas-Sanchez et al. (26)	3,179	2014–15	2–6	52.8/47.2	PREFIT
Spain	Castro-Piñeiro et al. (29)	2,752	2006–07	6–18	54.2/45.8	Léger
Spain	Gulias-Gonzalez et al. (51)	1,725	2010	6–12	50.3/49.7	Eurofit
United States (Wisconsin)	Carrel et al. (28)	20,631	2008–10	8–18	51.2/48.8	PACER

Note, see Section 4 for protocol descriptions.

Figure captions

Figure 1. Sample-weighted quadratic curve showing the age-related changes in criterion validity for gas analyzed peak $\dot{V}O_2$ (mL/kg/min) from 20mSRT performances only.

Note, black dots indicate data for boys, dark grey dots indicate data for girls, and light grey dots indicate data for both boys and girls. Validity data for children and youth are from 3,5,7,11,12,14,20,32,33,34,35,44,48,53,55,70,72,76,77,79,80,81,83,86,87,104,131,137,138,139,140.

Figure 2. Sample-weighted Lowess curves (tension=66) showing the age-related changes in health-related criterion-referenced standards for CRF (peak $\dot{V}O_2$ in mL/kg/min).

Note, black dots and curve indicate data for boys, and the light grey dots and curve indicate data for girls.