8-8-2017

Systematic review of the relationship between 20m shuttle run performance and health indicators among children and youth

Justin J. Lang

Kevin Belanger

Veronica Poitras

Ian Janssen

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

See next page for additional authors

Systematic review of the relationship between 20 m shuttle run performance and health indicators among children and youth

Justin J. Langa, Kevin Belanger, Veronica Poitras, Ian Janssen, Grant R.Tomkinson, Mark S. Tremblay

Abstract

Objective

This systematic review aimed to summarize research that assessed the associations between 20 m shuttle run test (20mSRT) performance and indicators of physiological, psychosocial and cognitive health among school-aged children and youth.

Design

Systematic review.

Methods

Five online databases were used to identify peer-reviewed studies published from 1980 to 2016. Studies were included if they matched these criteria: population (children and youth with a mean age of 5–17 years and/or in Grades 1–12), intervention/exposure (performance on the 20mSRT), and outcomes (health indicators: adiposity, cardiometabolic biomarkers, cognition, mental health, psychosocial health, self-esteem and physical self-perception, quality of life and wellbeing, bone health, musculoskeletal fitness, motor skill development, and injuries and/or harm). Narrative syntheses were applied to describe the results. A lack of homogeneity precluded a meta-analysis approach.

Results

Overall, 142 studies that determined an association between 20mSRT performance and a health indicator were identified, representing 319,311 children and youth from 32 countries. 20mSRT performance was favourably associated with indicators of adiposity, and some indicators of cardiometabolic, cognitive, and psychosocial health in boys and girls. Fewer studies examined the relationship between 20mSRT performance and measures of quality of life/wellbeing, mental health and motor skill development, and associations were generally inconsistent. The quality of the evidence ranged from very low to moderate across health indicators.

Conclusion and Implications

These findings support the use of the 20mSRT as a holistic indicator of population health in children and youth.

1. Introduction

Cardiorespiratory fitness (CRF) is measured as the highest rate of oxygen consumed ($\dot{V}O_2$) during a bout of maximal exercise to voluntary exhaustion, described as $\dot{V}O_{2peak}$ in children and youth.[1] CRF provides an indication of the function of the cardiovascular and respiratory systems and it provides a strong summative measure of physical health among children and youth.[2] Furthermore, youth with low CRF in late adolescence have an increased risk of developing cardiovascular disease in adulthood[3] and succumbing to premature mortality.[4] For these reasons, CRF may be a strong population health indicator and surrogate measure that could help researchers and public health officials understand the health of children and youth and how health varies across these populations. Objective measures of CRF can be obtained in lab-based settings by collecting expired gases during a graded exercise test. Limitations of this approach include the high cost, time intensiveness, laboratory infrastructure requirements, unfamiliar equipment, and unnatural environment yielding heavy participant burden, making these tests not suitable for mass testing. To help address these limitations, the 20 m shuttle run test (20mSRT)[5] was developed as a field-based measure to estimate CRF.

The 20mSRT is a progressive CRF test where children and youth are asked to run laps back and forth between two parallel lines 20 m apart. An audio recording paces the participants beginning at a speed of 8.5 km/h and increasing by 0.5 km/h every consecutive minute. Participants continue until they are no longer able to keep pace with the audio recording for two consecutive laps, at which point their score is recorded (i.e., total laps, stages, time, etc.).[5] The 20mSRT is the most widely used field-based assessment of CRF, with recent research indicating that at least 50 countries from six continents have used the assessment.[6, 7] This popularity is likely due to the low cost of equipment, simplicity in administering and scoring the test, flexibility in testing (e.g., surface, space, location, etc.), low participant burden and ability to simultaneously test large groups of children and youth.[8] These characteristics allow the 20mSRT to overcome many cultural and language barriers that limit other fitness assessments. Moreover, the 20mSRT demonstrates strong test-retest reliability[8] and moderate-to-strong validity[9] which suggests that the 20mSRT is a good measure of CRF. Although recent systematic reviews have explored the associations between physical fitness, physical activity and health among children and youth,[1, 10, 11] no study has comprehensively assessed the associations between 20mSRT performance and health indicators among children and youth.

Thus, this systematic review aimed to summarize and evaluate available research that assessed the association between 20mSRT performance and health indicators (e.g., physiological, psychosocial, cognitive) among children and youth aged 5–17 years (or in Grades 1–12).

2. Methods

The systematic review protocol is registered with the International Prospective Register of Systematic Reviews (PROSPERO; registration no. CRD42015032250: available from http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42015032250). This study was conducted and reported in accordance with the PRISMA 2009 statement for reporting systematic reviews.[12]

The participants, intervention, comparison, outcome and study designs (PICOS) framework[13] was used to help delineate the study parameters of our research question, and to develop the search strategy.

School-aged participants who were between the mean ages of 5–17 years and/or between grades 1–12, if in the North America (or equivalent grades in other countries), were included in this study.

Performance results for the 20mSRT protocol beginning at 8.5 km/h and increasing by 0.5 km/h every minute was our main exposure. Results reported as raw scores (e.g., speed, laps, stages, etc.) and/or predicted relative $\dot{V}O_{2peak}$ were included.

In certain cases 20mSRT performance can be dichotomized into healthy and unhealthy CRF using criterion-referenced standards. In these cases comparisons for health indicators would take place between the different CRF groups. Otherwise, the relationship between 20mSRT performance and continuous health indicators were assessed according to comparisons or relationships provided by the authors.

Health indicators for this study were selected based on a recent physical activity systematic review,[11] with additional health indicators being added based on the literature (i.e., previous research). Eleven health indicators were included that broadly represented three categories: physiological, psychosocial and cognitive health. Health indicators included: (i) adiposity (e.g., body mass index, waist circumference), (ii) cardiometabolic biomarkers (e.g., lipid profile, insulin resistance, blood pressure), (iii) cognition/academic achievement (e.g., concentration, memory, school grades), (iv) mental health (e.g., depression, anxiety, risk of mental illness), (v) psychosocial health (e.g., behaviour problems, self-efficacy, social development), (vi) self-esteem and physical self-perception (e.g., perceived competence), (vii) quality of life and wellbeing, (viii) bone health (e.g., bone mineral density, bone area), (ix) musculoskeletal fitness (e.g., muscular strength, muscular endurance, flexibility), (x) motor skill development (e.g., gross/fine motor skills, coordination), and (xi) injuries and/or harm (e.g., adverse events).

No restrictions were placed on study design. A minimum sample size for all study designs was set at 40 participants. Studies were screened if they were written in English only. Peer-reviewed original research articles that were published or in-press were eligible for inclusion. All grey literature, abstracts, reviews, theses and conference proceedings were excluded from this review. One additional variant of the 20mSRT protocol was also accepted (i.e., the Eurofit 20mSRT protocol begins at 8.0 km/h, increases to 9.0 km/h for the second minute, and then increases by 0.5 km/h every minute thereafter[14]). No restrictions were placed on the included health indicators (e.g., validity and reliability criteria).

A systematic search of the literature was conducted in MEDLINE (OVID interface; 1980 to March 2nd, 2016), EMBASE (OVID interface; 1980 to March 3rd, 2016), PsychINFO (OVID interface; 1980 to March 4th, 2016), Cochrane Central Register of Controlled Trials (CENTRAL, OVID interface; 1980 to March 4th, 2016) and SPORTDiscus (EBSCO interface; 1980 to March 2nd, 2016). The search strategy was developed by a librarian experienced in systematic review searching and peer-reviewed by another librarian using the PRESS standard.[15] Search strategies are presented in Supplement 1.

Database specific search strategies were executed by a single researcher through the OVID and EBSCOhost interfaces. All bibliographic records were extracted as text files and imported into Reference Manager Software (version 11; Thompson Reuters, San Francisco, CA, USA) where duplicate files were manually removed. Titles and abstracts for the remaining relevant articles were imported into DistillerSR (Evidence Partners; Ottawa, ON, Canada). All potentially relevant titles and abstracts were screened against the inclusion criteria by two reviewers. Exclusion by both reviewers was required to exclude records at level 1. The remaining records passed to level 2 where full-text articles were screened against the inclusion criteria independently by two reviewers. Consensus by both reviewers was required at level 2 to include records. All discrepancies were resolved between reviewers or by a third reviewer, if necessary. Furthermore, the reference lists of included articles and relevant reviews in the area were checked for additional studies not identified through the systematic search.

A study specific database was created in Excel (Microsoft Corp.; Redmond, Wash., USA) and piloted by reviewers. Data were extracted from included studies into the database by one reviewer, and subsequently checked for accuracy by a second reviewer.

Study specific descriptive characteristics were extracted (i.e., author, study year, study country, study design, sample size, participant age and sex), 20mSRT type (i.e., result units, equation used if predicted $\dot{V}O_{2peak}$ was calculated, CRF standard, etc.), health indicator (including measurement method and units), results (i.e., correlation coefficient, odds ratio, p-value, chi-square values, 95% CIs) and confounding variables (included in multivariate analyses). When studies presented multiple models, only the results from the fully adjusted model were extracted. Statistically significant findings were set at $p < 0.05$. Furthermore, correlations <0.1 were considered as trivial and correlations of 0.1, 0.3 and 0.5 were used as thresholds for weak, moderate and strong, respectively.

The risk of bias across all included studies was evaluated using the methods described in the Cochrane Handbook.[16] A modified version of the Cochrane Handbook was used to assess risk of bias in observational studies. Each study was assessed for selection bias (e.g., failure to use an appropriate sampling technique), performance bias (e.g., inappropriate measurement of exposure) and selective reporting bias (e.g., selective/incomplete reporting of results); detection (e.g., inappropriate measurement of outcomes) and attrition (e.g., amount and treatment of incomplete data) bias for each health indicator; and other sources of bias (e.g., inadequate control for important confounders).

The Grading of Recommendations Assessment, Development, and Evaluation (GRADE) framework[17] was used to assess the quality of evidence for each health indicator category across each study design. The GRADE framework categorizes the quality of evidence into 4 groups (high, moderate, low or very low). According to GRADE, quality of evidence is high for randomized controlled trials (RCTs) and low for all other study designs. Quality of evidence is downgraded if there is a serious risk of bias, imprecision (e.g., insufficiently precise findings to support a recommendation),[18] inconsistency (e.g., differences in the direction of the effects not explained by methodological differences),[19] indirectness (e.g., whether studies included the population, assessment, or outcomes of interest),[20] or other factors. If there is no justification for downgrading, quality of evidence can be upgraded if there is a large effect size, a dose-response gradient, or if all plausible confounders would decrease an apparent treatment effect.[21] Overall, quality of evidence was assessed independently by two reviewers and subsequently verified by the larger co-author team.

Meta-analyses were planned for sufficiently homogeneous health indicators with regard to statistical and methodological characteristics. Narrative syntheses of results were planned for health indicators that were considered to be too heterogeneous.

3. Results

A total of 11,745 records were identified through database searches. After de-duplication, 2,447 records remained. After screening titles and abstracts, 470 full-text articles were obtained for further review. A total of 328 studies were excluded for the reasons described in Fig. 1. Full consensus was attained between both reviewers for all included studies.

The remaining 142 studies met all inclusion criteria and represented 319,311 participants (301,105 from unique samples) from 32 different countries. Study-specific sample sizes ranged from 51 [S138] to 141,169 [S35] participants, and spanned a mean age from 5.1 to 17.9 years. Included studies employed either a longitudinal cohort (n = 14) or cross-sectional (n = 128) design. Individual study characteristics are summarized in Supplement 2, and Tables 1–11.

In total, 89 studies (63% of included studies) investigated the relationship between 20mSRT performance and measures of adiposity (Table 1, Supplement Table 1). Six studies used a longitudinal cohort study design, 82 studies used a cross-sectional study design, and the remaining study used a mixed longitudinal cohort and cross-sectional design. Five of seven longitudinal cohort studies reported that better 20mSRT performance at baseline was associated with lower adiposity levels at 1–9 years of follow-up [S1–S3, S6, S7], while two longitudinal cohort studies reported mixed favourable and null associations for girls [S4] and boys [S5]. The quality of the evidence for longitudinal cohort studies remained rated as low as there were no reasons to downgrade or upgrade the quality of the evidence. A total of 66 cross-sectional studies (out of 83) reported favourable relationships between 20mSRT performance and adiposity, 15 studies reported mixed favourable and null findings, and two studies reported mixed favourable and unfavourable findings (Table 1) [S16, S41]. Of the 89 included studies, 41 studies assessed adiposity using only BMI, 25 of the studies used various measures, including waist circumference, percent body fat, and sum of skinfolds, and the remaining 23 studies used a combination of BMI and other measures to assess adiposity. The quality of the evidence was upgraded from low to moderate because of a large effect and evidence of a dose-response relationship.

A total of 42 studies (30% of included studies) investigated the relationship between 20mSRT performance and various cardiometabolic biomarkers (e.g., HOMA-IR, cholesterol, C-reactive protein) in children and youth (Table 2, Supplement Table 2). One study used a longitudinal cohort design, 39 studies used a cross-sectional design, and the remaining two studies used mixed longitudinal cohort and cross-sectional designs. One longitudinal cohort study reported mixed null and favourable associations between 20mSRT performance and cardiometabolic biomarkers (HOMA-IR) at 2 years of follow-up [S91], and the remaining two longitudinal cohort studies reported null findings at follow-up [S90, S92]. The quality of the evidence for longitudinal cohort studies remained rated as low. Of the 41 cross-sectional studies, 13 reported favourable findings, 21 reported mixed favourable and null findings, 6 reported null findings, and one study [S97] reported a mixed null and unfavourable findings between 20mSRT performance and cardiometabolic biomarkers for boys only. The quality of the evidence for cross-sectional studies was also rated as low.

Fifteen studies (11% of included studies) investigated the associations between 20mSRT performance and cognition/academic achievement (Table 3, Supplement Table 3). Only one study used a longitudinal cohort design, and the remaining 14 studies used a cross-sectional design. The longitudinal cohort study reported that 20mSRT performance was favourably associated with attention, but null associations were reported for spatial working memory at 9 months of follow-up [S114]. The quality of the evidence for the longitudinal cohort study was downgraded from low to very low due to serious imprecision (insufficient evidence). Three cross-sectional studies assessed the relationship between 20mSRT performance and cognition, and reported favourable or mixed favourable and null findings [S9, S33, S74]. The remaining 11 cross-sectional studies investigated the association between 20mSRT performance and academic achievement. Six studies reported favourable associations [S9, S42, S116, S119–S121], four studies reported mixed favourable and null findings [S115, S117, S118, S122], one study reported null findings [S37], and one study reported mixed unfavourable and null findings [S123]. The quality of the evidence for cross-sectional studies was downgraded from low to very low due to a serious risk of bias. Specifically, nine studies had unknown reliability and validity for the outcome measures, and four studies did not control for sex or age differences in their analyses.

Two cross-sectional studies (1% of included studies) investigated the association between 20mSRT performance and mental health (Table 4, Supplement Table 4). One study reported no association between 20mSRT performance and depression in children [S124]. Another study reported a favourable association between 20mSRT performance and levels of depression in girls, but reported no association in boys [S125]. The quality of the evidence remained rated as low.

Five cross-sectional studies (4% of included studies) investigated the associations between 20mSRT performance and psychosocial health (Table 5, Supplement Table 5). Four of five studies reported favourable associations between 20mSRT performance and self-efficacy [S126–S128], enjoyment [S13], and outcome expectancy [S128]. One study reported favourable associations between 20mSRT performance and mood/emotions in girls, but the associations were null in boys [S129]. The majority of studies did not adequately control for differences in sex or age within their analyses [S13, S126–S128], and one study did not describe the sampling methods employed [S129], resulting in a risk of bias. As a result, the quality of the evidence was downgraded from low to very low.

Table 1
Association between 20mSRT performance and adiposity.

No. of studies	Design	Quality assessment					No. of participants	Absolute effect	Quality
		Risk of bias	Inconsistency	Indirectness	Imprecision	Other			
The range of mean ages was 5.1–17.4 years; 37 studies reported an age range in lieu of the mean, and in these cases the included ages ranged from 5 to 18 years. Data were collected cross-sectionally and up to 9 years of follow-up. Adiposity was assessed as body mass index (absolute, z-score or percentile), waist circumference, waist-to-hip ratio, waist-to-height ratio, hip circumference, % body fat (bioelectrical impedance, dual-energy X-ray absorptiometry, skinfolds) or Marshal Visual Rating Scale									
7	Longitudinal[a]	No serious risk of bias	No serious inconsistency	No serious indirectness	No serious imprecision	None	18,396	BMI (5 studies) 5 of 7 studies reported favourable associations between 20mSRT performance and **BMI** at 1 to 6.1 years of follow-up 2 studies also reported null findings between 20mSRT performance and **BMI** for girls but not boys at 1 year [S5] and boys but not girls at 18 months [S4] of follow-up Other measures (2 studies) 2 studies reported favourable associations between 20mSRT performance and **SSF** [S2] and **WC** [S3] at 5–9 years of follow-up	Low[b]
83	Cross-sectional[c]	No serious risk of bias	No serious inconsistency	No serious indirectness	No serious imprecision	Evidence of a large effect and a dose-response gradient. Example, better 20mSRT performance associated with lower adiposity[d]	220,971	BMI only (37 studies) 32 studies reported favourable relationships between 20mSRT performance and **BMI** 3 studies reported mixed favourable and null associations between 20mSRT performance and **BMI**. 2 studies reported null associations between 20mSRT performance and **BMI** Other measures (23 studies) 20 studies reported favourable relationships between 20mSRT and other measures of **adiposity** 2 studies reported mixed favourable and null associations between 20mSRT performance and other measures of **adiposity** 1 study reported null associations between 20mSRT performance and **%BF** BMI and other measures (23 studies) 14 studies reported favourable associations between 20mSRT performance and various measures of **adiposity** 7 studies reported mixed favourable and null associations between 20mSRT and various measures of **adiposity** 2 studies reported mixed favourable and unfavourable associations between 20mSRT performance and **adiposity** Due to heterogeneity in the measurement of 20mSRT performance and adiposity a meta-analysis was not possible	Moderate[f]

*Note 20mSRT, 20mshuttleruntest; BMI, bodymassindex; SSF, sumofskinfolds; WC, waistcircumference; %BF, percentbodyfat.

a. Includes 7 longitudinal studies [S1S7].

b. The quality of the evidence remained as "low" as there were no serious concerns about the quality of included longitudinal studies or reasons to increase the rating.

c. Includes 83 cross-sectional studies [S5, S8–S89].

d. The general favourable gradient and large effect of the findings (e.g., approximately 60% of reported unstandardized correlation coefficients were greater than r=0.30) suggests that better 20mSRT performance is associated with lower adiposity.

f. The quality of the evidence was upgraded from "low" to "moderate" because of the large effect observed and the evidence of adose-response gradient between 20mSRT performance and adiposity.

Table 2
Association between 20mSRT performance and cardiometabolic biomarkers.

No. of studies	Design	Quality assessment					No. of participants	Absolute effect	Quality
		Risk of bias	Inconsistency	Indirectness	Imprecision	Other			
The range of mean ages was 9–17.9 years; 13 studies reported an age range in lieu of the mean, and in these cases the included ages ranged from 7 to 18 years. Data were collected cross-sectionally and up to 4 years of follow-up. Lipid, Glucose/Insulin and inflammatory indicators were measured via fasting blood samples. Blood pressure was measured either manually or with an automated sphygmomanometer. Cardiometabolic risk score was calculated according to published methods using varying combinations of pertinent health indicators[a]									
3	Longitudinal[b]	No serious risk of bias	No serious inconsistency	No serious indirectness	No serious imprecision	None	1,115	Glucose, insulin, HOMA-IR 1 study reported mixed favourable and null associations between 20mSRT performance and **HOMA-IR** at 2 years follow-up [S91] 2 studies reported no association between 20mSRT performance and **HOMA-IR** at 9 months to 4 years follow-up [S90,S92] Inflammatory markers 1 study reported null findings between 20mSRT performance and **CRP** at 9 months follow-up [S90]	Low[c]
41	Cross-sectional[d]	No serious risk of bias	No serious inconsistency	No serious indirectness	No serious imprecision	None	52,102	Blood pressure (16 studies) 3 studies reported favourable associations between 20mSRT performance and **BP**. 4 studies reported mixed favourable and null associations between 20mSRT performance and **BP**. 9 studies reported null findings between 20mSRT performance and **BP** Blood lipids (TG, TC, HDL, LDL; 13 studies) 4 studies reported favourable associations between 20mSRT performance and **blood lipids** 7 studies reported mixed favourable and null associations between 20mSRT performance and **blood lipids** 2 studies reported null findings between 20mSRT performance and **blood lipids** Glucose, insulin, HOMA-IR (15 studies) 7 studies reported favourable associations between 20mSRT performance and **fasting glucose**, **fasting insulin**, and/or **HOMA-IR** levels 6 studies reported mixed favourable and null associations between 20mSRT performance and **fasting glucose**, **fasting insulin**, and/or **HOMA-IR** levels 2 studies reported null findings between 20mSRT performance and **fasting glucose**, **fasting insulin**, and/or **HOMA-IR** levels Inflammatory Markers (CRP, homocysteine, interleukin-6; 13 studies) 5 studies reported favourable associations between 20mSRT performance and **inflammatory markers** 5 studies reported mixed favourable and null associations between 20mSRT performance and **inflammatory markers** 3 studies reported null associations between 20mSRT performance and **inflammatory markers** Cardiometabolic risk (14 studies) 10 studies reported favourable associations between 20mSRT performance and **cardiometabolic risk** 3 studies reported mixed favourable and null associations between 20mSRT performance and **cardiometabolic risk** 1 study reported null findings between 20mSRT performance and **cardiometabolic risk** Due to heterogeneity in the measurement of 20mSRT performance and cardiometabolic biomarkers a meta-analysis was not possible	Low[e]

*Note 20mSRT, 20 m shuttle run test; BP, blood pressure; CRP, c-reactive protein; HDL, high density lipoprotein; HOMA-IR, homeostatic model assessment – insulin resistance; LDL, low density lipoprotein; MAP, mean arterial pressure; TC, total cholesterol; TG, triglycerides.

a. Different combinations of the following health indicators were used to calculate a cardiometabolic risk score: waist circumference, BMI z-score, HDL, LDL, glucose, insulin, HOMA-IR, MAP, triglycerides, blood pressure, total cholesterol: HDL ratio, c-reactive protein.

b. Includes 3 studies [S90–S92].

c. The quality of the evidence remained as "low" as there were no serious concerns about the quality of included longitudinal studies or reasons to increase the rating.

d. Includes 41 studies [S3, S11, S14, S26, S32, S34, S36, S38, S40, S45, S46, S52, S56, S63, S64, S67, S69, S71, S90, S91, S93–S113].

e. The quality of the evidence remained as "low" as there were no serious concerns about the quality of included cross-sectional studies or reasons to increase the rating.

Table 3
Association between 20mSRT performance and cognition/academic achievement.

No. of studies	Design	Quality assessment					No. of participants	Absolute effect	Quality
		Risk of bias	Inconsistency	Indirectness	Imprecision	Other			
The range of mean ages was 5.2 and 17.3 years; 5 studies reported an age range in lieu of the mean, and in these cases the included ages ranged from 8 to 17 years. Data were collected cross-sectionally and up to 9 months of follow-up. Cognition was measured using computer testing modalities and other tests or questionnaires: FitSmart, Intelligence and Development Scales, KHV-VK test, Flanker task, spatial n-back task, Stroop Colour and Word test, and Kaufman Brief Intelligence Test. Academic achievement was measured using student grade averages across various subjects, performance on standardized tests, WESTEST, Weschsler Individual Achievement Test, or the HSBC questionnaire									
1	Longitudinal[a]	No serious risk of bias	No serious inconsistency	No serious indirectness	Serious imprecision[b]	None	245	Niederer et al. found that 20mSRT performance at baseline was favourably associated with **attention** ($\beta = 0.25$, 95%CI 0.03–0.48, $p = 0.03$) but not associated with **spatial working memory** ($\beta = 0.08$, 95%CI −0.09 to 0.24, $p = 0.37$) at 9 months follow-up [S114]	Very low[c]
14	Cross-sectional[d]	Serious risk of bias[e]	No serious inconsistency	No serious indirectness	No serious imprecision	None	40,743	Cognition (3 studies) Williams et al. found favourable associations between 20mSRT performance **and health knowledge** ($\beta = 0.56$, $p < 0.001$) [S9] Scudder et al. found that children's 20mSRT performance was favourably associated with **working memory**, **reaction time** and **accuracy** but not **incongruent accuracy** during tasks of inhibitory control [S33] Buck et al. found that 20mSRT performance was favourably related to word, colour, and colour/word **concentration tasks**, but not interference concentration tasks. Performance on the 20mSRT was not related to IQ in children [S74] Academic achievement (11 studies) 5 studies found favourable associations between 20mSRT performance and measures of **overall academic achievement**, **English language arts**, **math**, **science**, **social studies**, **reading** or **spelling** courses [S42, S116, S119–S121] 4 studies reported mixed favourable associations between 20mSRT performance and **numeracy**, **writing** and **spelling**, but no association with **reading** skills [S115, S117, S118, S122] 1 study reported null associations between 20mSRT performance and **math**, **English** and **social study** grades [S37] 1 study found mixed null associations between 20mSRT performance and **academic achievement** in children, but unfavourable associations in adolescents [S123] Due to heterogeneity in the measurement of 20mSRT performance and cognition a meta-analysis was not possible	Very low[f]

*Note 20mSRT, 20 m shuttle run test; β indicates standardized beta; HSBC, Health Behaviour in School-Aged Children.

a. Includes 1 longitudinal study [S114].

b. Serious imprecision. Only 1 study investigated the longitudinal associations between 20mSRT performance and cognition.

c. The quality of the evidence was downgraded from "low" to "very low" because of serious imprecision.

d. Includes 14 cross-sectional studies [S9, S33, S37, S42, S74, S115–S123].

e. Serious risk of bias. A total of 10 out of 14 studies did not provide detail on the validity and reliability of outcome measures [S9, S37, S42, S115, S116, S119–S123]. Four studies did not control for sex or age in their analysis [S37, S42, S74, S116].

f. The quality of the evidence was downgraded from "low" to "very low" because of a serious risk of bias that diminished the level of confidence in the observed effects.

Table 4
Association between 20mSRT performance and mental health.

No. of studies	Design	Quality assessment					No. of participants	Absolute effect	Quality
		Risk of bias	Inconsistency	Indirectness	Imprecision	Other			
The range of mean ages was 12.4–12.5 years. Data were collected cross-sectionally. Depression was measured using the CES-DC									
2	Cross-sectional[a]	No serious risk of bias	No serious inconsistency	No serious indirectness	No serious imprecision	None	2,008	Rieck et al. found no association between 20mSRT performance and **depression** in girls (OR = 2.09, 95%CI 0.96–4.57) and boys (OR = 1.63, 95%CI 0.08–3.17) [S124] Greenleaf et al. reported a favourable association between 20mSRT performance and **depression** in girls ($F = 8.64$, $r = -0.54$, $p = 0.01$), but not boys ($F = 5.14$, $r = -0.25$) [S125] Due to heterogeneity in the measurement of 20mSRT performance and mental health a meta-analysis was not possible	Low[b]

*Note 20mSRT, 20 m shuttle run test; CES-DC, Center for Epidemiology Studies–Depression Scale for Children.

a. Includes 2 studies [S124, S125].

b. The quality of the evidence remained as "low" as there were no serious concerns about the quality of included cross-sectional studies or reasons to increase the rating.

Table 5
Association between 20mSRT performance and psychosocial health.

No. of studies	Design	Quality assessment					No. of participants	Absolute effect	Quality
		Risk of bias	Inconsistency	Indirectness	Imprecision	Other			
The range of mean ages was 9.5–13.3 years. Data were collected cross-sectionally. Psychosocial health was measured using self-report questionnaires: KIDSCREEN-52 and various adapted questionnaires based on Bandura's guide for constructing self-efficacy scales									
5	Cross-sectional[a]	Serious risk of bias[b]	No serious inconsistency	No serious indirectness	No serious imprecision	None	1,949	4 of 5 studies reported favourable relationships between 20mSRT performance and either **self-efficacy**, **enjoyment**, or **outcome expectancy** [S13, S126–S128] 1 study reported favourable associations between 20mSRT performance and **moods/emotions** for girls ($\beta=0.14$, $p=0.009$), but no association for boys ($\beta=0.05$, $p=0.43$) [S129] Due to heterogeneity in the measurement of 20mSRT performance and psychosocial health a meta-analysis was not possible	Very low[c]

*Note 20mSRT, 20 m shuttle run test.

a. Includes 5 studies [S13, S126–S129].

b. Serious risk of bias. 4 out of 5 studies did not control for sex and/or age in the analyses [S13, S126–S128].

c. The quality of the evidence was downgraded from "low" to "very low" because of a serious risk of bias that diminished the level of confidence in the observed effects.

Table 6
Association between 20mSRT performance and self-esteem and physical self-perception.

No. of studies	Design	Quality assessment					No. of participants	Absolute effect	Quality
		Risk of bias	Inconsistency	Indirectness	Imprecision	Other			
The range of mean ages was 8.2–13.7 years; 3 studies reported an age range in lieu of the mean, and in these cases the included ages ranged from 8 to 17 years. Data were collected cross-sectionally and up to 4 years of follow-up. Self-esteem and/or physical self-perception was measured using an adapted version of the Body Self-Esteem Scale for children, Perceived Competence in Physical Education Scale, Behavioural Regulation in Exercise-2 questionnaire, KIDSCREEN-52 questionnaire, Self-Description Questionnaire II, CSAPPA questionnaire, and the Intrinsic Motivation Inventory questionnaire									
1	Longitudinal[a]	Serious risk of bias[b]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	None	821	20mSRT performance at baseline was not associated with **body dissatisfaction** at 4 years follow-up in girls ($\beta=-4.96$, $p=0.21$) and boys ($\beta=-4.67$, $p=0.21$) [S130]	Very low[d]
9	Cross-sectional[e]	Serious risk of bias[f]	No serious inconsistency	No serious indirectness	No serious imprecision	None	6,973	Favourable associations Colquitt et al. found 20mSRT performance was favourably associated with **perceived usefulness** ($r=0.35$, $p<0.01$; $\beta=0.27$, $p=0.005$) [S13] Gao et al. found 20mSRT performance was favourably associated with **perceived sport competence** ($\beta=0.27$, $p<0.01$) [S131] Riiser et al. found 20mSRT performance was favourably associated with **self-determined motivation** ($\beta=2.08$, $p<0.01$) [S133] Greenleaf et al. found boys ($F=10.2$, $r=0.35$, $p<0.01$) and girls ($F=6.87$, $r=0.48$, $p=0.01$) in the Healthy Fitness Zone (HFZ) for 20mSRT performance reported better **self-perceived strength** than those not in the HFZ. Additionally, boys ($F=80.2$, $r=0.97$, $p<0.001$) and girls ($F=26.1$, $r=0.93$, $p=0.0001$) in the HFZ for 20mSRT performance reported better **self-perceived endurance** and better **body satisfaction** (Boys: $F=8.3$, $r=0.31$, $p=0.01$; Girls: $F=8.4$, $r=0.53$, $p=0.01$) than those not in the HFZ [S125] Cairney et al. found 20mSRT performance was favourably associated with **perceived adequacy** ($\beta=0.08$, $SE=0.19$, $p<0.001$) and **predilection** ($\beta=0.03$, $SE=0.1$, $p<0.05$) [S73] Shen et al. found 20mSRT performance favourably associated with **perceived competence** in physical education ($r=0.21$, $p<0.01$) [S134] Jurimae et al. found 20mSRT performance was favourably associated with **perceived endurance** (r range = 0.31–0.51, $p<0.05$) [S132] Olive et al. found 20mSRT performance was favourably associated with **body dissatisfaction** in boys ($\beta=-9.87$, $SE=2.6$, $p<0.001$) and girls ($\beta=-11.0$, $SE=3.5$, $p=0.002$) [S130] Mixed findings Morales et al. found 20mSRT performance was favourably associated with **self-perception** in girls ($\beta=0.13$, $p=0.01$) but not boys ($\beta=0.04$, $p=0.5$) [S129] Greenleaf et al. found girls in the HFZ for 20mSRT performance reported better **self-esteem** ($F=8.46$, $p=0.01$) compared to those not in the HFZ. No differences were observed in boys ($F=5.24$, p = not significant) [S125] Null findings Morales et al. found 20mSRT performance was not associated with **autonomy** in girls ($\beta=0.05$, $p=0.42$) and boys ($\beta=0.00$, $p=0.42$) [S129] Due to heterogeneity in the measurement of 20mSRT performance and self-esteem and physical self-perception a meta-analysis was not possible	Very low[g]

*Note 20mSRT, 20 m shuttle run test; HFZ, Healthy Fitness Zone; NIZ, Needs Improvement Zone.

a. Includes 1 longitudinal study [S130].

b. Serious risk of bias. 1 study scale used to measure Body Dissatisfaction was adapted by the authors and had an equal number of positive and negative items, but did not mention changing the Likert scale of responses (ranged from "very much like I feel" to "not at all like I feel") to match the item's direction [S130].

c. Serious imprecision. Only one longitudinal study was published examining the association between 20 m SRT performance and self-esteem and physical self-perception.

d. The quality of evidence was downgraded from "low" to "very low" because of serious imprecision and a serious risk of bias that diminished the level of confidence in the observed effects.

e. Includes 9 cross-sectional studies [S13, S73, S125, S129–S134].

f. Serious risk of bias. 5 of 9 studies reported using convenience sampling [S13, S130, S131, S133, S134], and it was unclear whether 3 of the remaining 4 studies used convenience sampling methods [S73, S125, S132]. Additionally, 3 studies did not control for age or sex differences in 20mSRT performance [S13, S131, S134].

g. The quality of evidence was downgraded from "low" to "very low" because of a serious risk of bias that diminished the level of confidence in the observed effects.

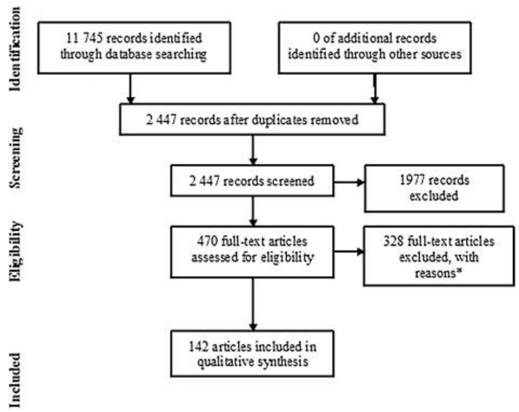

Fig. 1. PRISMA flowchart for the identification, screening, eligibility and inclusion of articles. *Reasons for exclusion: study did not include the proper 20mSRT protocol (n = 55); study did not assess the relationship between 20mSRT performance and a relevant health indicator (n = 137); sample size (n = 2); study participants were not in the target age range (n = 2); grey literature and/or not original research (i.e., review, conference proceeding, abstract; n = 82); populations with known disease or injury (n = 8); could not obtain full-text article (n = 12); non-English language article (n = 26); and duplicate paper discovered during level 2 screening (n = 4). Some articles were excluded for more than one reason.

Nine studies (6% of included studies) investigated the associations between 20mSRT performance and indicators of self-esteem and physical self-perception (Table 6, Supplement Table 6). One study used a mixed longitudinal cohort and cross-sectional design, and the remaining eight studies used a cross-sectional design. The longitudinal cohort study reported a null association between 20mSRT performance and body dissatisfaction after 4-years of follow-up [S130]. However, because this study used an un-validated scale to measure body dissatisfaction, there was a serious risk of bias. There was also serious imprecision (insufficient evidence). As a result, the quality of the evidence from the longitudinal cohort study was downgraded from low to very low. Six cross-sectional studies reported favourable associations between 20mSRT performance and various indicators of self-esteem and physical self-perception [S13, S128, S130, S132–S134]. Three studies reported mixed favourable and null associations between 20mSRT performance and various indicators of self-esteem or physical self-perception [S73, S125, S129]. The quality of the evidence was downgraded from low to very low as a result of a serious risk of bias in sampling methods and not controlling for age and sex differences in some studies [S13, S128, S134].

Six cross-sectional studies (4% of included studies) investigated the relationship between 20mSRT performance and indicators of quality of life and wellbeing (Table 7, Supplement Table 7). Three studies reported favourable associations between 20mSRT performance and various indicators of quality of life and wellbeing [S123, S133, S135], two studies reported mixed favourable and null associations [S123, S129], and one study reported mixed null and unfavourable findings [S10]. One study used a questionnaire that was not validated for the age groups in which it was used, thereby creating a serious risk of bias [S123]. There was also serious imprecision (i.e., wide confidence intervals) in the estimates reported by two studies [S123, S133]. As a result, the quality of the evidence was downgraded from low to very low.

One cross-sectional study (1% of included studies) investigated the relationship between 20mSRT performance and bone health (Table 8, Supplement Table 8) [S137]. This study found that 20mSRT performance was a favourably associated with bone mineral content in boys and girls. However, the study did not include lean body mass as a covariate in the analysis although it accounted for 67% of the variance in bone mineral content [S137]; this resulted in a serious risk of bias. In addition, the lack of studies in this area created a serious risk of imprecision. As a result, the quality of the evidence was downgraded from low to very low.

Nine cross-sectional studies (6% of included studies) investigated the relationship between 20mSRT performance and musculoskeletal fitness (Table 9, Supplement Table 9). The musculoskeletal fitness outcomes were divided into four categories: strength and power, muscular endurance, flexibility, and agility. Of the seven studies that investigated the association between 20mSRT performance and indicators of strength and power, four studies reported favourable findings [S13, S84, S108, S138], two studies reported mixed favourable and null findings [S12, S46], and one study reported null findings [S62]. All six studies assessing the relationship between 20mSRT performance and measures of muscular endurance reported favourable associations [S12, S13, S37, S42, S46, S62]. All four studies reported null associations between 20mSRT performance and flexibility [S13, S37, S42, S62]. Lastly, of the two studies that investigated the associations between 20mSRT performance and measures of agility, one reported favourable associations [S12] while the other reported null associations [S62]. The quality of the evidence was downgraded from low to very low as a result of a serious risk of bias. More specifically, the majority of studies used convenience sampling methods or did not describe their sampling procedures, and most did not describe the validity and reliability of their measurement instruments.

Four cross-sectional studies (3% of included studies) investigated the association between 20mSRT performance and motor skill development (Table 10, Supplement Table 10). Two of the four studies reported favourable associations between 20mSRT performance and motor skill development [S18, S139], and the remaining two studies reported null associations [S62, S140]. The quality of the evidence was downgraded from low to very low as a result of a serious risk of bias and serious imprecision. More specifically, three studies used convenience sampling methods, and one study did not control for sex differences in their analysis [S18, S62, S139], subsequently resulting in a serious risk of bias. Serious imprecision was recorded because one study reported wide confidence intervals [S139].

Two cross-sectional studies (1% of included studies) investigated the relationship between 20mSRT performance and the risk of injury/harm (Table 11, Supplement Table 11). One study reported a favourable association where children with better 20mSRT performance had lower odds of sustaining serious injury [S141]. Another study reported opposing findings where better 20mSRT performance was associated with higher odds of sustaining a bone fracture [S142]. The quality of the evidence was downgraded from low to very low due to this serious inconsistency.

A high-level summary of findings is reported in Table 12. Due to heterogeneity across all health indicators a meta-analysis of results was not possible.

4. Discussion

This systematic review combined results from 142 studies that investigated the associations between 20mSRT performance and various health indicators among children and youth aged 5–17 years. The included studies represented 319,311 children and youth (301,105 from unique samples) from 32 countries across 6 continents, which comprised of approximately 39% of the world population.[22] The quality of the evidence ranged from very low to moderate. Overall, the evidence suggests that better 20mSRT performance is associated with a variety of positive health indicators that range from lower adiposity levels to favourable cardiometabolic biomarker profiles and physical self-perception.

Some believe 20mSRT performance is strongly influenced by adiposity, suggesting that poor 20mSRT performers are on average the heavier children and youth.[23]However, our results indicate that the unadjusted correlation coefficients ranged widely from $r = -0.70$ (strong) [S47] to $r = -0.10$ (weak) [S28], with some studies having reported positive coefficients as high as $r = 0.13$ (weak unfavourable) [S83]. These varying results suggest that adiposity alone does not explain 20mSRT performance. Similarly, Olds and Dollamn reported that adiposity explained 40–60% of the longitudinal declines seen in 1.6 km running speed in 10- to 12-year-old children between 1985 and 1997.[24] In fact, it is likely that a combination of various physiological and psychological aspects help explain 20mSRT performance, including mechanical efficiency,[25] sustainable fractional of $\dot{V}O_{2max}$,[26] $\dot{V}O_2$ kinetics,[27] lactate threshold,[27] anaerobic capacity,[28] and motivation.[8]

In some cases 20mSRT performance was strongly associated with one subcategory, but not others, within similar health indicator categories. For instance, 20mSRT performance displayed favourable associations with some subcategories of the cardiometabolic biomarker category: among the most favourable were HOMA-IR (effect size: $r = 0.02$ [S101] to $r = -0.60$ [S26]; negligible to strong), C-reactive protein(effect size: $r = 0.01$ [S36] to $r = -0.15$ [S63]; negligible to weak) and aggregate indices of cardiometabolic disease risk (effect size: $r = 0.03$ [S101] to r = −0.31 [S52]; negligible to moderate). In addition, other health indicator subcategories also displayed consistent favourable associations: academic achievement (effect size: $r = -0.06$ [S117] to $r = 0.27$ [S121]; negligible to weak), muscular endurance (effect size: $r = 0.08$ [S37] to $r = 0.70$ [S13]; negligible to very strong), psychosocial health (effect size: $r = 0.23$ [S131] to $r = 0.71$ [S126]; weak to very strong) and physical self-perception (effect size: $r = 0.21$ [S134] to $r = 0.51$ [S132]; weak to strong); whereas, some health indicator subcategories displayed consistent null associations: mean arterial pressure (effect size: $r = 0.065$ [S3] to $r = -0.156$ [S11]; negligible to weak) and flexibility (effect size: $r = -0.01$ [S37] to $r = 0.20$ [S62]; negligible to weak). The variability across different subcategories within each health indicator was the main reason for mixed results and inconsistencies. However, it is apparent that very few studies reported unfavourable associations between 20mSRT performance and health indicators suggesting that there is an overall gradient between better performance and better holistic health among children and youth.

This study represents the largest effort to date to synthesize the associations between 20mSRT performance and health indicators in children and youth. The results from this study brought together findings from 142 studies using a rigorous systematic review process. Previous research has suggested that the 20mSRT would be a good measure to compare the CRF and health of children and youth across countries and continents.[6, 7] Indeed, the findings from our review support the 20mSRT as an appropriate population health indicator for children and youth.

A common issue with systematic reviews is that the synthesis of evidence is only as good as the papers that they contain.[30] Thus, the primary limitation of this review is that the quality of the evidence for all health indicator categories was rated as very low to moderate as a result of pooling findings that were always observational in nature. Nonetheless, the relative uniformity of findings with very few unfavourable relationships, despite the flaws inherent in several studies, should not prevent drawing confident conclusions about the strong and favourable relationships identified herein. Although the nature of the 20mSRT does not allow for randomized controlled trial (RCT) designs (the 20mSRT is a single measurement of a trait), the results from this study encourage the need for RCTs that investigate the sensitivity of 20mSRT performance improvements in children and youth and their relationship with changes in health indicators. In addition, a meta-analysis was not appropriate due to the heterogeneity between studies, which was mainly identified in the reporting of 20mSRT results as the exposure measure. For example, results were recorded as laps, stages, time, speed, distance, or predicted $\dot{V}O_{2peak}$, which in itself has many variations in the different equations used (e.g., Léger et al.[31] Ruiz et al.[32] Fitnessgram[33]). There is certainly a need for consistent reporting of 20mSRT results across studies, as recommended by Tomkinson et al.[7] Furthermore, we were not able to separate our results by sex or age groups because of the heterogeneity across studies, although study specific details regarding sex and age are available in Supplement 2. Lastly, we did not include studies written in languages other than English which may have introduced a bias.

This review also highlights many gaps in the literature. There is a need to further investigate the associations between 20mSRT performance and indicators of bone health, mental health, psychosocial health, and quality of life/wellbeing. These are areas of research that are important as mental health disorders continue to grow and result in healthcare challenges worldwide.[34] Furthermore, there is a need for more global and coordinated efforts to measure 20mSRT performance to better understand the health of children and youth across developed and developing countries. Evidence suggests that 20mSRT performance varies substantially worldwide[6]; thus, global measurement efforts could provide a proactive method to identify populations of youth at risk of poor health outcomes and future chronic diseases.[35]

5. Conclusion

Our study provides a systematic review of the associations between 20mSRT performance and health indicators among school-aged children and youth. The results demonstrate the broad utility of the 20mSRT as an indicator of not only physiological and physical health, but also psychosocial and cognitive health in children and youth. As a result, the 20mSRT would be an excellent tool to screen children and youth at risk of poor health and in need of increasing physical fitness levels. These results also support the notion that the 20mSRT could be used as an indicator to compare the holistic health of children and youth through temporal trends and between geographic locations (e.g., cities, provinces, states and countries).

Contributorship

Justin Lang, Kevin Belanger, Mark Tremblay and Grant Tomkinson developed the research questions and objectives. Justin Lang and Kevin Belanger contributed equally in screening articles, extracting data and producing result tables. Justin Lang drafted the manuscript. All authors contributed to interpreting results, editing, reviewing and approved the final manuscript.

Funding statement

This study received no financial support.

Ethical approval statement

Not applicable.

Data sharing statement

Not applicable.

Acknowledgements

We thank Margaret Sampson, MLIS, PhD, AHIP (Children's Hospital of Eastern Ontario) for developing the electronic search strategies, and Lise Dubas, MLIS student, University of Alberta, for peer review of the search strategies.

References

1. Armstrong N, Tomkinson GR, Ekelund U. Aerobic fitness and its relationship to sport, exercise training and habitual physical activity during youth. *Brit J Sports Med* 2011; 45: 849-858.

2. Ortega FB, Ruiz JR, Castillo M Jetal. Physical fitness in child hood and adolescence: a powerful marker of health. *Int J Obes* 2008; 32(1): 1–11.

3. Högström G, Nordström A, Nordström P. Highaerobic fitness in late adolescence is associated with a reduced risk of myocardial infraction later in life: a nationwide cohort study in men. *Eur Heart J* 2014; 35:3133-3140.

4. 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. http://dx.doi.org/10.1093/ije/dyv321. Published Online First: 20 December.

5. 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(2):64–69.

6. Lang JJ, Tremblay MS, Léger Letal. 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. *Sports Med* 2016. Published Online First: 20 September.

7. 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. *Brit J Sports Med* 2016. http://dx.doi.org/10.1136/bjsports-2016-095987. Published Online First: May 20th.

8. Tomkinson GR, Olds TS. Field tests of fitness, in *Paediatric exercise science and medicine*, Armstrong N, van Mechlen W, editors, New York, NY, Oxford University Press, 2000, p.109–128.

9. Mayorga-Vega M, Aguilar-Soto P, Viciana J. Criterion-related validity of the 20-M shuttle run test for estimating cardiorespiratory fitness: ameta-analysis. *J Sports Sci Med* 2015; 14:536–547.

10. Janssen I, Le Blanc AG. Systematic review of the health benefits of physical activity and fitness in school-aged children and youth. *Int J Behav Nutr Phys Act* 2010; 7:40.

11. Poitras VJ, Gray CE, Borghese M Metal. 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:S197–S239.

12. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009; 6(6): e1000097.

13. Schardt C, Adams MB, Owens T, Keits S, Fontelo P. Utilization of the PICO frame-work to improve searching Pub Med for clinical questions. *BMC Med Inf Decis Mak* 2007; 7:16.

14. Council of Europe. *Eurofit: handbook for the Eurofit tests of physical fitness*, Rome, Council of Europe, 1988.

15. Mcgowan J, Sampson M, Lefebvre C. An evidence based checklist for the peer review of electronic search strategies (PRESS EBC). *Evid Based Libr Inf Pract* 2010; 5(1):149–154.

16. Higgins J, Green S. *Cochrane handbook for systematic reviews of interventions version 5.1.0,* The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

17. Guyatt GH, Oxman AD, Akl EA, Kunz R, Vist G, Brozed Jetal. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol* 2011; 64(4):383–394.

18. Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind Detal. GRADE guidelines: 6. Rating the quality of evidence—imprecision. *J Clin Epidemiol* 2011; 64(12):1283–1293.

19. Guyatt GH, Oxman AD, Kunz R, Wood cock J, Brozek J, Helfand Metal. GRADE guidelines: 7. Rating the quality of evidence—inconsistency. *J Clin Epidemiol* 2011; 64(12): 1294–1302.

20. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand Metal. GRADE guidelines: 8. Rating the quality of the evidence—indirectness. *J Clin Epidemiol* 12 2011; 64: 1303–1310.

21. Guyatt GH, Oxman AD, Sultan S, Glasziou P, Akl EA, Alonso-Coello Petal. GRADE guidelines: 9. Rating up the quality of evidence. *J Clin Epidemiol* 2011; 64(12): 1311–1316.

22. United Nations, Department of Economic and Social Affairs, Population Division. *World population prospects: the 2015 revision, key findings and advancetables*. 2015; Working Paper No. ESA/P/WP. 241.

23. Rowland T. Declining cardiorespiratory fitness in youth: factor supposition? *Pediatr Exerc Sci* 2002; 14: 1–8.

24. Olds T, Dollman J. Are changes in distance-run performance of Australian children between 1985 and 1997 explained by changes in fatness? *Pediatr Exerc Sci* 2004; 16: 201–209.

25. Lussier L, Buskirk ER. Effects of an endurance training regimen of assessment of work capacity in prepubertal children. *Ann N Y Acad Sci* 1977; 301: 734–747.

26. Krahenbuhl GS, Pangrazi RP, Chomokos EA. Aerobic responses of young boys to submaximal running. *Res Q* 1979; 50: 413–421.

27. Péronnet F, Thibault G. Mathematical analysis of running performance and worldrecords. *J Appl Physiol* 1989; 67: 453–465.

28. Sjödin B. The relationship among running economy, aerobic power, and on set of blood lactate accumulation in young boys (11–15 years), in *Exercise and sport biology*, Komi PV, editor, Champaign, IL, Human Kinetics, 1982, p. 57–60.

30. Weir A, Rabia S, Ardern C. Trusting systematic reviews and meta-analyses: all that glitters is not gold! *Brit J Sports Med* 2016. Published online first: 15 October.

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

32. Ruiz JR, Ramirez-Lechuga J, Ortega Fetal. Artificial neural network-based equation for estimating VO2max for the 20 m shuttle run test in adolescents. *Artif Intell Med* 2008; 44: 233–245.

33. Meredith MD, Welk GJ *FITNESSGRAM/ACTIVITYGRAM test administration manual (updated 4th ed),* Meredith MD, Welk GJ, editors, Champaign, IL, Human Kinetics, 2010.

34. Whiteford HA, Degenhardt L, Rehm Jetal. Global burden of disease attributable to mental and substance used is orders: findings from the Global Burden of Disease Study 2010. *Lancet* 2013; 382(9904): 1575–1586.

35. Ruiz JR, Castro-Pinero J, Artero E Getal. Predictive validity of health-related fitness in youth: a systematic review. *Brit J Sports Med* 2009; 43(12): 909–923.