This is the peer reviewed author accepted manuscript (post print) version of a published work that appeared in final form in:

Fraser, Brooklyn J, Rollo, Scott, Sampson, Margaret, Magnussen, Costan G, Lang, Justin J, Tremblay, Mark S, Tomkinson, Grant R 2021 'Health-related criterion-referenced cut-points for musculoskeletal fitness among youth: a systematic review', Sports Medicine, vol. 51, no. 12, pp. 2629-2646

This un-copyedited output may not exactly replicate the final published authoritative version for which the publisher owns copyright. It is not the copy of record. This output may be used for non-commercial purposes.

The final definitive published version (version of record) is available at:
https://doi.org/10.1007/s40279-021-01524-8

Persistent link to the UniSA Research Outputs Repository:
https://researchoutputs.unisa.edu.au/11541.2/148382

General Rights

Copyright and moral rights for the publications made accessible in the UniSA Research Outputs Repository are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the persistent link identifying the publication in the UniSA Research Outputs Repository

If you believe that this document breaches copyright, please contact us and provide details. We will remove access to the work immediately and investigate your claim.

TITLE PAGE

Health-related criterion-referenced cut-points for musculoskeletal fitness among youth: A systematic review

Authors

Brooklyn J. FRASER[*1] ORCID: 0000-0002-1752-5431

Scott ROLLO[2] ORCID: 0000-0003-1122-2728

Margaret SAMPSON[2] ORCID: 0000-0003-2550-9893

Costan G. MAGNUSSEN[1,3,4] ORCID: 0000-0002-6238-5730

Justin J. LANG[2,5] ORCID: 0000-0002-1768-319X

Mark S. TREMBLAY[2] ORCID: 0000-0002-8307-3568

Grant R. TOMKINSON[*6,7] ORCID: 0000-0001-7601-9670

Institutional affiliations

[1]Menzies Institute for Medical Research, University of Tasmania, Hobart, TAS, Australia

[2]Healthy Active Living and Obesity Research Group, Children's Hospital of Eastern Ontario Research Institute, Ottawa, ON, Canada

[3]Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland

[4]Centre for Population Health Research, University of Turku and Turku University Hospital, Turku, Finland

[5]Centre for Surveillance and Applied Research, Public Health Agency of Canada, Ottawa, ON, Canada

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

[7]Alliance for Research in Exercise, Nutrition and Activity (ARENA), School of Health Sciences, University of South Australia, Adelaide, SA, Australia.

Running title

Health-related cut-points for musculoskeletal fitness among youth

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

Acknowledgements

We would like to thank Katie O'Hearn (Children's Hospital of Eastern Ontario Research Institute) for methodological assistance. We also acknowledge the help of the authors of the included studies who provided additional details.

ABSTRACT

Background: Musculoskeletal fitness is an excellent functional measure that is significantly related to health among youth.

Objective: To identify health-related criterion-referenced cut-points for musculoskeletal fitness (MSF) among youth.

Methods: A systematic search of two electronic databases (MEDLINE and SPORTDiscus) was conducted in September 2020. Only peer-reviewed studies that developed health-related criterion-referenced cut-points for MSF among youth were eligible provided they included: (a) youth aged 5–17 years from the general population; (b) at least one quantitative assessment of MSF (e.g., muscular strength); (c) at least one quantitative assessment of health (e.g., cardiometabolic risk); (d) a criterion for health; and (e) a quantitative analysis (e.g., receiver operating characteristic [ROC] curve) of at least one health-related cut-point for MSF. A narrative synthesis was used to describe the results of included studies.

Results: Collectively, 13 studies developed health-related criterion-referenced cut-points for MSF among 14,476 youth from 15 countries were included. Muscular strength demonstrated high discriminatory ability (median area under the curve [AUC]: ≥0.71) for cardiometabolic risk/metabolic syndrome, sarcopenic obesity risk and bone health, and moderate discriminatory ability (median AUC: 0.64–0.70) for asthma. Muscular power also demonstrated high discriminatory ability for bone health, but only moderate discriminatory ability for cardiometabolic risk/metabolic syndrome and low discriminatory ability (median AUC: 0.56–0.63) for cognition/academic performance. Both muscular endurance and flexibility demonstrated low discriminatory ability for musculoskeletal pain. Health-related cut-points for MSF that demonstrated significant discriminatory ability were generally higher for boys compared to girls (for muscular strength and power) and generally increased with age for muscular strength and power but remained stable for flexibility.

Conclusions: Currently, there are not enough data to establish universal health-related cut-points for MSF among youth. Despite variation in the health-related discriminatory ability of different MSF tests, handgrip strength and standing broad jump emerged as the two tests with the highest discriminatory ability. More research, using standardized testing protocols and health-risk definitions, is required to better triangulate universal health-related cut-points for MSF among youth.

PROSPERO registration number: CRD42020207458.

KEY POINTS

- Thirteen studies examined health-related criterion-referenced cut-points for MSF among youth, with variation in the discriminatory ability of different MSF tests observed. Given the heterogeneity of findings, no universal health-related cut-points for MSF could be supported.
- Handgrip strength and the standing broad jump were the MSF tests with the highest discriminatory ability.
- The small number of studies identified in this review highlight the need for additional research to improve our current understanding and move towards robust health-related cut-points. Standardized MSF test protocols and health-risk definitions need to be established in the pediatric population, with differences by sex, age, growth and maturation considered when quantifying health-related cut-points for MSF among youth.

1 INTRODUCTION

Musculoskeletal fitness (MSF), a health-related component of physical fitness, is a multidimensional construct that reflects the ability of a muscle or group of muscles to exert force maximally (muscular strength), quickly (muscular power), or repeatedly (muscular endurance). Flexibility, defined as the ability to move a joint through a full range of motion, is another component of MSF. The importance of MSF in childhood is becoming increasingly recognized, with physical activity guidelines now recommending that in addition to aerobic activities, children and adolescents (herein termed youth) aged 5–17 years should engage in muscle- and bone-strengthening activities at least three times per week [1]. Better MSF in youth has health-benefits related to improved cardiovascular health, skeletal and bone health, and self-esteem [2, 3]. Youth MSF is not only a marker of current health, but also a predictor of future health outcomes [4-6]. In a large Swedish cohort of over one million participants, higher levels of MSF in adolescence was inversely associated with premature mortality [7] and type 2 diabetes [8] in adulthood. Flexibility, however, is one component of MSF where inconclusive evidence exists regarding potential health benefits [9, 10]. It is unclear whether the absence of compelling evidence is due to lack of associations, or whether flexibility is being poorly measured and interpreted. To confirm or challenge existing evidence regarding flexibility, health, physical performance, and injury prevention, researchers have been encouraged to examine all aspects of flexibility and health in youth [9]. Despite the health-benefits of MSF becoming mostly better understood, MSF levels of youth have remained relatively stable and have even declined more recently [11, 12]. It is difficult to correct this decline and gain an understanding of how the MSF levels and overall health of contemporary youth fare with youth from previous generations, if youth physical fitness levels are not being routinely and accurately monitored.

The National Academy of Medicine (formerly called the Institute of Medicine) released a report in 2012 recommending that health-related physical fitness measures be incorporated in national youth fitness surveys, with MSF measures highlighted as a key component [9]. Most countries do not routinely monitor the physical fitness levels of youth and national surveillance systems are still required. Furthermore, MSF is often overlooked as a health-related component of physical fitness [13]. A lack of awareness, in addition to a lack of monitoring, means more work is required before national fitness surveillance systems that include MSF tests are routinely administered. Some MSF tests (e.g., handgrip strength, standing broad jump) are widely- and historically-used, practical, feasible, scalable, and ecologically valid, making them potential population health indicators that could complement other measures as part of national health surveillance systems [14]. However, a key barrier for adopting national fitness surveillance systems is the lack of evidence-based, health-related criterion-referenced MSF cut-points (also termed "standards" or "thresholds"), which help us to interpret MSF test results by identifying individuals with low/high levels of MSF on a given test. The absence of these cut-points means there is no standard way to identify youth who could be targeted with strategies aimed at increasing their MSF levels, providing a way to guide healthy public policy efforts.

International experts have endorsed MSF for both its predictive utility and its population-based health surveillance utility [9, 14], and recommend as a priority, that health-related criterion-referenced cut-points be established for both cardiorespiratory fitness (CRF) and MSF [9, 14]. Youth fitness levels have historically been compared against norm-referenced national/international standards, which allows youth to be compared against centile bands to identify expected, better than expected, or worse than expected developmental changes. Because these norm-referenced standards are time dependent and often outdated, such values

may not reflect norms of contemporary youth. This approach then potentially misclassifies youth and could provide a false indication of their overall fitness status [15]. Further, the degree to which existing normative standards are related to health indicators of interest is often unknown or poorly established. Given these limitations, there has been a concentrated effort to shift away from the norm-referenced framework and to establish health-related criterion-referenced cut-points, where youth fitness levels are compared against an absolute, often health-related, criterion [15]. Criterion-referenced cut-points can be, and are often, established using receiver operating characteristic curve (ROC) analysis, an analytic approach that identifies the optimal cut-point that balances sensitivity and specificity to discern a certain health outcome [16]. Previous systematic reviews have identified criterion-referenced cut-points for CRF linked with cardiovascular health outcomes [17, 18], however no such review exists focusing on MSF cut-points among youth. Therefore, this review seeks to identify all studies (not only those identified using ROC analysis and linked with cardiovascular health outcomes) that have established health-related criterion-referenced cut-points for MSF among youth.

2 METHODS

2.1 Protocol and registration

The review protocol was registered with the International Prospective Register of Systematic Reviews (PROSPERO; registration number CRD42020207458). Reporting follows the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement [19].

2.2 Eligibility criteria

One large systematic review of health-related criterion-referenced cut-points for CRF and MSF in youth was initially undertaken before being separated into two smaller reviews. For this review, we included studies that developed health-related criterion-referenced cut-points for MSF in youth. Studies were considered eligible if they included the following:

(1) participants were youth aged 5–17 years from the general population, with studies on targeted special populations (e.g., athletes, clinically obese) excluded;

(2) quantitative assessment of at least one MSF component (e.g., muscular strength, muscular power, muscular endurance, flexibility) was provided;

(3) quantitative assessment of at least one health indicator (e.g., cardiometabolic risk, bone health, cognition/academic performance, musculoskeletal pain) was provided;

(4) the health criterion (i.e., definition of health risk or test positivity) was provided; and

(5) quantitative analysis (e.g., receiver operating characteristic [ROC] curve) of at least one health-related cut-point for MSF was provided.

No study design (e.g., cross-sectional, case-control, cohort), date, or language restrictions were imposed, although we did impose a publication status restriction with only peer-reviewed original journal articles that were published or in-press eligible for inclusion.

2.3 Information sources

We identified studies by searching electronic databases, reference lists, topical systematic analyses/reviews, and personal libraries [19]. The electronic database search was performed by an experienced academic librarian on the 3rd and 4th of September 2020 using MEDLINE (OVID interface) and SPORTDiscus (EBSCOhost interface).

2.4 Search

The electronic database search strategy was designed by an academic librarian experienced in systematic reviews using a method designed to optimize term selection [20], with all researchers participating in generating candidate search terms. Search strategies are shown in Electronic Supplementary Material Appendix S1.

2.5 Study selection

All database records were downloaded and imported into a reference manager database where duplicates were removed, and then uploaded to Covidence systematic review software (Veritas Health Innovation, Melbourne, VIC, Australia), where a final duplicate detection pass was made. Record screening comprised two levels. Level 1 involved two researchers independently screening the titles and abstracts against inclusion criteria, with consensus required for further screening. Level 2 involved two researchers independently screening the full texts against inclusion criteria, with consensus required for final inclusion. When necessary, discrepancies were resolved by discussion or a third researcher prior to reaching consensus.

2.6 Data collection process

Descriptive data were extracted into a pre-piloted, study-specific Excel (Microsoft, Redmond, WA, USA) spreadsheet by one researcher and checked for accuracy by a second researcher. If required, additional information was requested from the publication's corresponding author via email (e.g., to clarify published results, request additional data, avoid double counting data).

2.7 Data items

The following study-specific descriptive data were extracted: author(s), country, year of publication, participant demographics (e.g., sex/gender, age, race, ethnicity), study design, sampling method, sample size, MSF measure (e.g., MSF component, MSF test, protocol), health indicator, health criterion, health-related MSF cut-points (e.g., cut-points, 95% confidence intervals [CI]), measures of diagnostic accuracy (e.g., area under the curve, sensitivity, specificity, predictive values), and prevalence of health risk.

2.8 Risk of bias in individual studies

Risk of bias was assessed at the study-level by the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool [21]. The QUADAS-2, which has been recommended [22] for use in systematic reviews to assess the risk of bias and applicability concerns of diagnostic accuracy studies, assesses four study domains: patient selection, index test, reference standard, and flow and timing. All domains are assessed for risk of bias, with the first three also assessed for applicability concerns. Signaling questions are included to help judge the risk of bias. Risk of bias and applicability concerns were independently assessed by two researchers as ‘low’, ‘high’ or ‘unclear’ in each domain, with discrepancies resolved by discussion or a third researcher prior to reaching consensus.

2.9 Summary measures and methods of analysis

We used a narrative synthesis to describe the results of included studies because the between-study methodological differences (e.g., the mix of MSF components, health indicators, health criteria) prevented a meta-analysis. The primary measures of interest were the health-related MSF cut-points and the area under the ROC curve [AUC]—a summary statistic that represents the discriminatory ability of the MSF test. To interpret the magnitude of AUC values, effect sizes of 0.56, 0.64 and 0.71 were used as thresholds for low, moderate and

high, respectively, with effect sizes <0.56 considered to be negligible [23]. AUC values were described by medians and ranges.

3 RESULTS

3.1 Study selection

A total of 3784 unique records were identified (4017 records were identified from the electronic database search plus one additional study from a topical systematic review, minus 234 duplicate records). Fifty-eight full-text articles were retained after the first level of screening, and 13 articles were retained after the second level and included in this systematic review. Figure 1 illustrates the PRISMA flowchart for included studies.

Insert Figure 1 here

3.2 Study characteristics

All 13 included studies were published in English between 2014 and 2020 and employed a cross-sectional [24-34], longitudinal cohort [35] or case-control [36] design. Collectively, these studies represented 14,476 youth aged 4–18 years from 15 countries (12 high-income and 3 upper-middle-income countries [37] across 3/7 continents). All studies: (a) quantitatively analyzed at least one health-related criterion-referenced cut-point for MSF in youth from the general population, with almost all [24-30, 32-36] using ROC analysis; (b) quantitatively assessed at least one MSF component, including muscular strength [25, 26, 28, 31, 32, 34-36], muscular power [24-26, 29, 30], muscular endurance [27, 33] and flexibility [27, 33]; (c) quantitatively assessed at least one health indicator, including cardiometabolic risk/metabolic syndrome [25, 26, 31, 32], bone health [24, 35], cognition/academic performance [29, 30], musculoskeletal pain [27, 33], sarcopenic obesity risk [28, 34] and asthma [36]; and (d) defined the health criterion/criteria. Individual study characteristics are summarized in Table 1.

Insert Table 1 here

3.3 Risk of bias within studies

The assessments of risk of bias and applicability concerns are summarized in Figure 2 (see also Electronic Supplementary Material Appendix S2). Patient selection risk of bias and applicability concerns were assessed as high for most studies. Although almost all studies [24-35] avoided a case-control design, most studies [24, 26, 29, 30, 32-35] did not use random or consecutive sampling for participant recruitment. All studies were assessed to have high risk of bias for the index test (the MSF test) because none of the health-related cut-points for MSF were pre-specified. There were low applicability concerns for index test for 12 studies, with high concern for one study [36] because the index test was not interpreted blind to the reference standard. Risk of bias and applicability concerns were low for the reference standard (the health indicator[s]). Flow of patients through the study and timing of the index tests and reference standard were also considered low risk in all studies.

Insert Figure 2 here

3.4 Synthesis of results

3.4.1 Muscular strength

Health-related criterion-referenced cut-points for muscular strength in youth were developed in eight studies [25, 26, 28, 31, 32, 34-36] which used a cross-sectional [25, 26, 28, 31, 32, 34], longitudinal cohort [35] or case-control [36] design (Table 1). Study sample sizes ranged from 140 to 4021 participants and participant ages ranged from 4 to 17 years. Muscular strength was measured as maximal handgrip strength using handheld dynamometry, expressed as absolute handgrip strength [25, 35, 36] or as handgrip strength relative to either body mass [25, 26, 31, 32] or body mass index (BMI) [28, 34].

Cut-points for muscular strength, quantified using either ROC analysis [25, 26, 28, 32, 34-36] or conditional inference tree analysis [31], were consistently associated with cardiometabolic risk/metabolic syndrome [25, 26, 31, 32], sarcopenic obesity risk [28, 34], bone health [35] and asthma [36]. From age 6 to 12 years, cut-points for absolute handgrip strength (kg) associated with bone health were similar for both sexes (range: 7.4–18.4 kg), with subsequent age-specific increases larger for boys compared to girls (range: boys, 21.9–34.4 kg; girls, 19.9–23.7 kg) (Figure 3; Table 1). Absolute handgrip cut-points for 11-year-olds were similar for bone health and asthma. Cut-points for handgrip strength expressed relative to body mass (kg/kg) and BMI (kg/kg/m^2) were consistently higher for boys compared to girls and generally increased with age. Cut-points associated with cardiometabolic risk/metabolic syndrome ranged from 0.33 to 0.59 kg/kg for boys and 0.28 to 0.44 kg/kg for girls aged 6–17 years, whereas those associated with sarcopenic obesity risk ranged from 0.67 to 1.04 kg/kg/m^2 for boys and 0.66 to 0.92 kg/kg/m^2 for girls aged 4–14 years.

Insert Figure 3 here

Collectively, muscular strength demonstrated high discriminatory ability (AUC: median=0.78, range=0.54–0.90), with 82% or 23/28 of AUC values ≥0.71 (Table 1) (see also Electronic Supplementary Material Appendix S3 for additional measures of diagnostic accuracy). High discriminatory ability was typically shown for both boys (AUC: median=0.80, range=0.56–0.90) and girls (AUC: median=0.79, range=0.54–0.86), and for cardiometabolic risk/metabolic syndrome (AUC: median=0.76, range=0.54–0.87), sarcopenic obesity risk (AUC: median=0.79, range=0.72–0.90), and bone health (AUC: median=0.80, range=0.71–0.81), with typically moderate discriminatory ability for asthma (AUC: median=0.69, range=0.65–0.73). One study [25] showed that handgrip strength relative to body mass (AUC: median=0.80, range=0.75–0.87) had significantly higher discriminatory

ability for detecting cardiometabolic risk/metabolic syndrome than absolute handgrip strength (AUC: median=0.60, range=0.54–0.66).

3.4.2 Muscular power

Five studies developed health-related criterion-referenced cut-points for muscular power in youth [24-26, 29, 30], all of which used cross-sectional designs and quantified cut-points using ROC analysis (Table 1). Study-specific sample sizes ranged from 114 to 2272 participants, with participant ages ranging from 6 to 17 years. Muscular power was measured using the standing broad jump [25, 26, 29], squat jump [30] or countermovement jump [24, 30] tests, and was expressed as absolute jump performance [25, 26, 29, 30], peak vertical jump power (calculated from countermovement jump height, age, body mass and height) [24] or as standing broad jump performance relative to body mass [25]. Cut-points for muscular power were consistently associated with cardiometabolic risk/metabolic syndrome [25, 26] and bone health [24]. Mixed results were observed when examining cognition/academic performance [29, 30]; cut-points were identified for the standing broad jump [29], whereas the squat jump and countermovement jump were not predictive [30]. Cut-points for muscular power associated with bone health, cardiometabolic risk/metabolic syndrome and cognition/academic performance were consistently higher for boys compared to girls (Figure 3; Table 1). For the standing broad jump (cm), health-related cut-points tended to increase with age. Cut-points associated with cardiometabolic risk/metabolic syndrome ranged from 104.5 to 184.2 cm for boys and 81.5 to 129.5 cm for girls aged 6–17 years, while cut-points associated with cognition/academic performance ranged from 151.7 to 202.8 cm for boys and 113.8 to 141.3 cm for girls aged 12–17 years. For 12- to 17-year-olds, standing broad jump cut-points for cardiometabolic risk/metabolic syndrome and cognition/academic performance

were dissimilar for boys (i.e., generally higher for cognition/academic performance) yet similar for girls.

Collectively, muscular power demonstrated moderate discriminatory ability (AUC: median=0.66, range=0.30–0.85), with similar effects for both boys (AUC: median=0.66, range=0.50–0.83) and girls (AUC: median=0.68, range=0.53–0.85). High discriminatory ability was generally shown for bone health (AUC: median=0.84, range=0.82–0.85), whereas moderate discriminatory ability was shown for cardiometabolic risk/metabolic syndrome (AUC: median=0.70, range=0.66–0.83) and low discriminatory ability for cognition/academic performance (AUC: median=0.60, range=0.30–0.75). One study [25] showed that absolute standing broad jump (AUC: median=0.72, range=0.66–0.78) and standing broad jump relative to body mass (AUC: median=0.70, range=0.68–0.72) had similar discriminatory ability for detecting cardiometabolic risk/metabolic syndrome.

3.4.3 Muscular endurance

Health-related criterion-referenced cut-points for muscular endurance in youth were developed in two studies [27, 33]. Both studies used a cross-sectional design (Table 1). Study-specific sample sizes ranged from 375 to 1364 participants and participant ages ranged from 9 to 17 years. Muscular endurance was measured using the dynamic curl-ups/sit-ups [27, 33], static curl-up [33], lateral plank [33] or prone plank [33] tests. Both studies examined the association between muscular endurance and musculoskeletal pain, assessed as self-reported low back pain. Cut-points were quantified using ROC analysis and only dynamic curl-ups and the static curl-up demonstrated statistically significant discriminatory ability at the 95% level. The cut-point for dynamic curl-ups was 48 repetitions for 14- to 15-year-old boys (i.e., Grades 9–10); the cut-point for the static curl-up was 306 seconds for 11-

to 13-year-old girls (i.e., Grades 6–8) (Figure 3; Table 1). Muscular endurance demonstrated low discriminatory ability (AUC: median=0.59, range=0.50–0.75), with similar effects for both sexes (AUC: boys, median=0.59, range=0.50–0.75; girls, median=0.61, range=0.50–0.73).

3.4.4 Flexibility

Two studies developed health-related criterion-referenced cut-points for flexibility in youth [27, 33]. Both studies used cross-sectional designs and examined cut-points using ROC analysis (Table 1). Participant ages ranged from 9 to 17 years and study-specific sample sizes ranged from 375 to 1364 participants. Flexibility was measured using the back-saver sit-and-reach [33], trunk extension [33] or sit-and-reach [27] tests. Cut-points for flexibility associated with musculoskeletal pain (assessed as self-reported low back pain) only demonstrated statistically significant discriminatory ability for the back-saver sit-and-reach and sit-and-reach tests for girls. Cut-points for the sit-and-reach ranged from 14 to 16 cm (past the toes) for 10- to 17-year-old girls, with 1.9 cm identified as the cut-point for the back-saver sit-and-reach for 9- to 10-year-old girls (i.e., Grades 4–5) (Figure 3; Table 1). Collectively, flexibility demonstrated low discriminatory ability (AUC: median=0.58, range=0.49–0.80), with effects similar for boys (AUC: median=0.58, range=0.49–0.61) and girls (AUC: median=0.58, range=0.52–0.80).

4 DISCUSSION

This study is the first to systematically review published papers that have developed health-related criterion-referenced cut-points for MSF among youth. In total, 13 studies developed cut-points for MSF, including measures of muscular strength, muscular power, muscular endurance, and flexibility, in relation to a range of health indicators, including

cardiometabolic risk/metabolic syndrome, asthma, musculoskeletal pain, bone health, sarcopenic obesity risk, and cognition/academic performance. All studies were published over the past six years, highlighting this topic as an emerging area of research. Although this systematic review is a good step forward in collating and summarizing existing knowledge, our results highlight a great deal of variation in the discriminatory ability of different MSF tests. Because the current evidence is incomplete, we are unable to establish universal health-related cut-points for MSF among youth.

4.1 Summary of evidence

The results of this systematic review were challenging to consolidate given the lack of standardization among MSF test protocols. For example, handgrip strength, a maximal isometric grip strength task that is good predictor of overall muscular strength [38], was the only measure of muscular strength used among the included studies [25, 26, 28, 31, 32, 34-36]. Unfortunately, no standardized handgrip strength test protocol was used. Handgrip strength was measured either seated or standing, and scored as the average of the better of two attempts for each hand or only the dominant hand. In addition, some researchers adjusted the handheld dynamometer for the participants' hand size while others did not. These protocol differences could have led to systematic differences in handgrip strength and affected between-study comparisons [39].

There were also between-study differences in the way in which handgrip strength was expressed. For example, cut-points for handgrip strength were expressed in measured units (kg) [35, 36] or normalized for different measures of body size, including body mass [25, 26, 31, 32] and BMI [28, 34]. Whether or not differences in body size should be accounted for in these analyses requires further attention. One study [25] compared the discriminatory ability

of absolute handgrip strength and handgrip strength relative to body mass [25] and showed that relative handgrip strength had significantly higher discriminatory ability for detecting cardiometabolic risk/metabolic syndrome. These findings suggest that if muscular strength is to be measured as part of a youth health-related fitness test battery, that handgrip strength should be adjusted for body size for better detection of cardiometabolic risk/metabolic syndrome. However, for universal muscular strength cut-points (such as handgrip strength) to be established, there needs to be consensus regarding the most appropriate, both practically and statistically, method for normalizing non-weight bearing muscular strength for body size and maturation.

Caution, however, must be taken when analyzing health-related cut-points for muscular strength among youth. For example, cut-points for relative handgrip strength associated with sarcopenic obesity risk were identified in two studies [28, 34]. However, both studies included BMI in their index test (handgrip strength was expressed relative to BMI) and, albeit implicitly, in their reference standard (increased risk of sarcopenic obesity, defined as two standard deviations lower than the mean muscle-to-fat ratio for the 3rd BMI quintile). This approach could have resulted in a spurious correlation between these two variables and as such, results should be interpreted with caution. Establishing criterion-referenced cut-points that are not only accurate and best reflect the health-related association of interest, but are also simple to interpret and implement, is a priority to help facilitate feasibility and scalability of these findings.

Developing cut-points for MSF is challenging especially in youth because of the inherent complexities resulting from growth and maturation. Of the health-related cut-points identified for MSF, some were presented stratified by age and/or sex, whereas others were presented for

age and sex combined. The decision to stratify by age and/or sex in these studies could have been driven by sample size (because small samples result in reduced statistical power), although it is essential that sex-, age-, growth- and maturation-related changes be considered in future studies when establishing health-related criterion-referenced cut-points for MSF among youth. One approach to account for age-related changes in MSF [40, 41] observed among the included studies [35] was the use of lambda-mu-sigma (LMS) parameters (developed from national survey data) to standardize MSF scores prior to determining age- and sex-specific cut-points. By establishing cut-points from LMS-derived standardized scores and then expressing them as unstandardized (raw) scores (i.e., in the measured units [e.g., handgrip strength in kilograms]), cut-points can be easily interpreted to provide instantaneous feedback to youth without the need for complex and time-consuming mathematical computations. Further research to better understand how developmental and maturational differences relate to health-related cut-points is desperately needed.

Because MSF is specific to the muscle group, type of muscle action, muscle contraction, contraction velocity, equipment, and joint range of motion, it unfortunately cannot be measured by a single test. Future research should therefore examine whether it is better to use single or combined scores from multiple MSF tests to predict health outcomes. One study [33] compared the discriminatory ability of single and combined scores from multiple MSF tests to predict musculoskeletal pain (self-reported low back pain) and found similar discriminatory ability. While combining scores from multiple MSF tests may provide a better overall representation of MSF, it is not as practical as using a single MSF test. Future research is therefore required to explore if combined scores from multiple MSF tests improves predictive utility.

Several studies examined whether the predictive utility of the MSF cut-points was impacted by covariates. For example, using logistic regression analysis, potential covariates such as pubertal age [26], CRF [25, 26, 31], self-reported physical activity levels [31], and socioeconomic status [29] were examined. Results indicated that covariates did not confound the associations between MSF and health outcomes and therefore, no covariates were included in the cut-point analyses. The fact that these studies indicated that covariates did not confound the associations further underlines the potential value of some of these MSF tests in the population setting.

4.2 Future research directions and knowledge translation for public health

Common public health measures used in health surveillance and clinical settings often have universally accepted cut-points to help identify groups and/or individuals at risk of poor health. Examples include waist circumference [42], BMI [43, 44], blood pressure [45], fasting blood glucose, triglycerides, high-density lipoprotein cholesterol [46], CRF [18] and movement behaviors [47]. Similar universal cut-points for MSF do not currently exist, despite the fact that the origins of some adverse health conditions (e.g., cardiovascular disease) are found in childhood, and low MSF in childhood and adolescence is associated with cardiometabolic disease risk during these years [3] and later in life [4-6]. Progressing toward universal health-related cut-points for MSF is incredibly important for these measures to be adopted into regular public health practice. Such universal cut-points would help our current understanding of temporal trends in MSF, which have largely been limited to trends in mean MSF [11, 12, 48] by allowing the estimation of trends in the proportion of youth meeting or not meeting the health-related cut-points, and thus, potentially improving the evaluation of public health policies and interventions. Given the low number of studies identified in our systematic review, it is apparent that more research is needed.

Indeed, the evidence that we identified demonstrates higher discriminatory ability of handgrip strength and standing broad jump compared to other MSF tests. In addition, both the handgrip strength and standing broad jump are easy to implement, cost effective, and scalable [14], suggesting that these measures may be the best candidates for public health practice. Future research should produce more ROC curve analyses for MSF (particularly handgrip strength and standing broad jump, for example, using accepted Eurofit protocols [49]), especially among youth from low- and middle-income countries, to better triangulate universal cut-points. There is also a need to better harmonize efforts by standardizing test protocols for both MSF and the various health indicators, to enable more consistent measurements, to facilitate data pooling, and to better identify youth meeting or not meeting cut-points. Collectively, these efforts could help bridge the gap between research and practice.

There is also a need to develop appropriate public health messaging when communicating health-related cut-points for MSF among young. For instance, these cut-points are not intended to replace existing clinical diagnostic procedures. Not attaining a MSF cut-point suggests that an individual is at potential risk of poor health and systems should be in place to provide assistance and support for remediation. For example, these cut-points could also be used to identify those in need of a physical activity intervention, while MSF testing could be applied to help track progress over time.

4.3 Strengths and limitations

A limitation of our systematic review is that one of the 13 included studies [36] used a case-control study design. Although this design gives an indication of the maximum

discriminatory ability of the MSF test (in this case, handgrip strength), results of case-control studies tend to inflate the discriminatory ability in professional practice [50]. A cross-sectional design better reflects actual practice and is more likely to provide a valid estimate of discriminatory ability [50, 51]. Of the remaining 12 included studies, 11 used a cross-sectional study design and one used a longitudinal cohort study design. Although cross-sectional study designs are key to addressing our research question, they cannot be used to infer causality because a temporal sequence cannot be established. Only one included study [26] has examined whether cut-points developed at one time-point are valid at a second time-point. An additional limitation of our systematic review was that the 13 included studies only represented youth from the Americas and Europe, highlighting a geographical gap in our current understanding. This, combined with the degree of heterogeneity that precluded meta-analysis, means that universal health-related cut-points for MSF were unable to be established.

Strengths of this study include the rigorous systematic review approach. All identified studies used appropriate analytical techniques (e.g., ROC curves, conditional inference tree analyses) to quantify cut-points for MSF and to determine the health-related discriminatory ability of MSF. Furthermore, included studies used acceptable, feasible, ecologically valid, and scalable measures of MSF, with handgrip strength and the standing broad jump emerging as the two measures with the highest discriminatory ability. Despite gaps in our current understanding, it is reassuring that all 13 identified studies were published over the past six years, highlighting this as an emerging field of research.

CONCLUSIONS

This systematic review identified few studies that established health-related criterion-referenced cut-points for MSF among youth. Owing to the variation in the discriminatory ability of different MSF tests, universal health-related cut-points for MSF could not be established. Despite this being an emerging area of research with only a small number of studies identified and variation in the health-related discriminatory ability of different MSF tests, the handgrip strength and standing broad jump emerged as the two tests with the highest discriminatory ability. For universal health-related MSF cut-points to be established, we require: (1) more data from broad populations of youth from around the world; (2) standardized MSF test protocols; (3) consensus among the definitions of health risk; and (4) consideration of sex-, age-, growth- and maturation-related changes.

DECLARATIONS

Funding

This systematic review was funded by the Public Health Agency of Canada (#4500414578).

Conflicts of interest/Competing interests

Brooklyn J. Fraser, Scott Rollo, Margaret Sampson, Costan G. Magnussen, Justin J. Lang, Mark S. Tremblay, and Grant R. Tomkinson declare they have no conflicts of interest.

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.

Availability of data and material

The datasets analyzed in this review are available from the corresponding authors on reasonable request.

Code availability

Not applicable.

Authors' contributions

BJF screened the records, extracted the data and took responsibility for the integrity of the data, synthesized the results, and co-wrote the manuscript. ASR screened the records, checked the data for accuracy and took responsibility for the integrity of the data, contributed to the interpretation of results, and critically reviewed the manuscript for important intellectual content. MS designed and executed the systematic search strategy, and critically reviewed the manuscript for important intellectual content. CGM contributed to the interpretation of results and critically reviewed the manuscript for important intellectual content. JJL developed the research question, designed the systematic review, contributed to the interpretation of results, and critically reviewed the manuscript for important intellectual content. MST developed the research question, designed the systematic review, contributed to the interpretation of results, and critically reviewed the manuscript for important intellectual content. GRT developed the research question, designed the systematic review, had full access to the data, synthesized the results, and co-wrote the manuscript. All authors have read and approved the final version of the manuscript, agree to be accountable for all aspects of the work, and agree with the order of presentation of the authors.

Ethics approval

Not applicable.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

REFERENCES

1. Bull FC, Al-Ansari SS, Biddle S, Borodulin K, Buman MP, Cardon G, et al. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. Br J Sports Med. 2020;54(24):1451–62.
2. Ortega FB, Ruiz JR, Castillo MJ, Sjöström M. Physical fitness in childhood and adolescence: a powerful marker of health. Int J Obes (Lond). 2008;32(1):1–11.
3. Smith JJ, Eather N, Morgan PJ, Plotnikoff RC, Faigenbaum AD, Lubans DR. The health benefits of muscular fitness for children and adolescents: a systematic review and meta-analysis. Sports Med. 2014;44(9):1209–23.
4. García-Hermoso A, Ramírez-Campillo R, Izquierdo M. Is muscular fitness associated with future health benefits in children and adolescents? A systematic review and meta-analysis of longitudinal studies. Sports Med. 2019;49(7):1079–94.
5. Fraser BJ, Blizzard L, Schmidt MD, Juonala M, Dwyer T, Venn AJ, et al. Childhood cardiorespiratory fitness, muscular fitness and adult measures of glucose homeostasis. J Sci Med Sport. 2018;21(9):935–40.
6. Fraser BJ, Huynh QL, Schmidt MD, Dwyer T, Venn AJ, Magnussen CG. Childhood muscular fitness phenotypes and adult metabolic syndrome. Med Sci Sports Exerc. 2016;48(9):1715–22.
7. Ortega FB, Silventoinen K, Tynelius P, Rasmussen F. Muscular strength in male adolescents and premature death: cohort study of one million participants. BMJ. 2012;345:e7279.
8. Crump C, Sundquist J, Winkleby MA, Sieh W, Sundquist K. Physical fitness among Swedish military conscripts and long-term risk for type 2 diabetes mellitus: a cohort study. Ann Intern Med. 2016;164(9):577–84.

9. Institute of Medicine (IOM). Fitness measures and health outcomes in youth. Washington (DC): National Academies Press (US); 2012.
10. Ruiz JR, Castro-Piñero J, Artero EG, Ortega FB, Sjöström M, Suni J, et al. Predictive validity of health-related fitness in youth: a systematic review. Br J Sports Med. 2009;43(12):909-23.
11. Kaster T, Dooley FL, Fitzgerald JS, Walch TJ, Annandale M, Ferrar K, et al. Temporal trends in the sit-ups performance of 9,939,289 children and adolescents between 1964 and 2017. J Sports Sci. 2020;38(16):1913–23.
12. Tomkinson GR, Kaster T, Dooley FL, Fitzgerald JS, Annandale M, Ferrar K, et al. Temporal trends in the standing broad jump performance of 10,940,801 children and adolescents between 1960 and 2017. Sports Med. 2021;51(3):531-48.
13. Faigenbaum AD, MacDonald JP, Stracciolini A, Rebullido TR. Making a strong case for prioritizing muscular fitness in youth physical activity guidelines. Curr Sports Med Rep. 2020;19(12):530–6.
14. Lang JJ, Smith JJ, Tomkinson GR. Global surveillance of cardiorespiratory fitness and musculoskeletal fitness. The Routledge Handbook of Youth Physical Activity. New York: Routledge; 2020. p. 47–68.
15. Zhu W, Mahar MT, Welk GJ, Going SB, Cureton KJ. Approaches for development of criterion-referenced standards in health-related youth fitness tests. Am J Prev Med. 2011;41(4 Suppl 2):S68–76.
16. McNeil BJ, Keller E, Adelstein SJ. Primer on certain elements of medical decision making. N Engl J Med. 1975;293(5):211–5.
17. 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. 2019;53(15):953–8.

18. Ruiz JR, Cavero-Redondo I, Ortega FB, Welk GJ, Andersen LB, Martinez-Vizcaino V. Cardiorespiratory fitness cut points to avoid cardiovascular disease risk in children and adolescents; what level of fitness should raise a red flag? A systematic review and meta-analysis. Br J Sports Med. 2016;50(23):1451–8.

19. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ. 2009;339:b2535.

20. Bramer WM, de Jonge GB, Rethlefsen ML, Mast F, Kleijnen J. A systematic approach to searching: an efficient and complete method to develop literature searches. J Med Libr Assoc. 2018;106(4):531–41.

21. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529–36.

22. Ma L-L, Wang Y-Y, Yang Z-H, Huang D, Weng H, Zeng X-T. Methodological quality (risk of bias) assessment tools for primary and secondary medical studies: what are they and which is better? Military Medical Research. 2020;7(1):1–11.

23. Rice ME, Harris GT. Comparing effect sizes in follow-up studies: ROC area, Cohen's d, and r. Law Hum Behav. 2005;29(5):615–20.

24. Baptista F, Mil-Homens P, Carita AI, Janz K, Sardinha LB. Peak vertical jump power as a marker of bone health in children. Int J Sports Med. 2016;37(8):653–8.

25. Castro-Piñero J, Laurson KR, Artero EG, Ortega FB, Labayen I, Ruperez AI, et al. Muscle strength field-based tests to identify European adolescents at risk of metabolic syndrome: the HELENA study. J Sci Med Sport. 2019;22(8):929–34.

26. Castro-Piñero J, Perez-Bey A, Cuenca-Garcia M, Cabanas-Sanchez V, Gómez-Martínez S, Veiga OL, et al. Muscle fitness cut points for early assessment of cardiovascular risk in children and adolescents. J Pediatr. 2019;206:134–41.
27. Dorneles RCG, da Rosa Oliveira HL, de Araújo Bergmann ML, Bergmann GG. Flexibility and muscle strength/resistance indicators and screening of low back pain in adolescents. Rev Bras Cineantropom Desempenho Hum. 2016;18(1):93–102.
28. Gontarev S, Jakimovski M, Georgiev G. Using relative handgrip strength to identify children at risk of sarcopenic obesity. Nutr Hosp. 2020;34(3):490–6.
29. Martínez-López EJ, Grao-Cruces A, De La Torre-Cruz MD, Ruiz-Ariza A. Associations between physical fitness and academic performance in teenagers. S Afr J Res Sport Phys Educ R. 2019;41(1):63–75.
30. Mazzoccante RP, de Luca Corrêa H, de Santana DS, Câmara MA, de Sousa BRC, de Sousa IRC, et al. Attention and executive function are predicted by anthropometric indicators, strength, motor performance, and aerobic fitness in children aged 6–10 years. Hum Mov. 2020;21(1):40–8.
31. Peterson MD, Zhang P, Saltarelli WA, Visich PS, Gordon PM. Low muscle strength thresholds for the detection of cardiometabolic risk in adolescents. Am J Prev Med. 2016;50(5):593–9.
32. Ramírez-Vélez R, Peña-Ibagon JC, Martínez-Torres J, Tordecilla-Sanders A, Correa-Bautista JE, Lobelo F, et al. Handgrip strength cutoff for cardiometabolic risk index among Colombian children and adolescents: the FUPRECOL Study. Sci Rep. 2017;7:42622.
33. Saint-Maurice PF, Welk GJ, Burns R, Plowman SA, Corbin CB, Hannon JC. The criterion-referenced validity of the FITNESSGRAM Trunk-Extension test. J Sports Med Phys Fitness. 2015;55(10):1252–63.

34. Steffl M, Chrudimsky J, Tufano JJ. Using relative handgrip strength to identify children at risk of sarcopenic obesity. PLoS One. 2017;12(5):e0177006.
35. Saint-Maurice PF, Laurson K, Welk GJ, Eisenmann J, Gracia-Marco L, Artero EG, et al. Grip strength cutpoints for youth based on a clinically relevant bone health outcome. Arch Osteoporos. 2018;13(1):92.
36. Latorre-Román PA, Navarro-Martínez AV, Mañas-Bastidas A, García-Pinillos F. Handgrip strength test as a complementary tool in monitoring asthma in daily clinical practice in children. Iran J Allergy Asthma Immunol. 2014;13(6):396–403.
37. The World Bank. World Bank Country and Lending Groups [cited 30 December 2020]; Available from: http://data.worldbank.org/about/country-and-lending-groups
38. Wind AE, Takken T, Helders PJ, Engelbert RH. Is grip strength a predictor for total muscle strength in healthy children, adolescents, and young adults? Eur J Pediatr. 2010;169(3):281–7.
39. Roberts HC, Denison HJ, Martin HJ, Patel HP, Syddall H, Cooper C, et al. A review of the measurement of grip strength in clinical and epidemiological studies: towards a standardised approach. Age and ageing. 2011;40(4):423-9.
40. 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 Suppl 2):S111–S6.
41. Cole TJ, Green PJ. Smoothing reference centile curves: the LMS method and penalized likelihood. Stat Med. 1992;11(10):1305–19.
42. Xi B, Zong X, Kelishadi R, Litwin M, Hong YM, Poh BK, et al. International waist circumference percentile cutoffs for central obesity in children and adolescents aged 6 to 18 years. J Clin Endocrinol Metab. 2020;105(4):e1569–83.

43. Cole TJ, Lobstein T. Extended international (IOTF) body mass index cut-offs for thinness, overweight and obesity. Pediatr Obes. 2012;7(4):284–94.
44. WHO Multicentre Growth Reference Study Group. WHO Child Growth Standards based on length/height, weight and age. Acta Paediatr Suppl. 2006;450:76–85.
45. Riley M, Hernandez AK, Kuznia AL. High blood pressure in children and adolescents. Am Fam Physician. 2018;98(8):486–94.
46. Al-Hamad D, Raman V. Metabolic syndrome in children and adolescents. Transl Pediatr. 2017;6(4):397–407.
47. Tremblay MS, Carson V, Chaput JP, Connor Gorber S, Dinh T, Duggan M, et al. Canadian 24-hour movement guidelines for children and youth: an integration of physical activity, sedentary behaviour, and sleep. Appl Physiol Nutr Metab. 2016;41(6 Suppl 3):S311–27.
48. Dooley FL, Kaster T, Fitzgerald JS, Walch TJ, Annandale M, Ferrar K, et al. A Systematic Analysis of Temporal Trends in the Handgrip Strength of 2,216,320 Children and Adolescents Between 1967 and 2017. Sports Med. 2020;50(6):1129-44.
49. Tomkinson GR, Carver KD, Atkinson F, Daniell ND, Lewis LK, Fitzgerald JS, 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):1445–56.
50. Leeflang MM. Systematic reviews and meta-analyses of diagnostic test accuracy. Clin Microbiol Infect. 2014;20(2):105–13.
51. Whiting P, Rutjes AW, Reitsma JB, Glas AS, Bossuyt PM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. Ann Intern Med. 2004;140(3):189–202.

52. Global Initiative for Asthma (GINA). Global strategy for asthma management and prevention 2006 [date last updated 2006; cited 7 January 2021]; Available from: www.ginasthma.org

53. Jolliffe CJ, Janssen I. Development of age-specific adolescent metabolic syndrome criteria that are linked to the Adult Treatment Panel III and International Diabetes Federation criteria. J Am Coll Cardiol. 2007;49(8):891–8.

54. Meredith MD, Welk G. Fitnessgram/Activitygram Test Administration Manual. 4 ed. Champaign, IL: Human Kinetics; 2010.

TABLES

Table 1. Summary of the included studies that developed health-related criterion-referenced cut-points for MSF in youth (presented in chronological order by publication date).

Study	Year	Country	Gender/Sex	Age (years)	n	Health indicator	Health criterion	MSF component	MSF test	Health-related cut-point	AUC (95%CI)
Latorre-Román et al. [36][a]	2014	Spain	Boys+Girls	11.4±0.9 (healthy) 11.1±1.0 (asthma)	140	Asthma	Diagnosed according to the GINA 2006 criteria [52]	Muscular strength	Handgrip strength	16.8 kg (asthma)[b] 15.1 kg (severity)[c]	0.73 (0.64–0.81) (asthma) 0.65 (0.52–0.79) (severity)
Saint-Maurice et al. [33][d]	2015	USA	Boys+Girls	10.1±0.7 to 15.3±0.6 (boys) 9.8±0.7 to 15.1±0.7 (girls)	375	Musculoskeletal pain (self-reported LBP)	Indicated presence of LBP on ≥7 of 14 questionnaire items	Muscular endurance	Dynamic curl-ups	48 repetitions (Boys, G9–10)[e]	≤0.59 (Boys, G4–5) ≤0.59 (Boys, G6–8) 0.75 (Boys, G9–10) 0.61 (Girls, G4–5) 0.63 (Girls, G6–8) 0.70 (0.50–0.83) (Girls, G9–10)
									Static curl-up	306 s (Girls, G6–8)[e]	≤0.59 (Boys, G4–5) 0.61 (Boys, G6–8) 0.63 (Boys, G9–10) 0.67 (Girls, G4–5) 0.71 (0.56–0.84) (Girls, G6–8) 0.73 (0.54–0.88) (Girls, G9–10)
									Lateral plank		≤0.59 (Boys, G4–5) ≤0.59 (Boys, G6–8) 0.50 (Boys, G9–10) ≤0.51 (Girls, G4–5) 0.52 (Girls, G6–8) 0.71 (0.51–0.86) (Girls, G9–10)
									Prone plank		≤0.59 (Boys, G4–5) ≤0.59 (Boys, G6–8) 0.59 (Boys, G9–10) ≤0.51 (Girls, G4–5) 0.50 (Girls, G6–8) 0.60 (Girls, G9–10)
								Flexibility	Back-saver sit-and-reach	1.9 cm (Girls, G4–5)[e f]	≤0.59 (Boys, G4–5) ≤0.59 (Boys, G6–8) 0.58 (Boys, G9–10) 0.80 (0.66–0.90) (Girls, G4–5) 0.68 (Girls, G6–8) 0.53 (Girls, G9–10)
									Trunk extension[g]		0.52 (0.37–0.67) (Boys, G4–5) 0.58 (0.45–0.71) (Boys, G6–8) 0.61 (0.42–0.78) (Boys, G9–10) 0.59 (0.44–0.71) (Girls, G4–5) 0.56 (0.40–0.71) (Girls, G6–8) 0.52 (0.33–0.71) (Girls, G9–10)

Peterson et al. [31][d]	2016	USA	Boys+Girls	10–12	1326	Cardiometabolic risk (%BF, SBP, triglycerides, HDL-C and glucose)	≥75th percentile ("high risk") of a continuous standardized score calculated with all cardiometabolic risk factors	Muscular strength	Handgrip strength (relative to body mass)	High risk: ≤0.33 kg/kg (Boys) ≤0.28 kg/kg (Girls) Intermediate risk: >0.33 to ≤0.45 kg/kg (Boys) >0.28 to ≤0.36 kg/kg (Girls)	
Dorneles et al. [27][d]	2016	Brazil	Boys+Girls	10–17	1364	Musculoskeletal pain (self-reported LBP)	Indicated presence of LBP/discomfort as "often and always"	Muscular endurance	Sit-ups		0.53 (0.46–0.60) (Boys) 0.52 (0.46–0.57) (Girls)
								Flexibility	Sit-and-reach	15 cm (Girls, 10 y)[e f] 15 cm (Girls, 11 y)[e f] 16 cm (Girls, 12 y)[e f] 15 cm (Girls, 13 y)[e f] 14 cm (Girls, 14 y)[e f] 14 cm (Girls, 15 y)[e f] 14 cm (Girls, 16 y)[e f] 14 cm (Girls, 17 y)[e f]	0.49 (0.42–0.56) (Boys) 0.58 (0.52–0.63) (Girls)
Baptista et al. [24][d]	2016	Portugal	Boys+Girls	8.5±0.4	114	Bone health (BMD)	≤–1 SD ("increased risk") of height-adjusted WBLH BMD	Muscular power	Peak vertical jump power (calculated from counter-movement jump height, age, body mass and height)	735 W (Boys) 621 W (Girls)	0.82 (0.68–0.95) (Boys) 0.85 (0.70–1.00) (Girls)
Steffl et al. [34][d]	2017	Czech Republic	Boys+Girls	4–14[h]	730	Sarcopenic obesity risk	≤–2 SD ("increased risk") of mean muscle-to-fat ratio for the 3rd BMI quintile	Muscular strength	Handgrip strength (relative to BMI)	0.72 kg/kg/m^2 (Boys, 4–9 y) 1.04 kg/kg/m^2 (Boys, 10–14 y) 0.68 kg/kg/m^2 (Girls, 4–9 y) 0.92 kg/kg/m^2 (Girls, 10–14 y)	0.72 (0.61–0.83) (Boys, 4–9 y) 0.90 (0.82–0.97) (Boys, 10–14 y) 0.79 (0.69–0.89) (Girls, 4–9 y) 0.79 (0.69–0.89) (Girls, 10–14 y)
Ramirez-Velez et al. [32][d]	2017	Colombia	Boys+Girls	9–17	1950	Cardiometabolic risk (WC, SBP, DBP, triglycerides, HDL-C and glucose)	≥1 SD ("increased risk") of a continuous standardized score computed with all cardiometabolic risk factors	Muscular strength	Handgrip strength (relative to body mass)	0.38 kg/kg (Boys, 9–12 y) 0.45 kg/kg (Boys, 13–17 y) 0.36 kg/kg (Girls, 9–12 y) 0.44 kg/kg (Girls, 13–17 y)	0.84 (0.74–0.94) (Boys, 9–12 y) 0.80 (0.68–0.92) (Boys, 13–17 y) 0.83 (0.71–0.95) (Girls, 9–12 y) 0.79 (0.70–0.89) (Girls, 13–17 y)
Saint-Maurice et al. [35][i j]	2018	USA	Boys+Girls	11, 13, 15, 17	433	Bone health (BMC)	≤–0.84 SD ("some risk") and <–1.5 SD ("high risk") of height-adjusted WBLH BMC	Muscular strength	Handgrip strength	High risk: <7.8 kg (Boys, 6 y)[e] <8.9 kg (Boys, 7 y)[e] <10.4 kg (Boys, 8 y)[e] <12.0 kg (Boys, 9 y)[e] <13.8 kg (Boys, 10 y)[e] <15.8 kg (Boys, 11 y)[e] <18.4 kg (Boys, 12 y)[e] <21.9 kg (Boys, 13 y)[e] <25.7 kg (Boys, 14 y)[e] <29.0 kg (Boys, 15 y)[e] <31.5 kg (Boys, 16 y)[e] <33.2 kg (Boys, 17 y)[e] <34.4 kg (Boys, 18 y)[e]	High risk: 0.80 (0.77–0.83) (Boys) 0.79 (0.76–0.81) (Girls)

										<7.4 kg (Girls, 6 y)[e] <8.5 kg (Girls, 7 y)[e] <9.9 kg (Girls, 8 y)[e] <11.6 kg (Girls, 9 y)[e] <13.6 kg (Girls, 10 y)[e] <16.0 kg (Girls, 11 y)[e] <18.2 kg (Girls, 12 y)[e] <19.9 kg (Girls, 13 y)[e] <21.2 kg (Girls, 14 y)[e] <22.1 kg (Girls, 15 y)[e] <22.7 kg (Girls, 16 y)[e] <23.3 kg (Girls, 17 y)[e] <23.7 kg (Girls, 18 y)[e] Some risk: 7.8–10.0 kg (Boys, 6 y)[e] 8.9–11.2 kg (Boys, 7 y)[e] 10.4–13.1 kg (Boys, 8 y)[e] 12.0–15.2 kg (Boys, 9 y)[e] 13.8–17.3 kg (Boys, 10 y)[e] 15.8–19.9 kg (Boys, 11 y)[e] 18.4–23.1 kg (Boys, 12 y)[e] 21.9–27.4 kg (Boys, 13 y)[e] 25.7–32.2 kg (Boys, 14 y)[e] 29.0–36.3 kg (Boys, 15 y)[e] 31.5–39.3 kg (Boys, 16 y)[e] 33.2–41.4 kg (Boys, 17 y)[e] 34.4–42.8 kg (Boys, 18 y)[e] 7.4–8.9 kg (Girls, 6 y)[e] 8.5–10.2 kg (Girls, 7 y)[e] 9.9–12.0 kg (Girls, 8 y)[e] 11.6–14.0 kg (Girls, 9 y)[e] 13.6–16.5 kg (Girls, 10 y)[e] 16.0–19.4 kg (Girls, 11 y)[e] 18.2–22.1 kg (Girls, 12 y)[e] 19.9–24.2 kg (Girls, 13 y)[e] 21.2–25.7 kg (Girls, 14 y)[e] 22.1–26.8 kg (Girls, 15 y)[e] 22.7–27.6 kg (Girls, 16 y)[e] 23.3–28.3 kg (Girls, 17 y)[e] 23.7–28.7 kg (Girls, 18 y)[e]	Some risk: 0.71 (0.67–0.74) (Boys) 0.81 (0.79–0.84) (Girls)
Castro-Piñero et al. [25][d]	2019	Austria, Belgium, France, Germany, Greece, Hungary, Italy, Spain and Sweden	Boys+Girls	12–17	969	MetS and cardiometabolic risk (WC, SBP, DBP, triglycerides, HDL-C, glucose and insulin)	MetS: diagnosed according to the Jolliffe and Janssen [53] criteria (i.e., ≥3 increased cardiometabolic risk factors); and cardiometabolic risk: >1 SD ("increased risk") of a continuous standardized score	Muscular strength	Handgrip strength		MetS: 0.56 (0.52–0.61) (Boys) 0.54 (0.50–0.58) (Girls) Cardiometabolic risk: 0.66 (0.61–0.70) (Boys) 0.64 (0.59–0.68) (Girls)
									Handgrip strength (relative to body mass)	MetS and cardiometabolic risk: 0.44 kg/kg (Boys, 13 y)[e] 0.48 kg/kg (Boys, 14 y)[e] 0.52 kg/kg (Boys, 15 y)[e] 0.56 kg/kg (Boys, 16 y)[e] 0.59 kg/kg (Boys, 17 y)[e]	MetS: 0.87 (0.84–0.90) (Boys) 0.76 (0.72–0.79) (Girls) Cardiometabolic risk: 0.83 (0.79–0.87) (Boys)

							computed with all cardiometabolic risk factors			0.41 kg/kg (Girls, 13 y)[e] 0.41 kg/kg (Girls, 14 y)[e] 0.41 kg/kg (Girls, 15 y)[e] 0.42 kg/kg (Girls, 16 y)[e] 0.42 kg/kg (Girls, 17 y)[e]	0.75 (0.71–0.79) (Girls)
								Muscular power	Standing broad jump	MetS and cardiometabolic risk: 135.4 cm (Boys, 13 y)[e] 151.5 cm (Boys, 14 y)[e] 165.4 cm (Boys, 15 y)[e] 175.9 cm (Boys, 16 y)[e] 184.2 cm (Boys, 17 y)[e] 118.1 cm (Girls, 13 y)[e] 121.8 cm (Girls, 14 y)[e] 123.0 cm (Girls, 15 y)[e] 126.0 cm (Girls, 16 y)[e] 129.5 cm (Girls, 17 y)[e]	MetS: 0.75 (0.71–0.79) (Boys) 0.66 (0.62–0.70) (Girls) Cardiometabolic risk: 0.78 (0.73–0.81) (Boys) 0.69 (0.65–0.73) (Girls)
									Standing broad jump (relative to body mass)		MetS: 0.72 (0.68–0.76) (Boys) 0.68 (0.63–0.72) (Girls) Cardiometabolic risk: 0.70 (0.66–0.74) (Boys) 0.70 (0.66–0.74) (Girls)
Martínez-López et al. [29][d]	2019	Spain	Boys+Girls	12–17	2272	Cognition/academic performance	≥7 of 10 ("high performance") for grades in mathematics and language (combined) and physical education	Muscular power	Standing broad jump	Mathematics+language grades: 151.7 cm (Boys, 12 y) 166.3 cm (Boys, 13 y) 167.3 cm (Boys, 14 y) 180.0 cm (Boys, 15 y) 180.8 cm (Boys, 16 y) 202.8 cm (Boys, 17 y) 117.8 cm (Girls, 12 y) 118.8 cm (Girls, 13 y) 125.8 cm (Girls, 14 y) 129.3 cm (Girls, 15 y) 117.8 cm (Girls, 16 y) 120.8 cm (Girls, 17 y) Physical education grades: 157.8 cm (Boys, 12 y) 164.5 cm (Boys, 13 y) 159.8 cm (Boys, 14 y) 161.5 cm (Boys, 15 y) 186.8 cm (Boys, 16 y) 193.0 cm (Boys, 17 y) 113.8 cm (Girls, 12 y) 117.8 cm (Girls, 13 y) 128.5 cm (Girls, 14 y) 141.3 cm (Girls, 15 y) 131.3 cm (Girls, 16 y) 125.8 cm (Girls, 17 y)	Mathematics+language grades: 0.58 (Boys, 12 y) 0.54 (Boys, 13 y) 0.53 (Boys, 14 y) 0.50 (Boys, 15 y) 0.56 (Boys, 16 y) 0.58 (Boys, 17 y) 0.58 (Girls, 12 y) 0.55 (Girls, 13 y) 0.64 (Girls, 14 y) 0.65 (Girls, 15 y) 0.53 (Girls, 16 y) 0.61 (Girls, 17 y) Physical education grades: 0.68 (Boys, 12 y) 0.55 (Boys, 13 y) 0.63 (Boys, 14 y) 0.66 (Boys, 15 y) 0.66 (Boys, 16 y) 0.72 (Boys, 17 y) 0.75 (Girls, 12 y) 0.73 (Girls, 13 y) 0.63 (Girls, 14 y) 0.71 (Girls, 15 y) 0.68 (Girls, 16 y) 0.71 (Girls, 17 y)
Castro-Piñero et al. [26][d j]	2019	Spain	Boys+Girls	6–10, 12–16	511	Cardiometabolic risk (sum of 2 skinfolds, SBP, HOMA-IR, triglycerides, total	>1 SD ("increased risk") of a continuous standardized score	Muscular strength	Handgrip strength (relative to body mass)	0.37 kg/kg (Boys, 6–10 y) 0.47 kg/kg (Boys, 12–16 y) 0.31 kg/kg (Girls, 6–10 y) 0.42 kg/kg (Girls, 12–16 y)	0.75 (0.62–0.88) (Boys, 6–10 y) 0.75 (0.60–0.89) (Boys, 12–16 y) 0.86 (0.75–0.96) (Girls, 6–10 y) 0.74 (0.63–0.85) (Girls, 12–16 y)

						cholesterol-to-HDL-C ratio)	computed with all cardiometabolic risk factors	Muscular power	Standing broad jump	104.5 cm (Boys, 6–10 y) 140.5 cm (Boys, 12–16 y) 81.5 cm (Girls, 6–10 y) 120.5 cm (Girls, 12–16 y)	0.83 (0.72–0.94) (Boys, 6–10 y) 0.67 (0.53–0.80) (Boys, 12–16 y) 0.74 (0.61–0.87) (Girls, 6–10 y) 0.68 (0.54–0.79) (Girls, 12–16 y)
Mazzoccante et al. [30][d]	2020	Brazil	Boys+Girls	6–10	271	Cognition/academic performance	Attention and executive function based on attention cancellation test and trail making test performance	Muscular power	Squat jump		Attention: 0.44 (0.36–0.52) Executive function: 0.33 (0.26–0.41)
									Countermovement jump		Attention: 0.42 (0.34–0.50) Executive function: 0.30 (0.23–0.36)
Gontarev et al. [28][d]	2020	Republic of North Macedonia	Boys+Girls	6–10	4021	Sarcopenic obesity risk	≤–2 SD ("increased risk") of mean muscle-to-fat ratio for the 3rd BMI quintile	Muscular strength	Handgrip strength (relative to BMI)	0.67 kg/kg/m^2 (Boys) 0.66 kg/kg/m^2 (Girls)	0.84 (0.83–0.86) (Boys) 0.77 (0.75–0.79) (Girls)

Abbreviations: 95%CI=95% confidence interval; AUC=area under the curve; %BF=percent body fat; BMC=bone mineral content; BMD=bone mineral density; BMI=body mass index; cm=centimeters; G=Grade; GINA=Global Initiative for Asthma [52]; HDL-C=high-density lipoprotein cholesterol; HOMA-IR=homeostatic model assessment of insulin resistance; kg=kilograms; LBP=low back pain; MetS=metabolic syndrome; MSF=musculoskeletal fitness; s=seconds; SBP=systolic blood pressure; SD=standard deviation; USA=United States of America; W=Watts; WBLH=whole-body less head; WC=waist circumference; y=years.

Notes: Year=year of publication; all values >1.0 were reported to a maximum of 1 decimal place, with all values <1.0 reported to 2 decimal places; age (years) values were reported as mean±SD or range; only health-related cut-points that were statistically significant at the 95% level were reported here; [a]case-control study design; [b]asthma=cut-point to identify the presence or absence of asthma (i.e., healthy youth vs. youth with asthma); [c]severity=cut-point to identify the severity of asthma (i.e., intermittent asthma vs. moderately persistent asthma); [d]cross-sectional study design; [e]health-related cut-points for MSF were calculated as standardized scores (e.g., standardized by sex, age) and converted to raw scores, with only the raw scores reported here; [f]sit-and-reach cut-points were corrected so that a score of 0 cm corresponded to the participant reaching their toes, with positive values indicating that the participant could reach past their toes; [g]FitnessGram's trunk extension test is thought to measure both muscular strength and flexibility, but was categorized here as a measure of flexibility as specified by Meredith and Welk [54]; [h]despite being outside the target age range of 5–17 years, health-related cut-points developed by Steffl et al. [34] were included because almost all youth met the target age range; [i]longitudinal cohort study design; [j]only sample data from which the health-related cut-points were developed were reported here; no AUC values were available for Peterson et al. [31] because they identified health-related cut-points using conditional inference tree analysis rather than receiver operating characteristic curve analysis; while Castro-Piñero et al. [26] used a longitudinal cohort study design, because the health-related cut-points were developed using the cross-sectional sample, it was classified here as a cross-sectional study design; Saint-Maurice et al. [33] developed health-related cut-points for both single and composite MSF tests, but only cut-points for single MSF tests were reported here.

FIGURE CAPTIONS

Figure 1. PRISMA flow chart outlining the flow of studies through the review.

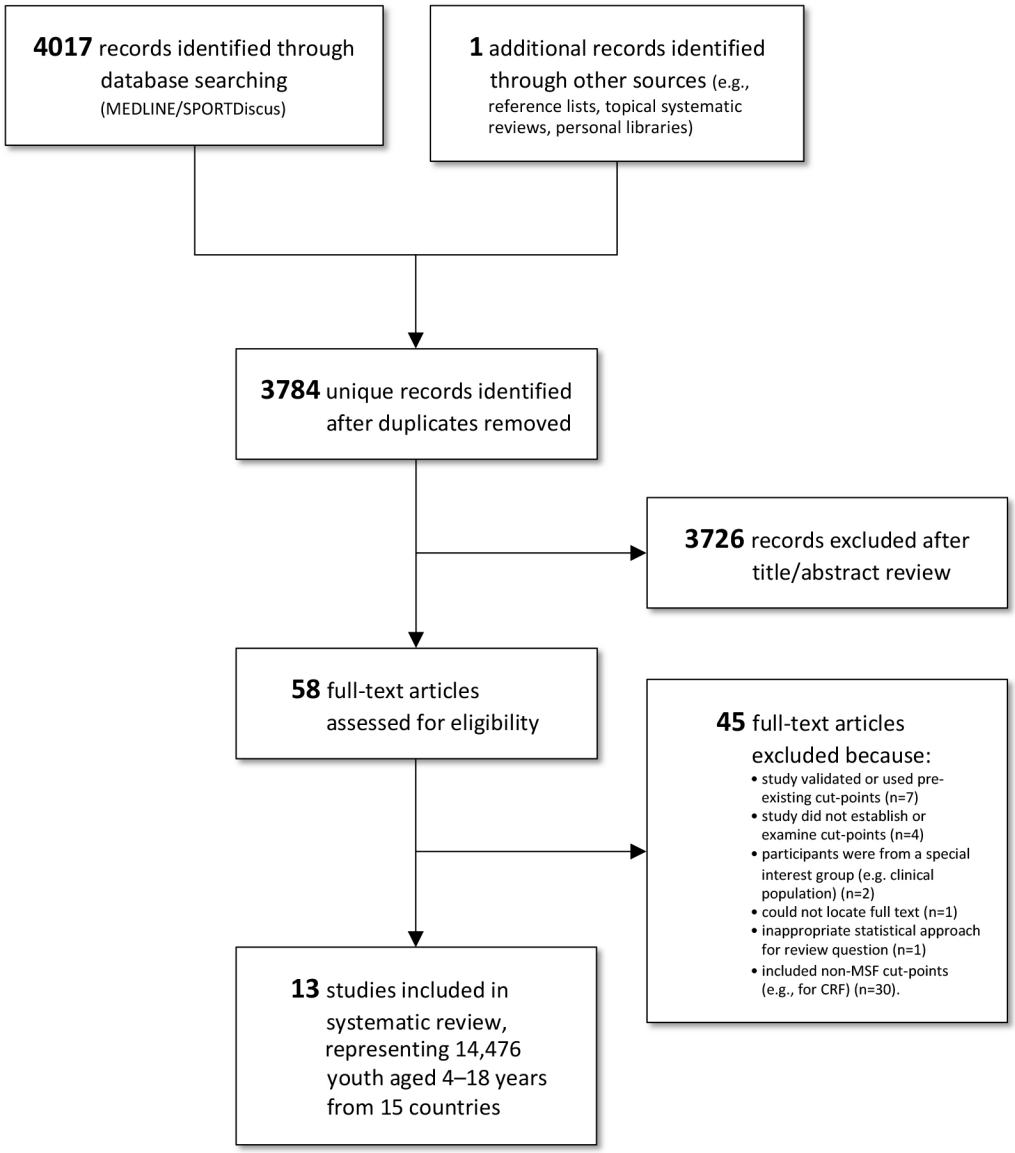
Abbreviations: CRF=cardiorespiratory fitness; MSF=musculoskeletal fitness.

Figure 2. Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) results for risk of bias and applicability criteria (n=13).

Figure 3. Scatterplots of health-related criterion-referenced cut-points for MSF in youth against age.

Abbreviations: BMI=body mass index; cm=centimeters; kg=kilograms; m^2=meters squared; reps=repetitions; s=seconds; W=Watts.

Notes: Only cut-points from (a) Saint-Maurice et al. [35] for high risk of insufficient whole-body less head bone mineral content, (b) Peterson et al. [31] for high cardiometabolic risk, and (c) Latorre-Román et al. [36] for the presence of asthma, were shown. Age-specific cut-points for dynamic curl-ups, static curl-up and back-saver sit-and-reach were estimated by adding 5 to the reported school grade-specific cut-points (e.g., the ages corresponding to grades 4–5 were estimated to be 9–10 years) reported by Saint-Maurice et al. [33]. ▼=asthma (boys+girls); ■=bone health (boys); □=bone health (girls); ●=cardiometabolic risk/metabolic syndrome (boys); ○=cardiometabolic risk/metabolic syndrome (girls); ◆=musculoskeletal pain (boys); ◇=musculoskeletal pain (girls); ▲=sarcopenic obesity risk (boys); △=sarcopenic obesity risk (girls); ×=cognition/academic performance (boys); +=cognition/academic performance (girls).


4017 records identified through database searching (MEDLINE/SPORTDiscus)
1 additional records identified through other sources (e.g., reference lists, topical systematic reviews, personal libraries)
3784 unique records identified after duplicates removed
3726 records excluded after title/abstract review
58 full-text articles assessed for eligibility
45 full-text articles excluded because:
• study validated or used pre-existing cut-points (n=7)
• study did not establish or examine cut-points (n=4)
• participants were from a special interest group (e.g. clinical population) (n=2)
• could not locate full text (n=1)
• inappropriate statistical approach for review question (n=1)
• included non-MSF cut-points (e.g., for CRF) (n=30).
13 studies included in systematic review, representing 14,476 youth aged 4–18 years from 15 countries

1

2 **Figure 1.**

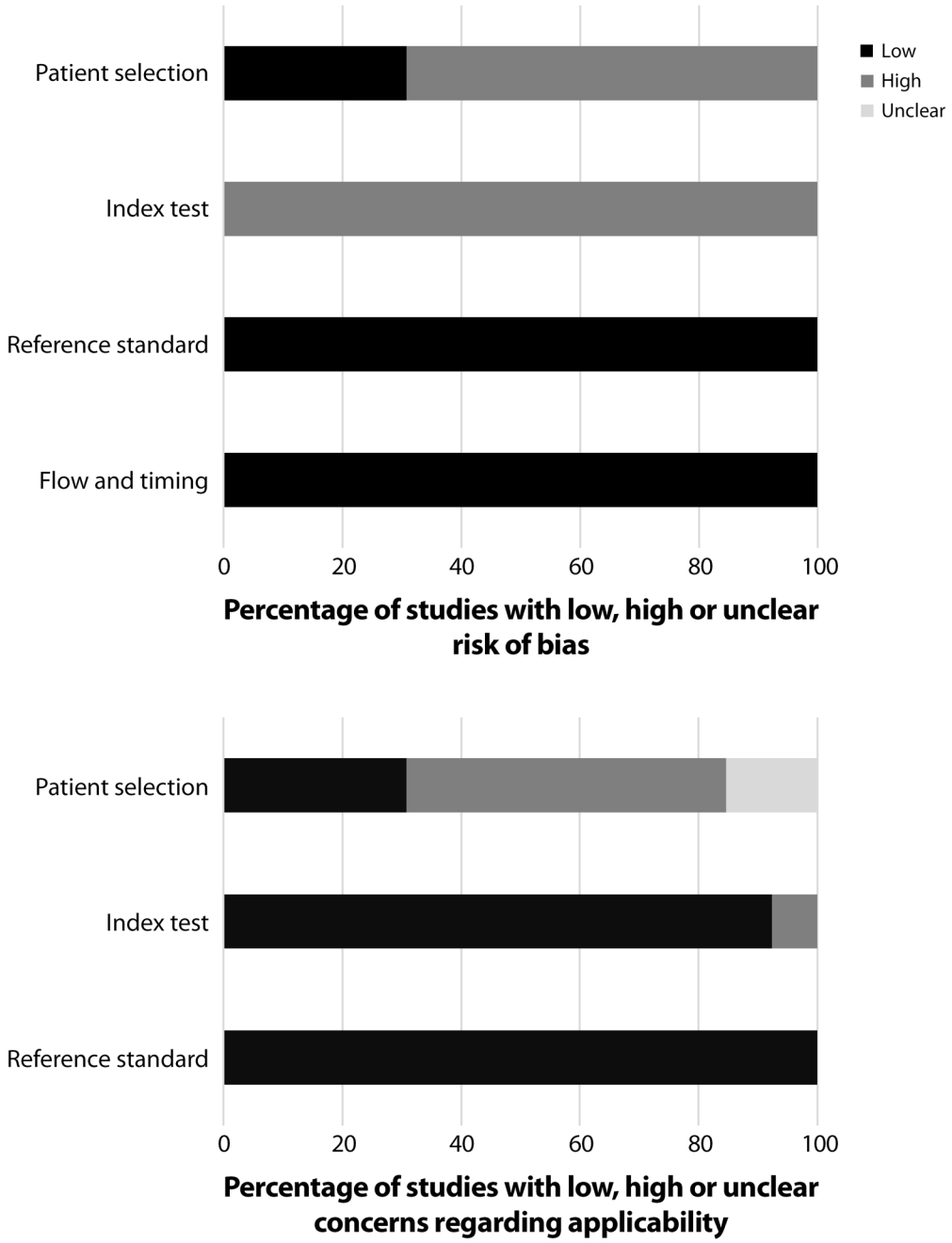

Patient selection
Index test
Reference standard
Flow and timing
Low
High
Unclear
0
20
40
60
80
100
Percentage of studies with low, high or unclear risk of bias
Patient selection
Index test
Reference standard
0
20
40
60
80
100
Percentage of studies with low, high or unclear concerns regarding applicability

Figure 2.

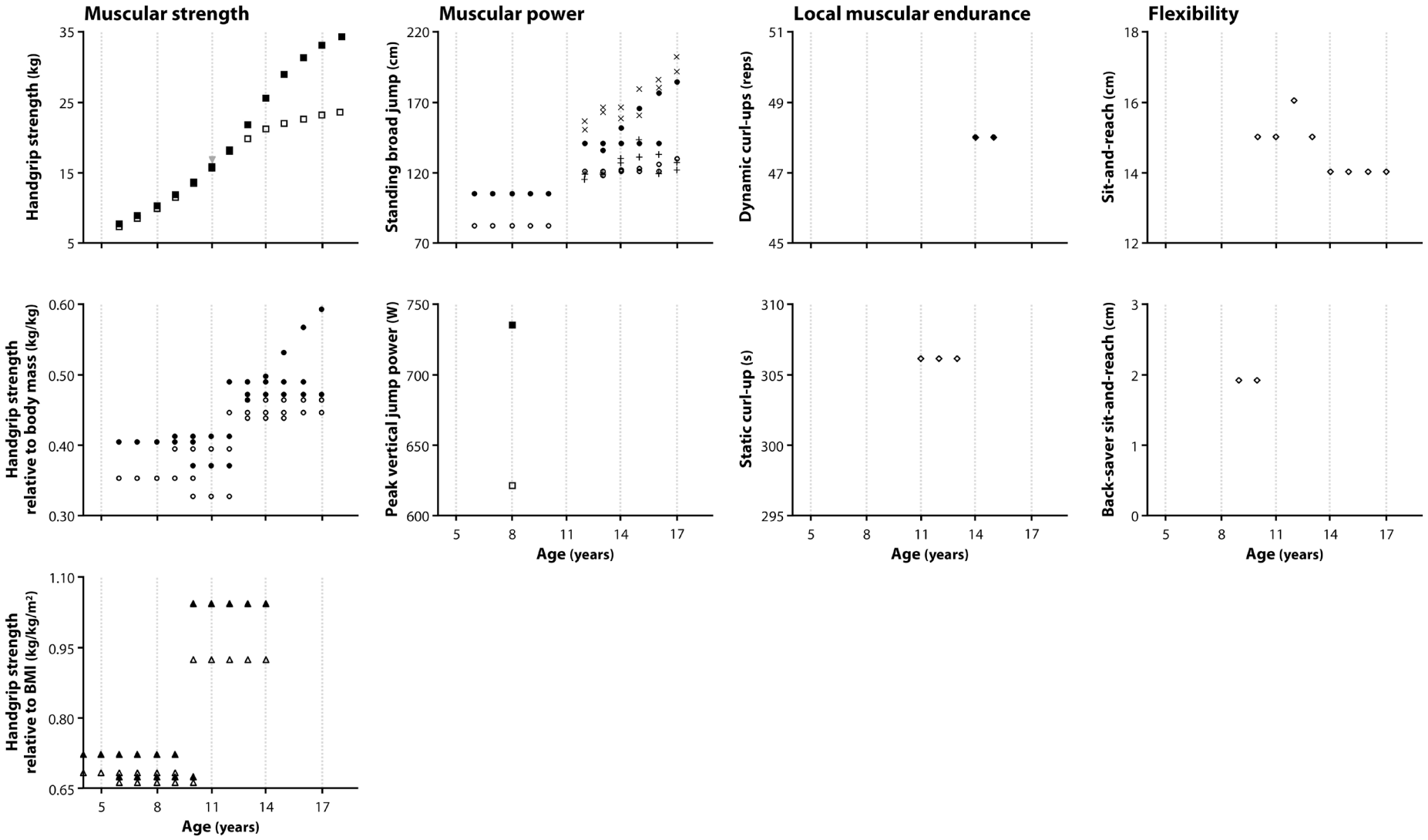

Muscular strength
Muscular power
Local muscular endurance
Flexibility
Handgrip strength (kg)
Standing broad jump (cm)
Dynamic curl-ups (reps)
Sit-and-reach (cm)
Handgrip strength relative to body mass (kg/kg)
Peak vertical jump power (W)
Static curl-up (s)
Back-saver sit-and-reach (cm)
Handgrip strength relative to BMI (kg/kg/m²)
Age (years)

Figure 3.

COMPLIANCE WITH ETHICAL STANDARDS

Funding

This systematic review was funded by the Public Health Agency of Canada (#4500414578).

Conflicts of Interest

Brooklyn J. Fraser, Scott Rollo, Margaret Sampson, Costan G. Magnussen, Justin J. Lang, Mark S. Tremblay, and Grant R. Tomkinson declare they have no conflicts of interest.

Author Contributions

BJF screened the records, extracted the data and took responsibility for the integrity of the data, synthesized the results, and co-wrote the manuscript. ASR screened the records, checked the data for accuracy and took responsibility for the integrity of the data, contributed to the interpretation of results, and critically reviewed the manuscript for important intellectual content. MS designed and executed the systematic search strategy, and critically reviewed the manuscript for important intellectual content. CGM contributed to the interpretation of results and critically reviewed the manuscript for important intellectual content. JJL developed the research question, designed the systematic review, contributed to the interpretation of results, and critically reviewed the manuscript for important intellectual content. MST developed the research question, designed the systematic review, contributed to the interpretation of results, and critically reviewed the manuscript for important intellectual content. GRT developed the research question, designed the systematic review, had full access to the data, synthesized the results, and co-wrote the manuscript. All authors have read and approved the final version of the manuscript, agree to be accountable for all aspects of the work, and agree with the order of presentation of the authors.

ELECTRONIC SUPPLEMENTARY MATERIAL APPENDIX S1

Article title: Health-related criterion-referenced cut-points for musculoskeletal fitness among youth: A systematic review; **Journal name:** Sports Medicine; **Author names and affiliations:** Brooklyn J. Fraser (University of Tasmania), Scott Rollo (Children's Hospital of Eastern Ontario Research Institute), Margaret Sampson (Children's Hospital of Eastern Ontario Research Institute), Costan G. Magnussen (University of Tasmania and University of Turku), Justin J. Lang (Public Health Agency of Canada and Children's Hospital of Eastern Ontario Research Institute), Mark S. Tremblay (Children's Hospital of Eastern Ontario Research Institute), and Grant R. Tomkinson (University of North Dakota and University of South Australia); **E-mail addresses of the corresponding authors:** fraserbj@utas.edu.au and grant.tomkinson@und.edu.

Supplement 1. Search strategy for databases.

Databases searched: MEDLINE and SPORTDiscus.

Searches performed: 3rd and 4th of September 2020.

Search strategy 1: OVID MEDLINE, 1946 to 3rd of September 2020.

Search string 1:

1 Cardiorespiratory fitness/ or ((Shuttle adj2 (run* or test)) or SRT or 20mSRT or ((Beep or Bleep or bip) adj3 test*) or PACER or Progressive Aerobic Cardiovascular Endurance or FITNESSGRAM or Eurofit test battery or Leger test or course navette or test Leger navette or test navette or ((Multi-stage or multistage) adj2 fitness) or Prefit).ti,ab,kf.

2 (((Multi-stage or multistage) adj2 test) or ((20 m or 20m or 20 metre) adj3 (test* or run)) or ((minute or mile or distance or timed) adj2 run) or ((fitness or exercise or endurance or step) adj2 test) or ((functional or aerobic or exercise) adj2 (capacity or

endurance)) or ((Submaximal or maximal or graded) adj2 (treadmill or ergometer)) or cardiorespiratory* or cardiovascular* or cardiopulmonary or endurance or aerobic* or maxim* oxygen or peak oxygen or VO2*).ti,ab,kf. and fit*.mp.

3 exp Muscle Strength/ or Muscle Strength Dynamometer/ or ((Musc* adj2 fit*) or handgrip strength or grip strength or dynamomet* or ((standing or sargent or sergeant or vertical or squat) adj2 jump) or Bosco or "sit-and-reach" or back-saver or shoulder stretch or push up* or pushup* or sit up* or situp* or curl up* or curl up* or plank or jump* height or trunk lift or (arm adj (hang* or suspension*)) or pull up* or pullup* or chin up* or chinup*).ti,ab,kf.

4 (strength or power or agility or speed or flexibility).ti,ab,kf. and fit*.mp.

5 (criterion-based or criterion-referenced or ROC curve or receiver operat* or AUC or "Area under the curve" or conditional inference or risk categorization or likelihood ratio or logistic regression or threshold or cut off or cut point or (fitness adj (category or classification)) or "sensitivity and specificity" or clinical* relevan*).ti,ab,kf. or Risk Factors/

6 (child* or adolesc* or youth or pediatr* or paediatr*).mp.

7 (1 or 2 or 3 or 4) and 5 and 6

Search strategy 2: EBSCOhost SPORTDiscus, searched 4th of September 2020.

Search string 2:

S1 DE "CARDIOPULMONARY fitness measurement" or DE "CARDIOVASCULAR fitness" OR DE "PHYSICAL fitness" OR DE "CARDIOPULMONARY fitness" OR ((TI Shuttle OR AB Shuttle N2 (TI run* OR AB run* OR TI test OR AB test)) OR TI SRT OR AB SRT OR TI 20mSRT OR AB 20mSRT OR ((TI Beep OR AB Beep OR TI Bleep OR AB Bleep OR TI bip OR AB bip) N3 (TI test* OR AB test*)) OR TI

1		PACER OR AB PACER OR TI "Progressive Aerobic Cardiovascular Endurance" OR
2		AB "Progressive Aerobic Cardiovascular Endurance" OR TI FITNESSGRAM OR AB
3		FITNESSGRAM OR TI "Eurofit test battery" OR AB "Eurofit test battery" OR TI
4		"Leger test" OR AB "Leger test" OR TI "course navette" OR AB "course navette" OR
5		TI "test Leger navette" OR AB "test Leger navette" OR TI "test navette" OR AB "test
6		navette" OR ((TI Multi-stage OR AB Multi-stage OR TI multistage OR AB multistage)
7		N2 (TI fitness OR AB fitness)) OR TI Prefit OR AB Prefit)
8	S2	(((TI Multi-stage OR AB Multi-stage OR TI multistage OR AB multistage) N2 (TI test
9		OR AB test)) OR ((TI m OR AB m OR TI 20m OR AB 20m OR TI metre OR AB
10		metre) N3 (TI test* OR AB test* OR TI run OR AB run)) OR ((TI minute OR AB
11		minute OR TI mile OR AB mile OR TI distance OR AB distance OR TI timed OR AB
12		timed) N2 (TI run OR AB run)) OR ((TI fitness OR AB fitness OR TI exercise OR AB
13		exercise OR TI endurance OR AB endurance OR TI step OR AB step) N2 (TI test OR
14		AB test)) OR ((TI functional OR AB functional OR TI aerobic OR AB aerobic OR TI
15		exercise OR AB exercise) N2 (TI capacity OR AB capacity OR TI endurance OR AB
16		endurance)) OR ((TI Submaximal OR AB Submaximal OR TI maximal OR AB
17		maximal OR TI graded OR AB graded) N2 (TI treadmill OR AB treadmill OR TI
18		ergometer OR AB ergometer)) OR TI cardiorespiratory* OR AB cardiorespiratory* OR
19		TI cardiovascular* OR AB cardiovascular* OR TI cardiopulmonary OR AB
20		cardiopulmonary OR TI endurance OR AB endurance OR TI aerobic* OR AB aerobic*
21		OR TI "maxim* oxygen" OR AB "maxim* oxygen" OR TI "peak oxygen" OR AB
22		"peak oxygen" OR TI VO2* OR AB VO2*) AND fit*
23	S3	DE "PHYSICAL fitness testing for children" or DE "PHYSICAL fitness testing " OR
24		DE "MUSCLE strength" OR DE "MUSCLE strength testing" OR DE "MUSCLE
25		strength measurement" or DE "ERGOMETRY" OR DE "DYNAMOMETER" OR ((TI

1		Musc* OR AB Musc* N2 TI fit* OR AB fit*) OR TI "handgrip strength" OR AB
2		"handgrip strength" OR TI "grip strength" OR AB "grip strength" OR TI dynamomet*
3		OR AB dynamomet* OR ((TI standing OR AB standing OR TI sargent OR AB sargent
4		OR TI sergeant OR AB sergeant OR TI vertical OR AB vertical OR TI squat OR AB
5		squat) N2 (TI jump OR AB jump)) OR TI Bosco OR AB Bosco OR TI sit-and-reach
6		OR AB sit-and-reach OR TI back-saver OR AB back-saver OR TI "shoulder stretch"
7		OR AB "shoulder stretch" OR TI "push up*" OR AB "push up*" OR TI pushup* OR
8		AB pushup* OR TI "sit up*" OR AB "sit up*" OR TI situp* OR AB situp* OR TI "curl
9		up*" OR AB "curl up*" OR TI "curl up*" OR AB "curl up*" OR TI plank OR AB
10		plank OR TI "jump* height" OR AB "jump* height" OR TI "trunk lift*" OR AB "trunk
11		lift"" OR (TI "arm" OR AB "arm" N1 (TI hang* OR AB hang* OR TI suspension* OR
12		AB suspension*)) OR TI "pull up*" OR AB "pull up*" OR TI pullup* OR AB pullup*
13		OR TI "chin up*" OR AB "chin up*" OR TI chinup* OR AB chinup*)
14	S4	(TI strength OR AB strength OR TI power OR AB power OR TI agility OR AB agility
15		OR TI speed OR AB speed OR TI flexibility OR AB flexibility) AND fit*
16	S5	(TI criterion-based OR AB criterion-based OR TI criterion-referenced OR AB
17		criterion-referenced OR TI "ROC curve" OR AB "ROC curve" OR TI "receiver
18		operat*" OR AB "receiver operat*" OR TI AUC OR AB AUC OR TI "Area under the
19		curve" OR AB "Area under the curve" OR TI "conditional inference" OR AB
20		"conditional inference" OR TI "risk categorization" OR AB "risk categorization" OR TI
21		"likelihood ratio" OR AB "likelihood ratio" OR TI "logistic regression" OR AB
22		"logistic regression" OR TI threshold OR AB threshold OR TI "cut off" OR AB "cut
23		off" OR TI "cut point" OR AB "cut point" OR ((TI "fitness" OR AB "fitness") N1 (TI
24		category OR AB category OR TI classification OR AB classification)) OR TI

"sensitivity and specificity" OR AB "sensitivity and specificity" OR TI "clinical* relevan*" OR AB "clinical* relevan*") OR (risk factors)

S6 (child* OR adolesc* OR youth OR pediatr* OR paediatr*)

S7 (S1 or S2 or S3 or S4) and S5 and S6

COMPLIANCE WITH ETHICAL STANDARDS

Funding

This systematic review was funded by the Public Health Agency of Canada (#4500414578).

Conflicts of Interest

Brooklyn J. Fraser, Scott Rollo, Margaret Sampson, Costan G. Magnussen, Justin J. Lang, Mark S. Tremblay, and Grant R. Tomkinson declare they have no conflicts of interest.

Author Contributions

BJF screened the records, extracted the data and took responsibility for the integrity of the data, synthesized the results, and co-wrote the manuscript. ASR screened the records, checked the data for accuracy and took responsibility for the integrity of the data, contributed to the interpretation of results, and critically reviewed the manuscript for important intellectual content. MS designed and executed the systematic search strategy, and critically reviewed the manuscript for important intellectual content. CGM contributed to the interpretation of results and critically reviewed the manuscript for important intellectual content. JJL developed the research question, designed the systematic review, contributed to the interpretation of results, and critically reviewed the manuscript for important intellectual content. MST developed the research question, designed the systematic review, contributed to the interpretation of results, and critically reviewed the manuscript for important intellectual content. GRT developed the research question, designed the systematic review, had full access to the data, synthesized the results, and co-wrote the manuscript. All authors have read and approved the final version of the manuscript, agree to be accountable for all aspects of the work, and agree with the order of presentation of the authors.

ELECTRONIC SUPPLEMENTARY MATERIAL APPENDIX S2

Article title: Health-related criterion-referenced cut-points for musculoskeletal fitness among youth: A systematic review; **Journal name:** Sports Medicine; **Author names and affiliations:** Brooklyn J. Fraser (University of Tasmania), Scott Rollo (Children's Hospital of Eastern Ontario Research Institute), Margaret Sampson (Children's Hospital of Eastern Ontario Research Institute), Costan G. Magnussen (University of Tasmania and University of Turku), Justin J. Lang (Public Health Agency of Canada and Children's Hospital of Eastern Ontario Research Institute), Mark S. Tremblay (Children's Hospital of Eastern Ontario Research Institute), and Grant R. Tomkinson (University of North Dakota and University of South Australia); **E-mail addresses of the corresponding authors:** fraserbj@utas.edu.au and grant.tomkinson@und.edu.

Supplement 2. Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) results for signaling questions, risk of bias and applicability criteria (n=13).

	Patient selection					Index test				Reference standard				Flow/timing				
Study	**Consecutive or random sample**	**Avoided case-control**	**Avoided inappropriate exclusions**	**Overall risk of bias**	**Concern regarding applicability**	**Interpreted blind to reference standard**	**Threshold pre-specified**	**Overall risk of bias**	**Concern regarding applicability**	**Correctly classifies target condition**	**Interpreted blind to index test**	**Overall risk of bias**	**Concern regarding applicability**	**Appropriate interval between index test and reference standard**	**All participants received a reference standard**	**All participants received same reference standard**	**All participants included in analysis**	**Overall risk of bias**
Latorre-Román et al. [36]	yes	no	yes	high	high	no	no	high	high	yes	yes	low	low	yes	yes	yes	yes	low
Saint-Maurice et al. [33]	no	yes	yes	high	high	unclear	no	high	low	yes	yes	low	low	yes	yes	yes	yes	low
Peterson et al. [31]	yes	yes	yes	low	low	yes	no	high	low	yes	yes	low	low	yes	yes	yes	yes	low
Dorneles et al. [27]	yes	yes	yes	low	low	yes	no	high	low	yes	yes	low	low	yes	yes	yes	yes	low
Baptista et al. [24]	no	yes	yes	high	high	yes	no	high	low	yes	yes	low	low	yes	yes	yes	yes	low
Steffl et al. [34]	no	yes	yes	high	high	yes	no	high	low	yes	yes	low	low	yes	yes	yes	yes	low
Ramirez-Velez et al. [32]	no	yes	yes	high	unclear	yes	no	high	low	yes	yes	low	low	yes	yes	yes	yes	low
Saint-Maurice et al. [35]	no	yes	yes	high	high	yes	no	high	low	yes	yes	low	low	yes	yes	yes	yes	low
Castro-Piñero et al. [25]	yes	yes	yes	low	low	yes	no	high	low	yes	yes	low	low	yes	yes	yes	yes	low
Martínez-López et al. [29]	no	yes	yes	high	unclear	yes	no	high	low	yes	yes	low	low	yes	yes	yes	yes	low
Castro-Piñero et al. [26]	no	yes	yes	high	high	yes	no	high	low	yes	yes	low	low	yes	yes	yes	yes	low
Mazzoccante et al. [30]	no	yes	yes	high	high	yes	no	high	low	yes	yes	low	low	yes	yes	yes	yes	low
Gontarev et al. [28]	yes	yes	yes	low	low	yes	no	high	low	yes	yes	low	low	yes	yes	yes	yes	low

COMPLIANCE WITH ETHICAL STANDARDS

Funding

This systematic review was funded by the Public Health Agency of Canada (#4500414578).

Conflicts of Interest

Brooklyn J. Fraser, Scott Rollo, Margaret Sampson, Costan G. Magnussen, Justin J. Lang, Mark S. Tremblay, and Grant R. Tomkinson declare they have no conflicts of interest.

Author Contributions

BJF screened the records, extracted the data and took responsibility for the integrity of the data, synthesized the results, and co-wrote the manuscript. ASR screened the records, checked the data for accuracy and took responsibility for the integrity of the data, contributed to the interpretation of results, and critically reviewed the manuscript for important intellectual content. MS designed and executed the systematic search strategy, and critically reviewed the manuscript for important intellectual content. CGM contributed to the interpretation of results and critically reviewed the manuscript for important intellectual content. JJL developed the research question, designed the systematic review, contributed to the interpretation of results, and critically reviewed the manuscript for important intellectual content. MST developed the research question, designed the systematic review, contributed to the interpretation of results, and critically reviewed the manuscript for important intellectual content. GRT developed the research question, designed the systematic review, had full access to the data, synthesized the results, and co-wrote the manuscript. All authors have read and approved the final version of the manuscript, agree to be accountable for all aspects of the work, and agree with the order of presentation of the authors.

ELECTRONIC SUPPLEMENTARY MATERIAL APPENDIX S3

Article title: Health-related criterion-referenced cut-points for musculoskeletal fitness among youth: A systematic review; **Journal name:** Sports Medicine; **Author names and affiliations:** Brooklyn J. Fraser (University of Tasmania), Scott Rollo (Children's Hospital of Eastern Ontario Research Institute), Margaret Sampson (Children's Hospital of Eastern Ontario Research Institute), Costan G. Magnussen (University of Tasmania and University of Turku), Justin J. Lang (Public Health Agency of Canada and Children's Hospital of Eastern Ontario Research Institute), Mark S. Tremblay (Children's Hospital of Eastern Ontario Research Institute), and Grant R. Tomkinson (University of North Dakota and University of South Australia); **E-mail addresses of the corresponding authors:** fraserbj@utas.edu.au and grant.tomkinson@und.edu

Supplement 3. Measures of diagnostic accuracy.

Study	Health indicator	Health criterion	MSF component	MSF test	AUC (95%CI)	Sensitivity	Specificity	Youden's J	PPV	NPV	LR+	LR−
Latorre-Román et al. [36][a]	Asthma	Diagnosed according to the GINA 2006 criteria [52]	Muscular strength	Handgrip strength	0.73 (0.64–0.81) (asthma)[b] 0.65 (0.52–0.79) (severity)[c]	52.9 (asthma) 53.6 (severity)	70.0 (asthma) 66.7 (severity)					
Saint-Maurice et al. [33][d]	Musculoskeletal pain (self-reported LBP)	Indicated presence of LBP on ≥7 of 14 questionnaire items	Muscular endurance	Dynamic curl-ups	≤0.59 (Boys, G4–5) ≤0.59 (Boys, G6–8) 0.75 (Boys, G9–10) 0.61 (Girls, G4–5) 0.63 (Girls, G6–8) 0.70 (0.50–0.83) (Girls, G9–10)	75.0 (Boys, G9–10)	74.1 (Boys, G9–10)					
				Static curl-up	≤0.59 (Boys, G4–5) 0.61 (Boys, G6–8) 0.63 (Boys, G9–10) 0.67 (Girls, G4–5) 0.71 (0.56–0.84) (Girls, G6–8) 0.73 (0.54–0.88) (Girls, G9–10)	82.4 (Girls, G6–8)	64.3 (Girls, G6–8)					
				Lateral plank	≤0.59 (Boys, G4–5) ≤0.59 (Boys, G6–8) 0.50 (Boys, G9–10) ≤0.51 (Girls, G4–5) 0.52 (Girls, G6–8) 0.71 (0.51–0.86) (Girls, G9–10)							
				Prone plank	≤0.59 (Boys, G4–5) ≤0.59 (Boys, G6–8) 0.59 (Boys, G9–10) ≤0.51 (Girls, G4–5) 0.50 (Girls, G6–8) 0.60 (Girls, G9–10)							
			Flexibility	Back-saver sit-and-reach	≤0.59 (Boys, G4–5) ≤0.59 (Boys, G6–8) 0.58 (Boys, G9–10) 0.80 (0.66–0.90) (Girls, G4–5) 0.68 (Girls, G6–8) 0.53 (Girls, G9–10)	75.0 (Girls, G4–5)	81.4 (Girls, G4–5)					
				Trunk extension[e]	0.52 (0.37–0.67) (Boys, G4–5) 0.58 (0.45–0.71) (Boys, G6–8) 0.61 (0.42–0.78) (Boys, G9–10) 0.59 (0.44–0.71) (Girls, G4–5) 0.56 (0.40–0.71) (Girls, G6–8) 0.52 (0.33–0.71) (Girls, G9–10)							
Peterson et al. [31][d]	Cardiometabolic risk (%BF, SBP, triglycerides, HDL-C and glucose)	≥75th percentile ("high risk") of a continuous standardized score calculated with all cardiometabolic risk factors	Muscular strength	Handgrip strength (relative to body mass)								
Domeles et al. [27][d]	Musculoskeletal pain (self-reported LBP)	Indicated presence of LBP/discomfort as "often and always"	Muscular endurance	Sit-ups	0.53 (0.46–0.60) (Boys) 0.52 (0.46–0.57) (Girls)							
			Flexibility	Sit-and-reach	0.49 (0.42–0.56) (Boys) 0.58 (0.52–0.63) (Girls)	53.5 (Girls)	62.0 (Girls)					
Baptista et al. [24][d]	Bone health (BMD)	≤–1 SD ("increased risk") of height-adjusted WBLH BMD	Muscular power	Peak vertical jump power (calculated from counter-movement jump height, age, body mass and height)	0.82 (0.68–0.95) (Boys) 0.85 (0.70–1.00) (Girls)	63.3 (Boys) 75.0 (Girls)	69.2 (Boys) 77.0 (Girls)					

Steffl et al. [34][d]	Sarcopenic obesity risk	≤–2 SD ("increased risk") of mean muscle-to-fat ratio for the 3rd BMI quintile	Muscular strength	Handgrip strength (relative to BMI)	0.72 (0.61–0.83) (Boys, 4–9 y) 0.90 (0.82–0.97) (Boys, 10–14 y) 0.79 (0.69–0.89) (Girls, 4–9 y) 0.79 (0.69–0.89) (Girls, 10–14 y)	75.0 (Boys, 4–9 y) 91.7 (Boys, 10–14 y) 76.5 (Girls, 4–9 y) 61.9 (Girls, 10–14 y)	57.7 (Boys, 4–9 y) 80.2 (Boys, 10–14 y) 60.2 (Girls, 4–9 y) 85.1 (Girls, 10–14 y)					
Ramirez-Velez et al. [32][d]	Cardiometabolic risk (WC, SBP, DBP, triglycerides, HDL-C and glucose)	≥1 SD ("increased risk") of a continuous standardized score computed with all cardiometabolic risk factors	Muscular strength	Handgrip strength (relative to body mass)	0.84 (0.74–0.94) (Boys, 9–12 y) 0.80 (0.68–0.92) (Boys, 13–17 y) 0.83 (0.71–0.95) (Girls, 9–12 y) 0.79 (0.70–0.89) (Girls, 13–17 y)	80.0 (Boys, 9–12 y) 78.9 (Boys, 13–17 y) 78.6 (Girls, 9–12 y) 92.9 (Girls, 13–17 y)	81.5 (Boys, 9–12 y) 77.0 (Boys, 13–17 y) 81.9 (Girls, 9–12 y) 55.1 (Girls, 13–17 y)	0.62 (Boys, 9–12 y) 0.56 (Boys, 13–17 y) 0.61 (Girls, 9–12 y) 0.48 (Girls, 13–17 y)			4.3 (Boys, 9–12 y) 3.4 (Boys, 13–17 y) 4.3 (Girls, 9–12 y) 2.1 (Girls, 13–17 y)	0.25 (Boys, 9–12 y) 0.27 (Boys, 13–17 y) 0.26 (Girls, 9–12 y) 0.13 (Girls, 13–17 y)
Saint-Maurice et al. [35][f,g]	Bone health (BMC)	≤–0.84 SD ("some risk") and <–1.5 SD ("high risk") of height-adjusted WBLH BMC	Muscular strength	Handgrip strength	High risk: 0.80 (0.77–0.83) (Boys) 0.79 (0.76–0.81) (Girls) Some risk: 0.71 (0.67–0.74) (Boys) 0.81 (0.79–0.84) (Girls)	High risk: 80.0 (28.4–99.5) (Boys) 63.6 (30.8–89.1) (Girls) Some risk: 81.4 (66.6–91.6) (Boys) 92.5 (79.6–98.4) (Girls)	High risk: 82.4 (79.5–85.0) (Boys) 80.9 (78.1–83.5) (Girls) Some risk: 53.6 (49.9–57.2) (Boys) 54.1 (50.6–57.6) (Girls)					
Castro-Piñero et al. [25][d]	MetS and cardiometabolic risk (WC, SBP, DBP, triglycerides, HDL-C, glucose and insulin)	MetS: diagnosed according to the Jolliffe and Janssen [53] criteria (i.e., ≥3 increased cardiometabolic risk factors); and cardiometabolic risk: >1 SD ("increased risk") of a continuous standardized score computed with all cardiometabolic risk factors	Muscular strength	Handgrip strength	MetS: 0.56 (0.52–0.61) (Boys) 0.54 (0.50–0.58) (Girls) Cardiometabolic risk: 0.66 (0.61–0.70) (Boys) 0.64 (0.59–0.68) (Girls)	MetS: 72.7 (Boys) Cardiometabolic risk: 66.7 (Boys) 84.2 (Girls)	MetS: 49.1 (Boys) Cardiometabolic risk: 59.9 (Boys) 40.8 (Girls)	MetS: 0.22 (Boys) Cardiometabolic risk: 0.27 (Boys) 0.25 (Girls)	MetS: 3.5 (Boys) Cardiometabolic risk: 10.8 (Boys) 10.4 (Girls)	MetS: 98.6 (Boys) Cardiometabolic risk: 96.1 (Boys) 97.0 (Girls)		
				Handgrip strength (relative to body mass)	MetS: 0.87 (0.84–0.90) (Boys) 0.76 (0.72–0.79) (Girls) Cardiometabolic risk: 0.83 (0.79–0.87) (Boys) 0.75 (0.71–0.79) (Girls)	MetS: 72.7 (Boys) 63.2 (Girls) Cardiometabolic risk: 80.0 (Boys) 65.8 (Girls)	MetS: 87.7 (Boys) 76.9 (Girls) Cardiometabolic risk: 79.7 (Boys) 78.0 (Girls)	MetS: 0.60 (Boys) 0.40 (Girls) Cardiometabolic risk: 0.60 (Boys) 0.44 (Girls)	MetS: 12.9 (Boys) 9.4 (Girls) Cardiometabolic risk: 22.2 (Boys) 19.5 (Girls)	MetS: 99.2 (Boys) 98.2 (Girls) Cardiometabolic risk: 98.2 (Boys) 96.6 (Girls)		
			Muscular power	Standing broad jump	MetS: 0.75 (0.71–0.79) (Boys) 0.66 (0.62–0.70) (Girls) Cardiometabolic risk: 0.78 (0.73–0.81) (Boys) 0.69 (0.65–0.73) (Girls)	MetS: 63.6 (Boys) 57.9 (Girls) Cardiometabolic risk: 73.3 (Boys) 44.7 (Girls)	MetS: 80.8 (Boys) 78.7 (Girls) Cardiometabolic risk: 80.7 (Boys) 83.1 (Girls)	MetS: 0.44 (Boys) 0.37 (Girls) Cardiometabolic risk: 0.54 (Boys) 0.28 (Girls)	MetS: 7.7 (Boys) 9.3 (Girls) Cardiometabolic risk: 21.6 (Boys) 17.7 (Girls)	MetS: 98.9 (Boys) 98.0 (Girls) Cardiometabolic risk: 97.7 (Boys) 94.9 (Girls)		
				Standing broad jump (relative to body mass)	MetS: 0.72 (0.68–0.76) (Boys) 0.68 (0.63–0.72) (Girls) Cardiometabolic risk: 0.70 (0.66–0.74) (Boys) 0.70 (0.66–0.74) (Girls)	MetS: 72.7 (Boys) 68.4 (Girls) Cardiometabolic risk: 70.0 (Boys) 57.9 (Girls)	MetS: 61.2 (Boys) 78.4 (Girls) Cardiometabolic risk: 62.1 (Boys) 79.5 (Girls)	MetS: 0.34 (Boys) 0.47 (Girls) Cardiometabolic risk: 0.32 (Boys) 0.37 (Girls)	MetS: 4.5 (Boys) 10.7 (Girls) Cardiometabolic risk: 11.8 (Boys) 18.6 (Girls)	MetS: 98.9 (Boys) 98.5 (Girls) Cardiometabolic risk: 96.6 (Boys) 95.9 (Girls)		
Martínez-López et al. [29][d]	Cognition/academic performance	≥7 of 10 ("high performance") for grades in mathematics and language (combined) and physical education	Muscular power	Standing broad jump	Mathematics+language grades: 0.58 (Boys, 12 y) 0.54 (Boys, 13 y) 0.53 (Boys, 14 y) 0.50 (Boys, 15 y) 0.56 (Boys, 16 y) 0.58 (Boys, 17 y) 0.58 (Girls, 12 y) 0.55 (Girls, 13 y) 0.64 (Girls, 14 y) 0.65 (Girls, 15 y)							

					0.53 (Girls, 16 y) 0.61 (Girls, 17 y) Physical education grades: 0.68 (Boys, 12 y) 0.55 (Boys, 13 y) 0.63 (Boys, 14 y) 0.66 (Boys, 15 y) 0.66 (Boys, 16 y) 0.72 (Boys, 17 y) 0.75 (Girls, 12 y) 0.73 (Girls, 13 y) 0.63 (Girls, 14 y) 0.71 (Girls, 15 y) 0.68 (Girls, 16 y) 0.71 (Girls, 17 y)							
Castro-Piñero et al. [26][f g]	Cardiometabolic risk (sum of 2 skinfolds, SBP, HOMA-IR, triglycerides, total cholesterol-to-HDL-C ratio)	>1 SD ("increased risk") of a continuous standardized score computed with all cardiometabolic risk factors	Muscular strength	Handgrip strength (relative to body mass)	0.75 (0.62–0.88) (Boys, 6–10 y) 0.75 (0.60–0.89) (Boys, 12–16 y) 0.86 (0.75–0.96) (Girls, 6–10 y) 0.74 (0.63–0.85) (Girls, 12–16 y)	80.9 (Boys, 6–10 y) 84.7 (Boys, 12–16 y) 93.4 (Girls, 6–10 y) 82.1 (Girls, 12–16 y)	61.5 (Boys, 6–10 y) 64.5 (Boys, 12–16 y) 67.7 (Girls, 6–10 y) 61.7 (Girls, 12–16 y)	0.42 (Boys, 6–10 y) 0.49 (Boys, 12–16 y) 0.61 (Girls, 6–10 y) 0.44 (Girls, 12–16 y)	5.0 (Boys, 6–10 y) 5.4 (Boys, 12–16 y) 3.4 (Girls, 6–10 y) 3.4 (Girls, 12–16 y)	75.0 (Boys, 6–10 y) 65.5 (Boys, 12–16 y) 75.7 (Girls, 6–10 y) 75.7 (Girls, 12–16 y)	2.1 (Boys, 6–10 y) 2.4 (Boys, 12–16 y) 2.9 (Girls, 6–10 y) 2.1 (Girls, 12–16 y)	0.31 (Boys, 6–10 y) 0.24 (Boys, 12–16 y) 0.10 (Girls, 6–10 y) 0.29 (Girls, 12–16 y)
			Muscular power	Standing broad jump	0.83 (0.72–0.94) (Boys, 6–10 y) 0.67 (0.53–0.80) (Boys, 12–16 y) 0.74 (0.61–0.87) (Girls, 6–10 y) 0.68 (0.54–0.79) (Girls, 12–16 y)	71.3 (Boys, 6–10 y) 74.4 (Boys, 12–16 y) 90.3 (Girls, 6–10 y) 86.1 (Girls, 12–16 y)	81.8 (Boys, 6–10 y) 67.8 (Boys, 12–16 y) 56.3 (Girls, 6–10 y) 60.7 (Girls, 12–16 y)	0.53 (Boys, 6–10 y) 0.42 (Boys, 12–16 y) 0.47 (Girls, 6–10 y) 0.47 (Girls, 12–16 y)	37.5 (Boys, 6–10 y) 37.5 (Boys, 12–16 y) 43.8 (Girls, 6–10 y) 38.0 (Girls, 12–16 y)	97.6 (Boys, 6–10 y) 90.2 (Boys, 12–16 y) 90.3 (Girls, 6–10 y) 88.1 (Girls, 12–16 y)	3.9 (Boys, 6–10 y) 2.3 (Boys, 12–16 y) 2.1 (Girls, 6–10 y) 2.2 (Girls, 12–16 y)	0.35 (Boys, 6–10 y) 0.38 (Boys, 12–16 y) 0.17 (Girls, 6–10 y) 0.23 (Girls, 12–16 y)
Mazzoccante et al. [30][d]	Cognition/academic performance	Attention and executive function based on attention cancellation test and trail making test performance	Muscular power	Squat jump	Attention: 0.44 (0.36–0.52) Executive function: 0.33 (0.26–0.41)							
				Countermovement jump	Attention: 0.42 (0.34–0.50) Executive function: 0.30 (0.23–0.36)							
Gontarev et al. [28][d]	Sarcopenic obesity risk	≤–2 SD ("increased risk") of mean muscle-to-fat ratio for the 3rd BMI quintile	Muscular strength	Handgrip strength (relative to BMI)	0.84 (0.83–0.86) (Boys) 0.77 (0.75–0.79) (Girls)	85.2 (79.2–90.0) (Boys) 90.1 (83.3–94.8) (Girls)	69.0 (66.8–71.1) (Boys) 55.3 (53.0–57.6) (Girls)		22.0 (Boys) 11.0 (Girls)	98.0 (Boys) 99.0 (Girls)	2.8 (Boys) 2.0 (Girls)	0.22 (Boys) 0.18 (Girls)

Abbreviations: 95%CI=95% confidence interval; AUC=area under the curve; %BF=percent body fat; BMC=bone mineral content; BMD=bone mineral density; BMI=body mass index; G=Grade; GINA=Global Initiative for Asthma [52]; HDL-C=high-density lipoprotein cholesterol; HOMA-IR=homeostatic model assessment of insulin resistance; LBP=low back pain; LR+= positive likelihood ratio; LR−= negative likelihood ratio; MetS=metabolic syndrome; MSF=musculoskeletal fitness; NPV=negative predictive value; PPV=positive predictive value; SBP=systolic blood pressure; SD=standard deviation; WBLH=whole-body less head; WC=waist circumference; y=years.

Notes: all values >1.0 were reported to a maximum of 1 decimal place, with all values <1.0 reported to 2 decimal places; [a]case-control study design; [b]asthma=cut-point to identify the presence or absence of asthma (i.e., healthy youth vs. youth with asthma); [c]severity=cut-point to identify the severity of asthma (i.e., intermittent asthma vs. moderately persistent asthma); [d]cross-sectional study design; [e]FitnessGram's trunk extension test is thought to measure both muscular strength and flexibility, but was categorized here as a measure of flexibility as specified by Meredith and Welk [54]; [f]longitudinal cohort study design; [g]only sample data from which the health-related cut-points were developed were reported here; no measures of diagnostic accuracy were available for Peterson et al. [31] because they identified health-related cut-points using conditional inference tree analysis rather than receiver operating characteristic curve analysis; while Castro-Piñero et al. [26] used a longitudinal cohort study design, because the health-related cut-points were developed using the cross-sectional sample, it was classified here as a cross-sectional study design; Saint-Maurice et al. [33] developed health-related cut-points for both single and composite MSF tests, but only measures of diagnostic accuracy related to cut-points for single MSF tests were reported here.

COMPLIANCE WITH ETHICAL STANDARDS

Funding

This systematic review was funded by the Public Health Agency of Canada (#4500414578).

Conflicts of Interest

Brooklyn J. Fraser, Scott Rollo, Margaret Sampson, Costan G. Magnussen, Justin J. Lang, Mark S. Tremblay, and Grant R. Tomkinson declare they have no conflicts of interest.

Author Contributions

BJF screened the records, extracted the data and took responsibility for the integrity of the data, synthesized the results, and co-wrote the manuscript. ASR screened the records, checked the data for accuracy and took responsibility for the integrity of the data, contributed to the interpretation of results, and critically reviewed the manuscript for important intellectual content. MS designed and executed the systematic search strategy, and critically reviewed the manuscript for important intellectual content. CGM contributed to the interpretation of results and critically reviewed the manuscript for important intellectual content. JJL developed the research question, designed the systematic review, contributed to the interpretation of results, and critically reviewed the manuscript for important intellectual content. MST developed the research question, designed the systematic review, contributed to the interpretation of results, and critically reviewed the manuscript for important intellectual content. GRT developed the research question, designed the systematic review, had full access to the data, synthesized the results, and co-wrote the manuscript. All authors have read and approved the final version of the manuscript, agree to be accountable for all aspects of the work, and agree with the order of presentation of the authors.