Title Page

Title:

Clinical outcome measures for monitoring physical function in children and adolescents with obesity: an integrative review

Authors:

Ryan Mahaffey1

Stewart C Morrison2

David Stephensen3

Wendy I Drechsler1

Affiliation:

1 School of Health, Sport and Bioscience, University of East London, London, UK

2 School of Health Sciences, University of Brighton, Brighton, UK

3 East Kent University Hospitals NHS Trust, Canterbury, UK

Keywords:

Children

Obesity

Overweight

Physical function

Measurement properties

Systematic review

Running Title:

Review physical function OW/OB children

Contact Info:

Ryan Mahaffey

School of Health, Sport and Bioscience

University of East London

Water Lane

London

E15 4LZ

r.mahaffey@uel.ac.uk

+44 (0)208 223 4110

+44 (0) 7708258174

Word Count:

4778

Funding:

The lead author (RM) is funded on a three-year post-doctoral research grant from the William M. Scholl Endowment Fund

Disclosure:

The authors declare that there are no conflicts of interest.

- What is already known about this subject? (or for **Review Proposals/Reviews,** what major reviews have already been published on this subject?)

- Physical function is described as the ability to move around and perform daily activities and has been measured by performance in flexibility, strength, aerobic, anaerobic, coordination and balance tasks which is suggested to be reduced in children with obesity.
- Measures of performance-based physical function provide data for comparisons with typical weight children and targets for clinical interventions but there is no gold standard for assessing physical function.
- With many measures available, a review of measurements properties of outcome measures in paediatric rehabilitation is needed to inform the clinical management and assessment of children and adolescents with obesity.

- What does your study add?

- Measurement properties of 66 outcome measures used to evaluate physical function in children who are overweight or obese have been evaluated. Rigorous evaluation of the measurement properties of outcome measures is required in larger collaborative studies, with a particular focus on repeatability and internal consistency.
- There is moderate evidence the six minute timed walk (6MTW) is suitable for measurement of physical function in children who are overweight or obese. But, evidence is low for the use of time and distance measures of aerobic and anaerobic performance, muscle strength and the MABC and BOT2 multi-item performance instruments and, very low for tests of flexibility, co-ordination and balance. Based on this review, measurement of physical function utilising the 6MTW is recommended. All measures except the 6MTW, require

further testing for repeatability and, multi-item instruments require assessment of internal consistency.

- Larger gender and age specific studies using both BMI z-score and % BF methods, together with standardised reference cut-offs to categorise overweight/obese, are needed to determine physical function outcome measures in children with obesity.

Abstract

Objective: Measuring physical function in children with obesity is important to provide targets for clinical intervention to reduce impairments and increase participation in activities. The objective of this integrative review was to evaluate measurement properties of performance-based measures of physical function in overweight and obese children.

Methods: An integrative review of literature published in Cochrane reviews, SPORTDiscus, CINAHL, Plos, Medline and Scopus was conducted.

Results: Twenty-eight studies were eligible and represented 66 performance-based measures of physical function. Assessments of repeatability and feasibility were not conducted in the majority of performance measures reported, only 6 minute-timed-walk (6MTW) was examined for test-retest repeatability. Measures of flexibility, strength, aerobic performance, anaerobic performance, coordination and balance demonstrated construct validity and responsiveness, but findings were inconsistent across all performance-based measures. Multi-item tests of physical function demonstrated acceptable construct validity and responsiveness, but internal consistency was not determined.

Conclusions: There is moderate evidence 6MTW is suitable for measurement of physical function in children with obesity. But, evidence is low for use of aerobic and anaerobic performance, muscle strength, MABC and BOT2 multi-item performance instruments and, very low for flexibility, co-ordination and balance tests. Based on this review, measurement of physical function utilising 6MTW is recommended.

INTRODUCTION

Increasing recognition of the health burden and long-term economic challenges of childhood obesity underpins the importance of optimum and effective healthcare interventions for children with obesity. Evidence shows that childhood obesity is associated with compromised musculoskeletal function and musculoskeletal pathologies [1 , 2, 3, 4] and warrants access to rehabilitation services. Childhood obesity is associated with poorer health related quality of life [5, 6, 7] and linked with a range of functional problems, including an increased risk of pain, discomfort and joint stiffness (particularly in the feet and lower extremities), [8, 9], postural deformities [10] and orthopaedic complications [11, 12]. Findings from a recent review looking at the impact of childhood obesity on physical function and disability, further highlighted problems in obese children with deficits in function that included impaired cardiorespiratory fitness, performance in motor tasks and decrements in muscle strength, gait and balance [13]. Physical activity is key to the prevention and management of obesity [14], maintaining good musculoskeletal health [1, 2] and mitigating orthopaedic co-morbidity and functional impairment [1]. Despite this, the complexities and poor robustness of measuring physical function and the ability to measure effective rehabilitation remains a challenge. There is a lack of consensus on what tools are appropriate to characterise and monitor physical function in obese children [13]. Furthermore, current understanding on the measurement properties of clinical tools used to evaluate features of physical function is lacking.

Physical function is a complex phenomenon and challenging to measure due to the multi-dimensional constructs [15]. Within the International Classification of Functioning, Disability and Health Framework (ICF) [16], the term function encompasses complex interactions between any given health condition, body structures and functions, activity and participation and contextual factors (Figure 1). Physical function fits into the "*activity*" domain of the ICF framework which relates to the execution of a task or action by an individual and has been described as the ability to "*move around*" and perform "*daily activities*" [16]. The choice of outcome measure for evaluating physical function is important to objectively determine baseline function prior to treatment and to evaluate the efficacy of intervention. Physical ability and function vary widely during childhood growth and outcome measures appropriate for both the young and adolescent child are necessary. Outcome measures for assessing physical function can be self-reported or standardised performance measures but, given the challenges with defining function, it is no surprise that multiple outcome measures have emerged in the scientific literature.

The aim of this study was to undertake an integrative review of the literature on the measurement properties (i.e. validity, internal consistency, repeatability, responsiveness and feasibility) of performance-based methods that have been used to assess physical function in children and adolescents with obesity and overweight. The findings from this integrative review will inform discussion around the optimum measures to use in this field through a robust synthesis and critical evaluation of current knowledge. This in turn will provide clinicians and investigators a basis to choose performance-based methods for use in clinical practice or, for future research.

METHODS

Using the search strategy described in online supporting material, a review of the following data sources from January 1990 to February 2014 was undertaken: Cochrane library, EBSCO (SPORTDiscus and CINAHL), Plos, Pubmed (Medline), Elsevier (Scopus). All titles and abstracts generated by the search were independently screened for inclusion by two authors (RM and SM). Disagreement between authors was discussed and consensus reached. The search was restricted to English language and met the following criteria: (1) cross-sectional study design; and (2) described the use of a clinician-assessed method, clinical evaluation or, a measurement tool (multi-item/task scored assessment), to record an aspect of physical function in children and adolescents age ≥5 to ≤19 years with overweight/ obesity (OW/OB). Methods recording (1) self-perceived patient reported physical function; (2) longitudinal or intervention studies; (3) measures requiring expensive sophisticated equipment such as three-dimensional gait analysis, isokinetic dynamomters or accelerometers; (4) no specific aim to study physical function in children with OW/OB; (5) endogenous obesity (e.g. Down's Syndrome) were excluded. In addition, conference proceedings, unpublished reports and case series reports, and studies where OW/OB was not standardised to reference population derived data sets and cut-offs (e.g. BMI Z-Scores) or adiposity estimates (e.g. skin folds, DEXA) were excluded.

Evaluation of measurement properties

The Consensus based standards for the selection of health status measurement instruments (COSMIN) study determined validity, reliability and responsiveness and were key to evaluating the methodological quality of studies on measurement properties. Standardised definitions and criteria to assess construct validity, internal consistency, repeatability, responsiveness together with feasibility were applied using previously established guidelines [17, 18]. Measurement properties were rated as positive (+), negative (-) or indeterminate (?) depending on the methods and results of the study. If no information was available, a zero was recorded.

Construct validity. The extent to which the result of a particular measure related to other measures, disease severity or clinical evidence. Results were determined using correlation or regression analysis (where there was a correlation with other measures of physical performance or disease activity, a *+* was scored; if alternate statistical method undertaken or statistical significance not reported a *?* was scored; where there was no correlation with other measures of physical performance or disease activity a – was scored; and if there was no information about construct validity a *0* was scored).

Internal consistency. The extent to which items in a subscale were interrelated and measured the same dimension; determined using Cronbach's alpha (where Cronbach's alpha ≥ 0.70 a *+* was scored; if alternate statistical method undertaken a *?* was scored; where Cronbach's alpha was < 0.70 a – was scored; and if there was no information about internal consistency a *0* was scored).

Repeatability. The extent to which repeated measurements were reported to be the same. Intraclass correlation coefficient (ICC) or kappa for ordinal/dichotomous data were considered an appropriate measure and, use of Pearson/ Spearman correlation coefficient and coefficient of variation was considered inadequate due to neglect of systematic error (where ICC or k was ≥ 0.70 a *+* was scored; if an alternate statistical method undertaken, a *?* was scored; where ICC or k was < 0.70 a – was scored; and if there was no information about test-retest reliability a *0* was scored).

Responsiveness. The ability of the measure to detect difference in the concept being measured or, to identify impairment (if significant *p*, 0.05 group difference in means ± SD or median ± IQR of group comparison a *+* was scored; if no significant group difference a – was scored; and where there was no statistical analysis or no information on responsiveness a *0* was scored).

Feasibility. The equipment, space, time, cost and training required to administer, record, score and interpret the outcome measure was recorded. The feasibility of the measures was not rated but the appropriate considerations are presented. Rating was not undertaken to determine feasibility because the relevance is dependent on the application and intended use of the outcome measure.

Data extraction

A description of the outcome measure was extracted from the included studies along with a description of the patient characteristics; age, gender, number of children assessed, including number of OW/OB and study design. Data on measurement properties were extracted and recorded in accordance with the criteria previously stated. *Measures* was defined as the physical function concept being measured (e.g. strength, aerobic performance) and *test* refers to the method, protocol or technique utilised (e.g. 6MTW, sit-up) to quantify the physical function measure.

Data synthesis

Identified outcome measures are described together with details of their use. Evidence of construct validity, internal consistency, test–retest repeatability, responsiveness and feasibility is reported and pooled data summarized. Risk of bias in relation to study design and the review process is discussed. Methodological quality of individual studies was assessed using the COSMIN guidelines and summarised using the Cochrane Reviews GRADE criteria. Four levels of quality are specified.

High - randomised studies with low risk of bias OR methodologically well-designed observational studies with large, consistent and precise estimates of the magnitude of an effect and there is a clear consistent dose-outcome response gradient

Moderate - randomised studies with unclear bias OR well-designed observational studies with large, consistent and precise estimates of the magnitude of an intervention effect

Low - randomised studies with high risk of bias OR sound observational studies with consistent estimates of the magnitude and direction of an effect

Very low - randomised or observational studies with serious methodological limitations and high risk of bias

SUMMARY OF EVIDENCE

Search strategy and selection of articles

The initial search strategy resulted in 4511 publications. Following screening of titles and abstracts, 103 studies met the inclusion criteria and were accessed for review of the full-text, of which, 28 eligible studies where included in the review (see Table 1 and Figure 2). Full text studies were excluded for a number of reasons: (1) the study lacked an aim relating to performance based or measurement properties of physical function (34 studies); (2) the study adopted sophisticated equipment which would not typically be available in the clinical environment (17 studies); and (3) the study lacked definition of OW/OB according to referenced population data sets and cut-offs (14 studies).

Participant characteristics

Outcome measures were administered to 36,279 children (18,262 girls) between the ages of 5 and 18 years (Table 1). The number of children with OW/OB in all studies was 9,806 based on BMI z-scores and percentage body fat (% BF). The proportion of children with OW/OB in each study ranged from 12 % to 100 % of the total participants. Only two of the 28 studies evaluated children with OW/OB exclusively [19, 20]. Of the remaining studies, 25.7 % of the participants were OW/OB, and the remainder were HW children.

Outcome measures

Table 2 provides a description of the 66 performance tests used to evaluate physical function in OW/OB children. The measurement domains can be summarised as those evaluating;

- flexibility (8 studies, 26,487 children): sit and reach, v-sit, shoulder stretch [21-28]

- strength (16 studies, 32,739 children): sit-ups, partial sit-ups, bent-leg sit up, horizontal jump, vertical jump, push-ups, bent-arm hang, trunk lifts, sit-to-stand, muscle force (BOT2), throwing [21-36]
- aerobic performance (17 studies, 34,185 children): time and distance limited run/walks [20-28, 30, 37-42]
- anaerobic performance (9 studies, 6,545 children): shuttle and sprint runs, timed up and down stairs, timed up-and-go [19, 22, 26, 27, 29, 31-34]
- coordination and balance (9 studies, 2,914 children): plate taping, agility tasks, one-leg balance, timed-up-and-go, step tests [22, 23, 30, 34, 41, 43-46]
- multi-item instruments recording performance scores (8 studies, 2,334 children) [27, 28, 30, 32, 38, 43, 45, 46]

The multi-item instruments can be divided into two general areas of physical function; firstly, physical fitness (general physical fitness [GPF], The president's physical fitness awards [PPF], Fitnessgram) and secondly, motor skills (Körperkoordinationstest für Kinder [KTK], Movement Assessment Battery for Children [MABC], Bruininks-Oseretsky Test of Motor Proficiency [BOT2] and Community Balance and Mobility Scale [CB&M]). Four studies normalised the performance measure to scores based on variation within the cohort [29, 31, 36, 41] (timed up-and-down stairs only in [41]), and 9 studies normalised the performance measure to referenced standardised scores [27, 28, 30, 34, 38, 42, 43, 45, 46].

Measurement properties

The measurement properties for the 66 performance tests are detailed in Table 1 and summarised in Table 2. When comparing between studies, the term performance was used rather than the specific units used to measure physical performance to avoid confusion between different measurement concepts (e.g. aerobic performance can be greater if more distance is covered in a set time or, less time is taken to cover a set distance).

Construct validity

Construct validity was assessed in all six of the domains of physical function.

Flexibility: Reduced sit-and-reach distance was associated with OW/OB in boys and girls determined by % BF [21, 24] but when determined by BMI z-score, no relationship was found with OW/OB [21]. This highlights the challenge of synthesising data where OW/OB is determined by different methods. No correlation was found when the v-sit version of the sit and reach test was utilised [27]. The sit and reach and v-sit protocol both reflect flexibility of the back, legs and shoulders, however the v-sit version of the test was performed without the need for a box to standardise and stabilise the position of the foot which may have contributed to the lack of association.

Strength: Reduced sit-up performance was associated with OW/OB determined by BMI z-score [21, 30], but not when OW/OB was determined by % BF [21], further highlighting discrepancies when different definitions of OW/OB are utilised. No correlation was found when partial sit-ups [27] and push ups [27, 30] were used as measures of strength. OW/OB was associated with reduced performance of a horizontal jump in boys and girls [24, 31], vertical jump [36] in girls but not boys [32] and sit-to-stand where only girls were assessed [30]. An explanation for the contrast in findings is not clear, however, studies showing no association recruited lower participant numbers (i.e. n<100) compared to the studies which did find associations (n range 116 to 13,500) [21, 24, 31, 32, 34]. Reduced strength, measured by the BOT2 including push-ups, sit-ups, horizontal jumps, was associated with OW/OB [34]. One study found reduced throwing performance was associated with OW/OB in girls measured by % BF [32], but one study found no association in girls and boys measured by BMI z-score [35]. The reason for gender differences is not clear, however, comparing strength values and performance in boys is challenging due to pubertal influences; the gender differences found indicates the importance of controlling for pubertal status when studying boys.

Aerobic performance: Reduced performance of 9 minute run/ walk [21], ¼/ ½/ 1 mile run [27], 800/ 1600 m walk/ run [24] and Leger shuttle run [37] was associated with OW/OB children. Distance walked in 6 minutes correlated with OB/OB in two studies [40, 41], but not a third study which recruited a smaller cohort (n=86 compared to n=347 and 239 in studies 40 and 41, respectively), [30].

Anaerobic performance: Mixed findings on relationships between anaerobic performance and OW/OB were found. Reduced shuttle run performance was associated with OW/OB in two studies [31, 34] but not associated in two other studies [27, 32]. The two studies to find associations both utilised 10 x 5 m shuttles, whereas the two studies that did not find associations utilised only 4 x 5 m shuttles [32] or, no information was provided [27]. It is possible that fewer shuttle runs were insufficient to draw out performance related relationships with OW/OB. Reduced sprint performance was associated with OW/OB measured by % BF [32] and power output during step climbing measured by fat-free mass in OW/OB [19].

Co-ordination and balance: Reduced static balance [34] and time taken to ascend and descend a set of stairs [41] was associated with OW/OB and was not associated in other studies [27, 30]. Differences in timed up-and-down stairs may be related to sample size. Tsiros et al [41] reported associations with BMI SD on 239 children, whereas Nunez-Gaunaurd et al [30] reported for HW (n = 48) and OW/OB (n = 39), separately. Reduced performance of dynamic coordination tasks in the MABC [45] (one-leg balance/ jumping/ walking) and the BOT2 [34] (tapping foot and finger, jumping jacks, walking forward on a line, standing on one leg on a balance beam, throwing a ball at a target, catching a tossed ball) along with ball skills (catching/ rolling/ throwing) were associated with OW/OB. But, manual dexterity (posting/ threading/ writing) from the MABC was not associated with OW/OB [45].

Multi-item performance instruments: All instruments were associated with poorer performance in OW/OB children compared to HW [27, 28, 30, 32, 38, 43, 45, 46], although Čapkauskienė and colleagues [32] found associations in girls but not in boys.

Internal consistency

No study investigated internal consistency of the seven multi-item outcome measures (GPF, KTK, MABC, PPF, Fitnessgram, BOT2 and CB&M) in OW/OB children.

Test-retest repeatability

Of the 66 measures, repeatability in OW/OB children was only evaluated for the 6 minute timed walk test (6MTW) [20, 40]. Indeterminate ratings were given for many studies because ICC or Kappa statistics were not reported. Use of the coefficient of variation (CoV) is considered an inadequate measure of test-retest repeatability due to its dependency on the scale of measurement [47].

Responsiveness

Responsiveness was assessed in all six of the domains of physical function.

Flexibility: Findings were inconsistent for flexibility performance with two studies reporting poorer performance in OW/OB compared to HW [26, 28] and five studies finding no differences between OW/OB and HW [22-26]. Joshi et al [28] was the only study to measure flexibility of the upper body (shoulder stretches); the difference in this flexibility measure may have led to the contrasting finding between other flexibility studies which used sit-and-reach technique. The one study to find significant differences in sit-and-reach performance between HW and OW/OB stratified the sample by gender and age group (6 to 9 years and 10 to 13 years) [26]. Differences between groups in other studies may have been hidden in variation between age groups and genders.

Strength: Most studies (n=21) reported strength performance was significantly lower in OW/OB [22-25, 28-30, 33-35] however, two studies found significantly higher performance in OW/OB [29, 35] and seven studies reported no differences between OW/OB and HW groups [23, 28, 30, 36, 42]. Responsiveness of strength tests for sit-ups, horizontal jump, vertical jump, push-ups, bent-arm hangs and sit-to-stand demonstrated lower performances in children with OW/OB compared to HW [22-25, 28, 30, 33-35]. Only strength measures where body mass was not lifted against gravity (i.e. throwing), did OW/OB perform better than HW [29, 35].

Aerobic performance: Most studies (n=14) reported aerobic performance was significantly lower in OW/OB [22, 24-26, 28-30, 38, 40-42], but three studies found no differences between OW/OB and HW groups [26, 39, 42]. Three studies found significant differences in 6MTW performance between OW, OB and HW groups [30, 40, 41]. In contrast, Guinhoya et al [39] found no significant difference in 6MTW when comparing a group of OW girls to a healthy group of girls however, the lack of an obese group may have influenced these findings.

Anaerobic performance: Performance of anaerobic tests was significantly lower in OW/OB in five tests [22, 29, 33, 34] and no differences were found in one test [26]. The possible reason for finding no significant difference in anaerobic performance in the study by Malina et al [26] was the low sample size of OW/OB participants; of girls and boys age 6-7 to 8-9 years, 15 OW/OB participants were reported in each gender and, of girls and boys age 9-10 to 11-12 years, 27 and 22 OW/OB participants, respectively, were reported.

Coordination and balance: There was significantly poorer performance in coordination and balance tests in children with OW/OB for 15 tests [34, 41, 43-46] and no differences were found in 15 tests [22, 23, 30, 45, 46]. The range and variation of tests for coordination and balance may underlie the reason for contrasting findings between studies. The tests ranged from predominantly coordination based measures such as plate tapping and posting/ treading/ writing to whole body dynamic balance measures such as jumping side-to-side and hopping. Whilst coordination and balance are commonly examined together during the same test, there is a need to appropriately define the measure for between study comparisons.

Multi-item performance instruments: All instruments that measured responsiveness (four tests) found significantly poorer performance of OW/OB compared to HW children [30, 38, 43, 45].

Feasibility

Table 3 presents the data provided on the conditions under which measurements were recorded, the time taken to carry out the procedures, the expertise of the person(s) administering the tests and the equipment requirements. In most of the studies, measures of time (stopwatch), distance (tape measure, sit-and-reach box) and repetitions were included, but information on the specific measurements was often lacking in many of the studies reviewed.

Study design

Insufficient reporting of study population in terms of BMI z-scores reference data and % BF measures, together with differences in the reference values and cut-offs utilised for OW/OB categories limit interpretability of this review. Of the five studies to define OW/OB by % BF measures, 71 % of the ratings were positive compared to only 27 % rating when OW/OB was defined by BMI z-score. Sophisticated body composition measures such as DEXA [41] may provide better sensitivity for defining OW, OB and HW groups compared to skinfold, bioelectrical impedance analysis (BIA) and BMI z-score, but places participants at risk of radiation exposure thereby reducing its clinical application. Eleven studies (44 %) utilised the International Obesity task force(IOTF) reference data set to calculate BMI z-scores, seven studies (28 %) used the Centre for Disease

Control and Prevention (CDC) reference data set, six (24 %) referred BMI to national data sets and one study did not report the reference data set. Cut-offs were also defined differently between studies, fifteen studies (60 %) used the IOTF cut-offs for HW, OW and OB (based on adult cut-offs of 25 kg/m^2 and 30 kg/m^2 for women and men at 18 years, respectively); five (20 %) used CDC cut-offs (≥ 95th percentile = Obese, 85th to < 95th percentile = Overweight); two (8 %) referenced national cut-offs; and three (12 %) did not reference their cut-offs. Many regression type analysis of physical function were excluded from this review because BMI was not presented as a BMI z-score despite the studies using BMI z-score to define the number of participants with OW/OB.

All studies included boys and girls in their samples however, only nine studies (14 %) separated results by gender. Many studies included a comparison of physical function between boys and girls but did not separate the cohort into gender categories for comparison between groups of HW and OW/OB. Performance differences in all physical function domains (except coordination and balance) were found to differ in the studies that stratified their sample into boys and girls, suggesting gender differences in physical function exist in OW/OB children.

The majority of the studies analysed performance using the raw outcome data, ten studies converted the raw data to scores/ centiles based on standardised reference values, four studies referenced the raw data to centiles within the cohort of children in the study sample. Of the four studies to report outcome data to within cohort variation, three [29, 31, 41] found significant construct validity and/ or responsiveness results for all measures of physical function. Hamlin et al [36] found no construct validity for boys and no responsiveness for boys or girls, but the sample size was low (n = 54), meaning cohort variation may have been small. Of the nine studies to report outcome data relative to standardised reference values, seven studies [27, 28, 30, 38, 43, 45, 46] produced multi-item scores all of which demonstrated significant construct validity and/ or responsiveness. The two studies that did not produce multi-item scores [34, 42] were based on reported multi-item tests (BOT-2 and FITNESSGRAM) and only the individual tests were reported.

The benefit of referring outcome measures to standardised values, is that cut-offs for the level of performance (e.g. poor, normal, high) can be used to define the amount of physical function. However, standardisation relies on the reference sample being of large enough numbers and spread of values that cut-offs (which maybe arbitrarily chosen) can be derived. Consideration of the type of analysis should be undertaken based on the design and aim of the study.

Limitations of the summary of evidence

It is acknowledged that there are no standardised criteria to assess the measurement properties of performance-based physical function and alternative criteria applied to the data may produce different interpretations. Non-English publications, conference proceedings, unpublished reports and case series reports were not included. The search was limited to cross-sectional studies in order to establish the baseline measurement properties. It is acknowledged that the omission of longitudinal and intervention studies may lead to bias and an underestimation of the responsiveness of measures. However, understanding the effects of interventions on measurement properties would require further review and analysis. A further review of longitudinal and intervention studies should be conducted to fully analyse outcome measures of physical function in children with OW/OB. Only Malina et al [26] stratified their sample population by age and found differences between age groups 6-7 to 8-9 years and 9-10 to 11-12 years in boys and girls (Mexican school grades 1-3 and 4-6), respectively. Differences in age groups may underlie some of the contrasting findings in this review. Fourteen studies (50 %) sampled only children age 5-12 years, twelve studies (43 %) sampled children and adolescents (age range crossed 12-13 years) and two studies (7 %) sampled only adolescents age 13-18 years. Future studies should consider recruiting or stratifying samples by child/ adolescent age ranges. Of particular interest to this review was the exclusion of sophisticated equipment such as three dimensional (3D) motion capture. Whilst the inclusion of measurement properties of sophisticated equipment would be useful, it was beyond the scope of this review which focussed on tools commonly used in clinical practice. Most studies in this review were observational cohort studies and did not set out any hypotheses to evaluate measurement properties which may explain why many received indeterminate or negative ratings. Hypothesis testing is needed to establish which outcomes should be used for answering specific clinical and research questions.

KEY SUMMARY OF EVIDENCE

- Of the 66 measures identified, only the 6-minute timed walk test was examined for all measurement properties (internal consistency not appropriate).
- There is moderate evidence that the 6-minute timed walk test is suitable for monitoring physical function.
- There is low evidence that time and distance measures of aerobic performance, anaerobic performance measured with: 10 x 5 m sprints (but not when few repetitions are utilised) and, multi-item performance instruments (in particular the MABC and BOT2) may be suitable for monitoring physical function in OW/OB children (but test-retest repeatability is lacking). Internal consistency of the multi-item tests has been determined in other paediatric populations [48, 49], but, it has not been evaluated specifically in OW/OB children.
- There is low evidence in OW/OB girls and insufficient evidence in boys that, strength measured with a full sit-up and horizontal jump is suitable for monitoring physical function in OW/OB children, (but test-retest repeatability is lacking).
- There is very low evidence that measurement of flexibility, co-ordination and balance tests are suitable methods of monitoring physical function in OW/OB children.
- More positive ratings were given to studies that measured OW/OB by % BF compared to studies that used BMI z-score, although to date, no established cut-offs for defining OW/OB by % BF have been determined.

- Few of the studies included in this review set out with the aim to investigate the measurement properties of outcome measures for physical function. Future work is needed, in particular, to understand the repeatability of most of the tests included in this review (with the exception of 6MWT). Similarly, no study on multi-item test scores for physical function examined internal consistency specifically in paediatric OW/OB. Rigorous evaluation of the measurement properties of outcome measures used to monitor physical function in children with OW/OB in larger collaborative studies is required.

CONCLUSION AND RECOMMENDATIONS

The synthesis of current evidence on the measurement properties of physical function will assist measurement selection by clinicians and researchers with an interest in evaluating physical function in children with OW/OB. Few of the studies included in this review set out with the aim to investigate the measurement properties of outcome measure for physical function. Rigorous evaluation of the measurement properties of outcome measures used to monitor physical function in OW/OB children in larger collaborative studies is required with a particular focus on those not assessed in the current review: Measurement error, validity (content, structural, cross-cultural and criterion), and interpretability. There is moderate evidence the 6MTW is suitable for measurement of physical function in children with obesity. But, evidence is low for the use of time and distance measures of aerobic and anaerobic performance, muscle strength and the MABC and BOT2 multi-item performance instruments and very low for tests of flexibility, co-ordination and balance. Based on this review, measurement of physical function utilising the 6MTW is recommended.

All tests, except the 6MTWrequire further testing for repeatability and multi-item instruments require assessment of internal consistency in OW/OB children. Larger gender and age specific studies using both BMI z-score and % BF methods together with standardised reference cut-offs to categorise OW/OB are needed to determine physical function outcome measures in OW/OB children. The selection of age and gender-matched reference data and cut-offs for OW/OB requires careful consideration as this appears to influence comparisons of physical function tests.

References

1. Shultz SP, Anner J, Hills AP. Paediatric obesity, physical activity and the musculoskeletal system. *Obes Rev.* 2009*;* 10(5):576-82.

2. Ebbeling CB, Pawla DB, Ludwig DS. Childhood obesity: Public-health crisis, common sense cure. *Lancet.* 2002; 360(9331):473-82.

3. Paulis, W. D., Silva, S., Koes, B. W., & Middelkoop, M. (2014). Overweight and obesity are associated with musculoskeletal complaints as early as childhood: a systematic review. *Obesity Rev.* 2014; *15*(1): 52-67.

4. Krul, M., van der Wouden, J. C., Schellevis, F. G., van Suijlekom-Smit, L. W., & Koes, B. W. Musculoskeletal problems in overweight and obese children. *Ann Fam Med*. 2009; *7*(4): 352-356.

5. Pinhas-Hamiel, O., Singer, S., Pilpel, N., Fradkin, A., Modan, D., & Reichman, B. Health-relted quality of life among children and adolescents: associations with obesity. *Int J Pediatric Obes.* 2006; *30*(2): 267-272.

6. Schwimmer, J. B., Burwinkle, T. M., & Varni, J. W. Health-related quality of life of severely obese children and adolescents. *JAMA.* 2003*;* 289 (14), 1813-1819.

7. Bout-Tabaku, S., Briggs, M. S., & Schmitt, L. C. Lower extremity pain is associated with reduced function and psychosocial health in obese children. *Clin Orthop Rel Res.* 2013; *471*(4): 1236-1244.

8. De Sá Pinto, A. L., De Barros Holanda, P. M., Radu, A. S., Villares, S. M., & Lima, F. R. Musculoskeletal findings in obese children. *Paediatr Child Health*. 2006; *42*(6): 341-344.

9. Stovitz, S. D., Pardee, P. E., Vazquez, G., Duval, S., & Schwimmer, J. B. Musculoskeletal pain in obese children and adolescents. *Acta Paediatr.* 2008, *97*(4): 489-493.

10. Wearing, S. C., Hennig, E. M., Byrne, N. M., Steele, J. R., & Hills, A. P. The impact of childhood obesity on musculoskeletal form. *Obesity Rev.* 2006; *7*(2): 209-218.

11. Wills, M. Orthopedic complications of childhood obesity. *Ped Phys Ther.* 2004*; 16*(4): 230-235.

12. Taylor, E. D., Theim, K. R., Mirch, M. C., Ghorbani, S., Tanofsky-Kraff, M., Adler-Wailes, D. C., Brady, S., Reynolds, J.C., Calis, K.A., & Yanovski, J. A. Orthopedic complications of overweight in children and adolescents. *Pediatr*, 2006; *117*(6): 2167-2174.

13. Tsiros MD, Coates AM, Howe PR, Grimshaw PN, Buckley JD. Obesity: the new childhood

disability? *Obes Rev*. 2011;12(1):26-36.
14. Trost, S. G., Kerr, L. M., Ward, D. S., & Pate, R. R. Physical activity and determinants of physical activity in obese and non-obese children. *Int J Obes Relat Metab Disord*. 2001; *25*(6): 822-829.

15. Dobson F, Hinman RS, Roos EM, Abbott JH, Stratford P, Davis AM, et al. OARSI recommended performance-based tests to assess physical function in people diagnosed with hip or knee osteoarthritis. *Osteoarthritis Cartilage*. 2013;21(8):1042-52.
16. World Health Organisation. Towards A Common Language for Functioning, Disability and Health: International Classification for Functioning, Disability and Health. World Health Organisation: Geneva, 2002.

17. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. BMC Med Res Methodol. 2010; 10(22): 160-5

18. Terwee CB, Mokkink LB, Steultjens MP, Dekker J. Performance-based methods for measuring the physical function of patients with osteoarthritis of the hip or knee: a systematic review of measurement properties. *Rheumatology*. 2006;45(7):890-902.

19. Sartorio A, Agosti F, De Col A, Lafortuna CL. Age- and gender-related variations of leg power output and body composition in severely obese children and adolescents. *J Endocrinol Invest*. 2006;29(1):48-54.

20. Vanhelst J, Fardy PS, Salleron J, Beghin L. The six-minute walk test in obese youth: reproducibility, validity, and prediction equation to assess aerobic power. *Disabil Rehabil*. 2013;35(6):479-82.

21. Andreasi V, Michelin E, Rinaldi AE, Burini RC. Physical fitness and associations with anthropometric measurements in 7 to 15-year-old school children. *J Pediatr*. 2010;86(6):497-502.

22. Casajus JA, Leiva MT, Villarroya A, Legaz A, Moreno LA. Physical performance and school physical education in overweight Spanish children. *Ann Nutr Metab*. 2007;51(3):288-96.

23. Graf C, Koch B, Dordel S, Schindler-Marlow S, Icks A, Schuller A, et al. Physical activity, leisure habits and obesity in first-grade children. *Eur J Cardiovasc Prev Rehabil*. 2004a;11(4):284-90.

24. Liao Y, Chang SH, Miyashita M, Stensel D, Chen JF, Wen LT, et al. Associations between health-related physical fitness and obesity in Taiwanese youth. *J Sports Sci*. 2013;31(16):1797-804.

25. Mak KK, Ho SY, Lo WS, Thomas GN, McManus AM, Day JR, et al. Health-related physical fitness and weight status in Hong Kong adolescents. *BMC public health*. 2010;10:88.

26. Malina RM, Pena Reyes ME, Tan SK, Little BB. Physical fitness of normal, stunted and overweight children 6-13 years in Oaxaca, Mexico. *Eur J Clin Nutr*. 2011;65(7):826-34.

27. Gray A, Smith C. Fitness, dietary intake, and body mass index in urban Native American youth. *J Am Diet Assoc*. 2003;103(9):1187-91.

28. Joshi P, Bryan C, Howat H. Relationship of body mass index and fitness levels among schoolchildren. *J Strength Cond Res*. 2012;26(4):1006-14.

29. Bovet P, Auguste R, Burdette H. Strong inverse association between physical fitness and overweight in adolescents: a large school-based survey. *Int J Behav Nutr Phys Act*. 2007;4:24.

30. Nunez-Gaunaurd A, Moore JG, Roach KE, Miller TL, Kirk-Sanchez NJ. Motor proficiency, strength, endurance, and physical activity among middle school children who are healthy, overweight, and obese. *Pediatr Phys Ther*. 2013;25(2):130-8.

31. Ara I, Sanchez-Villegas A, Vicente-Rodriguez G, Moreno LA, Leiva MT, Martinez-Gonzalez MA, et al. Physical fitness and obesity are associated in a dose-dependent manner in children. *Ann Nutr Metab*. 2010;57(3-4):251-9.

32. Čapkauskienė S, Visagurskienė K, Bakienė R, Vitkienė I, Vizbaraitė D. Peculiarities of physical fitness and body composition of 5-7 year-old children of several kaunas preschools and interaction between those indexes. Educ. *Physical Train. Sport*. 2009(73):14-20.

33. Gonzalez-Suarez CB, Grimmer-Somers K. The association of physical activity and physical fitness with pre-adolescent obesity: an observational study in Metromanila, Philippines. *J Phys Act Health*. 2011;8(6):804-10.

34. Poulsen AA, Desha L, Ziviani J, Griffiths L, Heaslop A, Khan A, et al. Fundamental movement skills and self-concept of children who are overweight. *Int J Pediatr Obes*. 2011;6(2-2):e464-71

35. Riddiford-Harland DL, Steele JR, Baur LA. Upper and lower limb functionality: are these compromised in obese children? *Int J Pediatr Obes*. 2006;1(1):42-9

36. Hamlin MJ, Fraser M, Lizamore CA, Draper N, Blackwell G, Shearman J. Effects of Bioelectrical Impedance-Derived Fat and Lean Mass on Fitness Levels in 8- to 13-Year-Old Children. *Asian Journal of Exercise & Sports Science*. 2014;11(1):36-45.

37. Burke V, Beilin LJ, Durkin K, Stritzke WG, Houghton S, Cameron CA. Television, computer use, physical activity, diet and fatness in Australian adolescents. *Int J Pediatr Obes*. 2006;1(4):248-55.

38. Graf C, Koch B, Kretschmann-Kandel E, Falkowski G, Christ H, Coburger S, et al. Correlation between BMI, leisure habits and motor abilities in childhood (CHILT-project). *Int J Obes Relat Metab Disord*. 2004b;28(1):22-6.

39. Guinhouya B. Outcomes and cardiac response of overweight prepubescent to the 6 minutes walk test. *Minerva Pediatr*. 2011;63(5):375-84.

40. Morinder G, Mattsson E, Sollander C, Marcus C, Larsson UE. Six-minute walk test in obese children and adolescents: reproducibility and validity. *Physiother Res Int*. 2009;14(2):91-104

41. Tsiros MD, Buckley JD, Howe PR, Olds T, Walkley J, Taylor L, et al. Day-to-day physical functioning and disability in obese 10- to 13-year-olds. *Pediatr Obes*. 2013;8(1):31-41

42. Mota J, Flores L, Flores L, Ribeiro JC, Santos MP. Relationship of single measures of cardiorespiratory fitness and obesity in young schoolchildren. *Am J Hum Biol*. 2006;18(3):335-41.

43. D'Hondt E, Deforche B, Vaeyens R, Vandorpe B, Vandendriessche J, Pion J, et al. Gross motor coordination in relation to weight status and age in 5- to 12-year-old boys and girls: a cross-sectional study. *Int J Pediatr Obes*. 2011;6(2-2):e556-64

44. Pathare N, Haskvitz EM, Selleck M. Comparison of measures of physical performance among young children who are healthy weight, overweight, or obese. *Pediatr Phys Ther*. 2013;25(3):291-6.

45. D'Hondt E, Deforche B, De Bourdeaudhuij I, Lenoir M. Relationship between motor skill and body mass index in 5- to 10-year-old children. *Adapt Phys Activ Q*. 2009;26(1):21-37

46. Wright MJ, Bos C. Performance of children on the Community Balance and Mobility Scale. *Phys Occup Ther Pediatr*. 2012;32(4):416-29.

47. Stephensen D, Drechsler WI, Scott OM. Outcome measures monitoring physical function in children with haemophilia: a systematic review. *Haemophilia*. 2014;20(3):306-21

48. Wuang YP, Su CY. Reliability and responsiveness of the Bruininks-Oseretsky Test of Motor Proficiency-Second Edition in children with intellectual disability. Research in developmental disabilities. 2009;30(5):847-55

49. Wuang YP, Su JH, Su CY. Reliability and responsiveness of the Movement Assessment Battery for Children-Second Edition Test in children with developmental coordination disorder. Developmental medicine and child neurology. 2012;54(2):160-5.

Figures

Figure 1. International Classification of Functioning, Disability and Health Framework [10]

Figure 2. Study selection diagram

Tables

Online supporting material: Search criteria

Table 1. Measurement properties and outcome measures

Table 2. Summary of measurement properties of outcome measures

Table 3. Characteristics of outcome measures