Bryant, M., Ashton, L., Nixon, J., Jebb, S., Wright, J., Roberts, K. ... Summerbell, Carolyn. Framework of outcome measures recommended for use in the evaluation of childhood obesity treatment interventions: the CoOR framework. *Pediatric Obesity*, 9, e116-e131.

Framework of outcome measures recommended for use in the evaluation of childhood obesity treatment interventions: The CoOR framework

Running title: Childhood Obesity Outcome Measures Framework

Maria Bryant[1 *], Lee Ashton[1,2], Jane Nixon[1], Susan Jebb[3], Judy Wright[4], Katharine Roberts[5] and Julia Brown [1], in collaboration with the CoOR Scientific advisory group[6]

[1] Clinical Trials Research Unit, University of Leeds, Leeds, LS2 9JT, UK

[2] The University of Newcastle, Callaghan NSW 2308, Australia

[3] MRC Human Nutrition Research, Cambridge, UK

[4] Institute of Health Sciences, University of Leeds, Leeds, UK

[5] National Obesity Observatory (NOO), Oxford, UK

[6] Ashley Cooper[i], Steven Cummings[ii], Ulf Ekelund[iii], Claudia Gorecki[iv], Lucy Griffiths[v], Julian Hamilton-Shield[vi], Andrew Hill[vii], Paul Kind[viii], John Reilly[ix] and Carolyn Summerbell[x].

[i] Centre for Exercise, Nutrition and Health Sciences, School for Policy Studies, University of Bristol, Bristol

[ii] Department of Social & Environmental Health Research, London School of Hygiene & Tropical Medicine, London, UK

[iii] Department of Sport Medicine, Norwegian School of Sport Sciences, Oslo, Norway and MRC Epidemiology Unit, University of Cambridge, Cambridge, UK

[iv] Clinical Trials Research Unit, University of Leeds, Leeds, LS2 9JT, UK

[v] MRC Centre of Epidemiology for Child Health , UCL Institute of Child Health, 30 Guilford Street, London, UK

[vi] NIHR Bristol Biomedical Research Unit in Nutrition, Bristol, UK; and University of Bristol, Bristol, UK

[vii] Institute of Health Sciences, University of Leeds, Leeds, UK

[viii] Institute of Health Sciences, University of Leeds, Leeds, UK

[ix] Physical Activity for Health Group, School of Psychological Sciences and Health University of Strathclyde, Glasgow, UK

[x] Department of Medicine, Pharmacy and Health, University of Durham, Durham, UK

*Corresponding author:

Maria Bryant

Clinical Trials

Clinical Trials Research Unit

Leeds Institute of Clinical Trials Research

University of Leeds

Leeds

LS2 9JT

Tel: 0113 343 7632

Fax: 0113 343 1471

Email: m.j.bryant@leeds.ac.uk

Website: http://ctru.leeds.ac.uk

Keywords: outcome measures, framework, systematic review, recommendation, quality

What is already known about this subject

- Systematic reviews on the effectiveness of treatment programmes for childhood obesity consistently report inadequacies in the conduct and reporting of trials
- There is a lack of consensus on which outcome measures to use in trials of childhood obesity treatments
- Lack of consensus and inadequate reporting limits our understanding of which treatments are effective

What this study adds

- This study identified existing outcome measures that have been used in the evaluation of childhood obesity treatments and those that have been developed for such use
- Identified measures were appraised for quality in order to develop a framework of recommended outcome measures
- This framework (The CoOR framework) will enable consistency in outcome measures used in future evaluations of childhood obesity treatments, to researchers and those working in policy and practice

Abstract

Introduction
Consensus is lacking in determining appropriate outcome measures for assessment of childhood obesity treatments. Inconsistency in the use and reporting of such measures impedes comparisons between treatments and limits consideration of effectiveness. This study aimed to produce a framework of recommended outcome measures; The Childhood obesity treatment evaluation Outcomes Review (CoOR) framework.

Methods
A systematic review including 2 searches was conducted to identify (1) existing trial outcome measures; (2) manuscripts describing development/evaluation of outcome measures. Outcomes included: Anthropometry, diet, eating behaviours, physical activity, sedentary time/behaviour, fitness, physiology, environment, psychological well-being and Health Related Quality of Life. Eligible measures were appraised by the internal team using a system developed from international guidelines; followed by appraisal from national external expert collaborators.

Results
25,486 papers were identified through both searches. Eligible Search 1 trial papers cited 417 additional papers linked to outcome measures, of which 56 were eligible. A further 297 outcome development/evaluation papers met eligibility criteria from Search 2. Combined, these described 191 outcome measures. After internal and external appraisal, 52 measures across 10 outcomes were recommended for inclusion in the CoOR framework.

Conclusion
Application of the CoOR framework will ensure greater consistency in choosing robust outcome measures that are appropriate to population characteristics.

Introduction
The degree to which weight management leads to reduction of obesity is reflected by measuring change in outcomes following participation in an intervention compared to controls. Outcomes either directly measure a definitive clinical change, (e.g. primary outcome of obesity such as weight loss); or assess proximal/secondary outcomes (e.g. change in diet) that impact on the primary outcome. In trial design, choosing appropriate outcomes is essential. Inappropriate outcomes will result in data that are inaccurate or biased and that do not indicate the effectiveness. Moreover, collection of data using poorly chosen outcomes is a waste of resources for the researchers and participants [(1)]. Inappropriate selection of outcomes in childhood obesity research is likely due to the uncertainty about which are most relevant to children and their families [(2)]. Furthermore, there is a lack of knowledge of which can be most reliably measured.

The lack of consensus in determining appropriate outcome measures for the reliable and valid assessment of childhood obesity interventions restricts comparisons between interventions, partly because of a shortage of validated outcome measures available, but also because selected outcome measures differ between studies. Such a lack of consistency impedes the progress of childhood obesity research. The aim of this study was to perform a systematic review to identify existing outcome measures used in childhood obesity treatment evaluations and appraise their quality in order to create a framework; 'The Childhood obesity Outcomes Review (CoOR) framework' containing recommended primary and secondary outcome measures.

Methods
A mixed methods approach was used to develop the CoOR framework, including a systematic review, followed by a quality appraisal study to identify robust measures.
Systematic Review:
Two literature searches were performed. Search 1 identified randomised controlled trials, pilot and feasibility studies of childhood obesity treatment evaluation studies with the intent of identifying outcome measures (and corresponding citations) already used in trials. Included outcomes and outcome measures are shown in Table 1. Search 2 aimed to directly identify manuscripts describing the development and/or evaluation of outcome measures intended for use in childhood obesity intervention evaluations. Both searches were conducted from August 2011 to October 2011 in databases: MEDLINE, MEDLINE in process, EMBASE, PsycINFO, HMIC, AMED, Global Health, Maternity and Infant Care (all Ovid); Cinahl (EbscoHost); Science Citation Index (WoS); and the Cochrane Library (Wiley) from the date of inception, with no language restrictions. Unpublished literature were sought by searching Inside Conferences, Systems for Information in Grey Literature (SIGLE), Web of Science Conference Proceedings Citation Index- Science (Thomson) and ClinicalTrials.gov. The same eligibility criteria were applied for each of these additional sources. Two reviewers (MB and LA) conducted both searches (with agreement of 98% and 96% for Searches 1 and 2 respectively). Disagreements were resolved by discussion. Full search strategies for all searches are available upon request from the author.

Eligible Search 1 studies included children (≤18 years) and should describe the evaluation of any intervention to treat obesity, including lifestyle, drug and surgery interventions. Studies without a primary outcome of obesity reduction (e.g. weight loss, BMI or adiposity reduction) were not included. For Search 2, methodological studies describing the development and evaluation of outcome methods were eligible, including quantitative measurement, qualitative assessment, feasibility and psychometric studies. As with Search 1, studies had to include evaluation in children. Studies comparing different cut off points, population equations or standards of population based criteria were not included.

In addition to study characteristics, extraction forms gathered data related to outcome measurement development (e.g. conceptual framework, involvement of users) and outcome measure evaluation. Specific sections within reliability included: Internal reliability (e.g. internal consistency); test re-test; and inter-rater reliability. Validity sections included: Internal validity (e.g. factor analysis); criterion validity (with pre-specified 'permitted' gold standard/criterion measures); convergent validity (described here as the association with another measure, aimed at assessing the same or similar construct(s)); and construct validity (i.e. ability of a tool to measure the concept being studied). Data describing face and content validity were extracted as part of the outcome measurement development. Sample size was recorded for each type of evaluation. Data were 'double-extracted' by two authors (MB and LA) to reach 100% agreement.

Quality appraisal of outcome measures

Measures were appraised for quality in order to identify those which demonstrated rigorous methods in both development and evaluation procedures. Appraisal involved two stages: (1) Internal appraisal; and (2) external appraisal.

Internal appraisal: Principles of international guidelines [3, 4] were drawn upon to appraise rigour (i.e. development and measurement properties) of outcome measures meeting eligibility criteria. Measures within outcome domains were specifically appraised according to its construct and/or clinical context since strict adherence to any individual guideline was not always appropriate. For example, physiological measures such as blood lipids were not expected to have involved patients at the development stage, nor present data on item reduction. Specific characteristics that were included were: concepts being measured; number of items; conceptual framework; intended use; population for intended use; data collection method; administration mode; response options; recall period; scoring; weighting; format; and response burden. A scoring system was also applied to secondary outcome measures, based on quality in the conduct and results of evaluation where appropriate and ranged from 1-4 (with 1 being the lowest). These were developed from criteria set by the international guidelines [3, 4], in addition to previous research conducted by the lead author [5]. Two members of the CoOR internal team (MB and LA) then used this information to classify each of the primary and secondary measures into one of 3 categories (by discussion and consensus) in relation to their confidence of whether or not each measure should be recommended for inclusion into the CoOR framework: (1) 'certain: good evidence, fit for purpose' (2) 'certain: poor evidence, not fit for purpose' and (3) 'uncertain, requiring further consideration'. Tools were only placed into Category 1 or 2 providing mutual agreement had been established. Category 1 was only assigned when the tool was clearly highly robust in terms of development and evaluation. Similarly, Category 2 was only assigned when the tools was very poorly developed and evaluated. Any disagreements were placed into Category three to be further discussed at the expert appraisal meeting.
External appraisal: Results of the systematic review and internal appraisal were reviewed by the CoOR Scientific Advisory group; consisting of obesity experts with specific proficiency in each outcome, in addition to methodological experts. Prior to a face-to-face meeting, experts were asked to consider factors such as: appropriateness domain of categorisation; obvious omissions (including knowledge of modified versions of outcomes); and personal and theoretical experience of use of outcome measures. During the meeting, discussions began by reviewing the internal appraisal decisions 1 (certain, fit for purpose) and 2 (certain, unfit for purpose). Disagreements resulted in measures being re-categorised as 3 (uncertain, requiring further consideration). All outcome measures that had been given an appraisal decision of 3, were then more fully discussed. Justifications for decisions were provided at the meeting and final rulings were based on consensus. All final decisions contributed towards the development of a provisional framework, which was later forwarded to each expert to secure their final agreement.

Results and Discussion
Combined, Searches 1 and 2 identified 25,486 manuscripts. A further 25 were identified through hand searching (grey literature and review citations). Of these, 14,419 were Search 1 trial manuscripts and 11,092 were Search 2 methodology manuscripts. Screening for eligibility at both the title and abstract stage and the full paper review resulted in the inclusion of 200 trial manuscripts from Search 1. After data were extracted from these papers, 417 further linked citations for outcome measures used by the trials were identified. However, only 56 (13%) correctly cited manuscripts which met eligibility criteria for inclusion as methodology papers. Screening of Search 2 methodology papers resulted in the inclusion of 320 manuscripts meeting eligibility criteria. Combined with Search 1, a total of 376 manuscripts were identified that described 180 outcome measures. Discrepancies between the number of manuscripts and the number of studies was a result of manuscripts evaluating more than one measure, and measures in which there were multiple manuscripts describing their evaluation.

Of the 180 measures that were appraised, 52 outcome measures across 10 outcome domains were recommended for inclusion to the CoOR outcome measures framework (Table 2).

Recommended primary outcome measures were BMI and DXA. Inclusion of BMI was, in part, based on the feasibility of its use and the ability to ensure comparability between evaluations. Fifty seven per cent of the eligible trials identified by the review reported using BMI (or a derivative of BMI) as a primary outcome. While the evidence of validity offered by the methodology studies within the review was inconsistent for BMI, experts agreed that it can be reliably measured, provided that administrators are well-trained and equipment is regularly calibrated. However, limitations of BMI were also acknowledged. Primarily, BMI does not provide any information about body composition

(including adiposity) or fat distribution. This caveat needs to be considered particularly in studies that evaluate interventions focused on physical activity (especially those focussed on strength training). However, most childhood obesity programmes are multi-faceted, composing a variety of lifestyle interventions. If feasible, the CoOR framework also advocates the use of DXA, which is also a proxy measure of adiposity, but is able to provide estimates that differentiates between fat and lean tissue. The equipment needed to conduct DXA measurements is expensive and though widely available in hospital settings, may not always be available for research purposes, especially in community settings; thus, the CoOR framework suggests that DXA is supported with measurement of BMI to allow comparisons between intervention evaluations.

Recommended secondary outcomes have been provided for all included outcome domains. However, researchers are advised to only include measures that will assess what they expect to change following an intervention, or what they believe will mediate such changes. Thus, it is not necessary to include a measure from all outcome domains in every programme evaluation. Similarly, where multiple measures are advocated within an outcome domain, researchers are advised to consider which measures are most closely aligned to the intervention targets and, where available, choose a measure that has been developed in a population most similar to the intended sample.

Experts agreed that objective measurements must be used in all outcome domains where available (i.e. activity monitors instead of self-reported physical activity) and where objective measures are available, no self-reported measures were recommended. Although findings from the systematic review indicated that some self-reported measures were well developed (e.g. [6-8]) validity evidence was generally less strong compared to objective measurements. The dependence of weight status on reporting (likely attributable to social desirability bias) was apparent in findings from self-reported measures (e.g. [9-11]) and was an issue discussed by experts incorporating wider evidence [12, 13]. For some outcome domains, it is not possible (e.g. psychological well-being) or feasible (e.g. dietary assessment) to use objective measures.

Measures identified by the review included those that assess sedentary behaviour, which would capture specific sedentary activities (e.g. time/frequency of watching TV); and sedentary time, which measures the total time spent being inactive. Only one sedentary time outcome measure; 'accelerometry' was recommended by experts. Accelerometers are not able to measure sedentary behaviours; only sedentary time. Thus, experts have only recommended a measure of sedentary time. Experts agreed that objective measurement of physical activity and sedentary time (and behaviour) will continue to improve and newer measures such as Actiheart and Sensewear bands may be recommended in the future.

Caveats for almost all recommended dietary measures are noted; primarily related to the need to conduct further evaluation for validity and reliability evidence. All of the recommended dietary measures were food frequency questionnaires (FFQs). Exclusion of food diaries and recall methodologies was based on evidence presented by the review, suggesting that validity of these measures was poor in obese children. Additionally, evidence of reliability was lacking, with no test re-test reliability evaluation conducted in the identified food diary studies and in only two studies evaluating it for recall methodologies in obese children. Overall however, it was difficult to identify a measure of diet that all experts agreed they would highly recommend. It was acknowledged that many decisions made by experts were applicable to the intended use as a secondary outcome measures in trials evaluating childhood obesity treatments, which may not apply in other study designs or different populations. For example, experts did not suggest that food diaries should not be advocated in studies with a primary outcome of diet. When choosing a diet measure, the original methodology manuscript should be reviewed to ensure that it is robust for nutrients or foods that will be targets for change in an intervention, since validity and reliability findings usually differ between these.

Experts noted that physiological outcomes have potential to be primary outcomes given that they are indicators of cardiovascular health are associated with obesity. Furthermore, evidence presented by the review and wider evidence outside of obesity research indicates that many physiological outcomes can be measured with a high degree of precision (and are often feasible to obtain from routine clinical measurement). However, within evidence specific to research in obese children, only 'indices of insulin sensitivity' offered a sufficient degree of validity evidence (with many studies demonstrating criterion validity comparing against a gold standard of the Hyperinsulinaemic-euglycaemic clamp test). It is important to note that there was considerable debate around what constitutes a clinically meaningfulness change of physiological measures for childhood obesity

treatment evaluations. A further scoping search was conducted by the team with inclusion of terms specific to all physiological measures and criteria /cut-offs to determine whether wider evidence of what is clinically meaningful existed outside the knowledge of the experts. However, this did not identify any further data within an obesity paediatric population. Given that other outcome domains also lack information on what is clinically meaningful, the team decided to continue to advocate 'indices of insulin sensitivity' to the framework. Experts agreed that these offer good surrogates for insulin sensitivity, provided pubertal status is taken into account. There was some concern about the sensitivity of these indices in small samples, and other methods to assess insulin sensitivity may be more appropriate for individuals or small groups (eg. hyperglycemic clamp). However, there are clear practical limitations to their use in children.

All identified QoL measures in the review lacked preference weights and are therefore not able to calculate Quality Adjusted Life Years (QALYs). Instead, these measures derive scores for varying dimensions of health statuses. They have been defined as Health Related Quality of Life measures within the CoOR framework. They should not be considered as outcome measures specifically for economic evaluation unless used in cost effectiveness evaluations of interventions with a primary target of QoL. However, for evaluations of childhood obesity interventions, a more likely measure to establish cost effectiveness is that of the primary outcome (i.e. cost per unit of reduction in BMI). The team are aware of research in which utility measures are being developed for use in obese paediatric populations. Unfortunately, these were not available at the time of the review.

To conclude, the CoOR framework provides clear guidance to researchers and those working in policy and practice regarding recommended measures for use in evaluations of childhood obesity treatment interventions. This should encourage a greater adoption of well validated tools and ensure comparability between different studies or treatment interventions. It is recommended that further research should be conducted in the development and evaluation of preference based measures for cost utility analysis in line with NICE guidance. Further research is also recommended to ascertain responsiveness of the recommended measures (ability of a measure to measure a clinically important change). Ascertainment of a minimally important difference is also recommended and should be based on consensus by clinical and academic experts and by children and their parents. Finally, there is also a lack of consistency within measures used in the evaluation of treatment of obesity in adults; and it is suggested that similar work to is conducted to fill this gap in evidence.

Conflict of Interest

Author SB is an independent advisory committee member for Tanita Ltd. All other authors have no conflicts of interest. All authors had some financial support from an NIHR HTA grant for the submitted work, but have had no financial relationships with any organisations that might have an interest in the submitted work in the previous 3 years and no other relationships or activities that could appear to have influenced the submitted work.

Acknowledgements

All authors contributed to the design and management of the study throughout. JW and LA led the systematic review strategy and process, which was reviewed by all authors. MB and LA led the appraisal process and all collaborators contributed to the external appraisal and final decision making. All authors were involved in writing the paper had final approval of the submitted and published versions.

This work was funded by an NIHR HTA grant (09/127/07). This paper presents independent research commissioned by the National Institute for Health Research (NIHR). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. No funding bodies had any role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Table legend

Table 1. Included primary and secondary outcomes

Table 2. The CoOR outcome measures framework

Figure 1. Flowchart of methodology

References

1. Holloway RG, Dick AW. Clinical trial end points: on the road to nowhere? *Neurology*. 2002.**58**(5):679-86.

2. Sinha I, Jones L, Smyth RL, Williamson PR. A Systematic Review of Studies That Aim to Determine Which Outcomes to Measure in Clinical Trials in Children *PLoS Medicine*. 2008.**5**(4):e96.

3. US Department of Health and Human Services FDA. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. *Health Qual Life Outcomes*. 2006.**4**:doi: 10.1186/477-7525-4-79.

4. Scientific Advisory Committee of the Medical Outcomes Trust. Assessing health status and quality of life instruments: Attributes and review criteria. *Quality of Life Research*. 2002.**11**:193-205.

5. Vaughn A, Tabak R, Bryant M, Ward D. Measuring parent food practices: a systematic review of existing measures and examination of instruments. *International Journal of Behavioral Nutrition and Physical Activity*. 2013.**10**(1):61. PubMed PMID: doi:10.1186/1479-5868-10-61.

6. Moore JB, Hanes Jr JC, Barbeau P, Gutin B, Trevino RP, Yin Z. Validation of the physical activity questionnaire for older children in children of different races. *Pediatr Exerc Sci*. 2007. Feb;**19 (1)**:6-19. PubMed PMID: 2007365587.

7. Janz KF, Lutuchy EM, Wenthe P, Levy SM. Measuring activity in children and adolescents using self-report: PAQ-C and PAQ-A. *Medicine and science in sports and exercise*. 2008.**40**(4):767.

8. Sallis JF, Buono MJ, Roby JJ, Micale FG, Nelson JA. 7-DAY RECALL AND OTHER PHYSICAL-ACTIVITY SELF-REPORTS IN CHILDREN AND ADOLESCENTS. *Medicine and Science in Sports and Exercise*. 1993. Jan;**25**(1):99-108. PubMed PMID: WOS:A1993KG53700014. 10.1249/00005768-199301000-00014.

9. Sjoberg A, Slinde F, Arvidsson D, Ellegard L, Gramatkovski E, Hallberg L, Hulthen L. Energy intake in Swedish adolescents: validation of diet history with doubly labelled water.

Eur J Clin Nutr. 2003. Dec;**57**(12):1643-52. PubMed PMID: WOS:000186844300018. 10.1038/sj.ejcn.1601892.

10. Maffeis C, Schutz Y, Zaffanello M, Piccoli R, Pinelli L. Elevated energy expenditure and reduced energy intake in obese prepubertal children: paradox of poor dietary reliability in obesity? *J Pediatr*. 1994. Mar;**124**(3):348-54. PubMed PMID: 8120702.

11. Sithole F, Veugelers PJ. Parent and child reports of children's activity. *Health Rep*. 2008. Sep;**19**(3):19-24. PubMed PMID: 18847142.

12. Klesges LM, Baranowski T, Beech B, Cullen K, Murray DM, Rochon J, Pratt C. Social desirability bias in self-reported dietary, physical activity and weight concerns measures in 8- to 10-year-old African-American girls: results from the Girls health Enrichment Multisite Studies (GEMS). *Preventive Medicine*. 2004.**38, Supplement**(0):78-87. http://dx.doi.org/10.1016/j.ypmed.2003.07.003.

13. Goran MI. Measurement issues related to studies of childhood obesity: Assessment of body composition, body fat distribution, physical activity and food intake. *Pediatrics*. 1998.**101**:505-18. PubMed PMID: A-3.

14. Rockett HRH, Berkey CS, Colditz GA. Comparison of a short food frequency questionnaire with the Youth/Adolescent Questionnaire in the Growing Up Today Study. *Int J Pediatr Obes*. 2007.**2**(1):31-9. PubMed PMID: WOS:000245117100004. 10.1080/17477160601095417.

15. Rockett HR, Wolf AM, Colditz GA. Development and reproducibility of a food frequency questionnaire to assess diets of older children and adolescents. *J Am Diet Assoc*. 1995. Mar;**95**(3):336-40. PubMed PMID: 7860946. Epub 1995/03/01. S0002-8223(95)00086-0 [pii]

10.1016/S0002-8223(95)00086-0.

16. Rockett HRH, Colditz GA. Assessing diets of children and adolescents. *Am J Clin Nutr*. 1997. Apr;**65 (4 SUPPL.)**:1116S-22S. PubMed PMID: 1997106983.

17. Perks SM, Roemmich JN, Sandow-Pajewski M, Clark PA, Thomas E, Weltman A, Patrie J, Rogol AD. Alterations in growth and body composition during puberty. IV. Energy intake estimated by the youth-adolescent food-frequency questionnaire: validation by the doubly labeled water method. *Am J Clin Nutr*. 2000.**72**(6):1455-60. PubMed PMID: 20003036192.

18. Lanfer ALA, Hebestreit A, Ahrens W, Krogh V, Sieri S, Lissner L, Eiben G, Siani A, Huybrechts I, Loit HM, Papoutsou S, Kovacs E, Pala V, Consortium I. Reproducibility of food consumption frequencies derived from the Children's Eating Habits Questionnaire used in the IDEFICS study. *Int J Obes (Lond)*. 2011. Apr;**35**:S61-S8. PubMed PMID: WOS:000289515100008. 10.1038/ijo.2011.36.

19. Huybrechts IHI, Bornhorst C, Pala V, Moreno LA, Barba G, Lissner L, Fraterman A, Veidebaum T, Hebestreit A, Sieri S, Ottevaere C, Tornaritis M, Molnar D, Ahrens W, De Henauw S, Consortium I. Evaluation of the Children's Eating Habits Questionnaire used in the IDEFICS study by relating urinary calcium and potassium to milk consumption frequencies among European children. *Int J Obes (Lond)*. 2011. Apr;**35**:S69-S78. PubMed PMID: WOS:000289515100009. 10.1038/ijo.2011.37.

20. Watson JF, Collins CE, Sibbritt DW, Dibley MJ, Garg ML. Reproducibility and comparative validity of a food frequency questionnaire for Australian children and adolescents. *International Journal of Behavioral Nutrition and Physical Activity*. 2009. Sep;**6**. PubMed PMID: WOS:000270482400001. 62

10.1186/1479-5868-6-62.

21. Burrows T, Warren JM, Baur LA, Collins CE. Impact of a child obesity intervention on dietary intake and behaviors. *Int J Obes (Lond)*. 2008. Oct;**32**(10):1481-8. PubMed PMID: 18607380.

22. Prochaska JJ, Sallis JF, Rupp J. Screening measure for assessing dietary fat intake among adolescents. *Preventive medicine*. 2001.**33**(6):699-706.

23. Metcalf PA, Scragg RK, Sharpe S, Fitzgerald ED, Schaaf D, Watts C. Short-term repeatability of a food frequency questionnaire in New Zealand children aged 1-14 y. *Eur J Clin Nutr*. 2003. Nov;**57**(11):1498-503. PubMed PMID: 14576765. Epub 2003/10/25. 10.1038/sj.ejcn.1601717

1601717 [pii].

24. Baughcum AE, Powers SW, Johnson SB, Chamberlin LA, Deeks CM, Jain A, Whitaker RC. Maternal Feeding Practices and Beliefs and Their Relationships to Overweight in Early Childhood. *Journal of Developmental & Behavioral Pediatrics*. 2001.**22**(6):391-408.

25. van Strien T, Oosterveld P. The children's DEBQ for assessment of restrained, emotional, and external eating in 7- to 12-year-old children. *Int J Eat Disord*. 2008.**41**(1):72-81. PubMed PMID: 2009745090. Language: English. Entry Date: 20080418. Revision Date: 20091218. Publication Type: journal article.

26. Banos RM, Cebolla A, Etchemendy E, Felipe S, Rasal P, Botella C. Validation of the dutch eating behavior questionnaire for children (DEBQ-C) for use with spanish children. *Nutr Hosp*. 2011. Jul-Aug;**26**(4):890-8. PubMed PMID: WOS:000293152000032. 10.3305/nh.2011.26.4.5238.

27. Braet C, Soetens B, Moens E, Mels S, Goossens L, Van Vlierberghe L. Are two informants better than one? Parent-child agreement on the eating styles of children who are overweight. *Eur*. 2007. Nov;**15**(6):410-7. PubMed PMID: 17960860.

28. Caccialanza R, Nicholls D, Cena H, Maccarini L, Rezzani C, Antonioli L, Dieli S, Roggi C. Validation of the Dutch Eating Behaviour Questionnaire parent version (DEBQ-P) in the Italian population: a screening tool to detect differences in eating behaviour among obese, overweight and normal-weight preadolescents. *Eur J Clin Nutr*. 2004. Sep;**58**(9):1217-22. PubMed PMID: 15054434.

29. Braet C, Van Strien T. Assessment of emotional, externally induced and restrained eating behaviour in nine to twelve-year-old obese and non-obese children. *Behav Res Ther*. 1997. Sep;**35**(9):863-73. PubMed PMID: 9299807.

30. Tanofsky-Kraff M, Theim KR, Yanovski SZ, Bassett AM, Burns NP, Ranzenhofer LM, Glasofer DR, Yanovski JA. Validation of the emotional eating scale adapted for use in children and adolescents (EES-C). *Int J Eat Disord*. 2007. Apr;**40 (3)**:232-40. PubMed PMID: 2007150664. http://dx.doi.org/10.1002/eat.20362.

31. Birch LL, Fisher JO, Grimm-Thomas K, Markey CN, Sawyer R, Johnson SL. Confirmatory factor analysis of the Child Feeding Questionnaire: a measure of parental attitudes, beliefs and practices about child feeding and obesity proneness. *Appetite*. 2001. Jun;**36**(3):201-10. PubMed PMID: 11358344.

32. Haycraft EL, Blissett JM. Maternal and paternal controlling feeding practices: Reliability and relationships with BMI. *Obesity (Silver Spring)*. 2008. Jul;**16**(7):1552-8. PubMed PMID: WOS:000257325300013. 10.1038/oby.2008.238.

33. Anderson CB, Hughes SO, Fisher JO, Nicklas TA. Cross-cultural equivalence of feeding beliefs and practices: the psychometric properties of the child feeding questionnaire among Blacks and Hispanics. *Prev Med*. 2005. Aug;**41**(2):521-31. PubMed PMID: 15917048.

34. Corsini N, Danthiir V, Kettler L, Wilson C. Factor structure and psychometric properties of the Child Feeding Questionnaire in Australian preschool children. *Appetite*. 2008. Nov;**51**(3):474-81. PubMed PMID: 18499301.

35. Polat S, Erci B. Psychometric Properties of the Child Feeding Scale in Turkish Mothers. *Asian Nurs Res*. 2010. Sep;**4**(3):111-21. PubMed PMID: WOS:000288578800001.

36. Boles RE, Nelson TD, Chamberlin LA, Valenzuela JM, Sherman SN, Johnson SL, Powers SW. Confirmatory factor analysis of the Child Feeding Questionnaire among low-income African American families of preschool children. *Appetite*. 2010. Apr;**54**(2):402-5. PubMed PMID: WOS:000276005100024. 10.1016/j.appet.2009.12.013.

37. Thompson AL, Mendez MA, Borja JB, Adair LS, Zimmer CR, Bentley ME. Development and validation of the Infant Feeding Style Questionnaire. *Appetite*. 2009. Oct;**53**(2):210-21. PubMed PMID: WOS:000270475900008. 10.1016/j.appet.2009.06.010.

38. Sleddens EFC, Kremers SPJ, Thijs C. The Children's Eating Behaviour Questionnaire: Factorial validity and association with Body Mass Index in Dutch children aged 6-7. *International Journal of Behavioral Nutrition and Physical Activity*. 2008. 20 Oct;**5**(49). PubMed PMID: 2009017453. http://dx.doi.org/10.1186/1479-5868-5-49.

39. Wardle J, Guthrie CA, Sanderson S, Rapoport L. Development of the children's eating behaviour questionnaire. *Journal of Child Psychology and Psychiatry and Allied Disciplines*. 2001.**42 (7)**:963-70. PubMed PMID: 2001380198.

40. Corsini N, Wilson C, Kettler L, Danthiir V. Development and preliminary validation of the Toddler Snack Food Feeding Questionnaire. *Appetite*. 2010. Jun;**54**(3):570-8. PubMed PMID: Peer Reviewed Journal: 2010-12652-003. http://dx.doi.org/10.1016/j.appet.2010.03.001.

41. Monnery-Patris S, Rigal N, Chabanet C, Boggio V, Lange C, Cassuto DA, Issanchou S. Parental practices perceived by children using a French version of the Kids' Child Feeding Questionnaire. *Appetite*. 2011. August;**57 (1)**:161-6. PubMed PMID: 2011345940. http://dx.doi.org/10.1016/j.appet.2011.04.014.

42. Carper JL, Orlet Fisher J, Birch LL. Young girls' emerging dietary restraint and disinhibition are related to parental control in child feeding. *Appetite*. 2000.**35**(2):121-9. 10.1006/appe.2000.0343.

43. Murashima M, Hoerr SL, Hughes SO, Koplowitz S. Confirmatory factor analysis of a questionnaire measuring control in parental feeding practices in mothers of Head Start children. *Appetite*. 2011. Jun;**56**(3):594-601. PubMed PMID: WOS:000291912700007. 10.1016/j.appet.2011.01.031.

44. Tanofsky-Kraff M, Ranzenhofer LM, Yanovski SZ, Schvey NA, Faith M, Gustafson J, Yanovski JA. Psychometric properties of a new questionnaire to assess eating in the absence of hunger in children and adolescents. *Appetite*. 2008. Jul;**51**(1):148-55. PubMed PMID: 18342988. Pubmed Central PMCID: Source: NLM. NIHMS51668

Source: NLM. PMC2424281.

45. Guinhouya CB, Apete GK, Hubert H. Diagnostic quality of Actigraph-based physical activity cut-offs for children: what overweight/obesity references can tell? *Pediatr Int*. 2009. Aug;**51**(4):568-73. PubMed PMID: 19674366.

46. Coleman KJ, Saelens BE, Wiedrich-Smith MD, Finn JD, Epstein LH. Relationships between TriTrac-R3D vectors, heart rate, and self-report in obese children. *Med Sci Sports Exerc*. 1997. Nov;**29**(11):1535-42. PubMed PMID: 9372493.

47. Noland M, Danner F, DeWalt K, McFadden M, Kotchen JM. The measurement of physical activity in young children. *Res Q Exerc Sport*. 1990. Jun;**61**(2):146-53. PubMed PMID: 2094925.

48. Pate RR, Almeida MJ, McIver KL, Pfeiffer KA, Dowda M. Validation and calibration of an accelerometer in preschool children. *Obesity (Silver Spring)*. 2006. Nov;**14**(11):2000-6. PubMed PMID: 17135617.

49. Kelly LA, Reilly JJ, Fairweather SC, Barrie S, Grant S, Paton JY. Comparison of two accelerometers for assessment of physical activity in preschool children. *Pediatr Exerc Sci*. 2004. Nov;**16**(4):324-33. PubMed PMID: WOS:000225223500003.

50. Duncan MJ, Al-Nakeeb Y, Woodfield L, Lyons M. Pedometer determined physical activity levels in primary school children from central England. *Prev Med*. 2007. May;**44**(5):416-20. PubMed PMID: WOS:000247161600008. 10.1016/j.ypmed.2006.11.019.

51. Kilanowski CK, Consalvi AR, Epstein LH. Validation of an electronic pedometer for measurement of physical activity in children. *Pediatr Exerc Sci*. 1999. Feb;**11**(1):63-8. PubMed PMID: WOS:000078375000006.

52. Treuth MS, Sherwood NE, Butte NF, McClanahan B, Obarzanek E, Zhou A, Ayers C, Adolph A, Jordan J, Jacobs DR, Rochon J. Validity and reliability of activity measures in African-American girls for GEMS. *Med Sci Sports Exerc*. 2003. Mar;**35**(3):532-9. PubMed PMID: 12618587. Epub 2003/03/06. 10.1249/01.MSS.0000053702.03884.3F.

53. Jago R, Watson K, Baranowski T, Zakeri I, Yoo S, Baranowski J, Conry K. Pedometer reliability, validity and daily activity targets among 10- to 15-year-old boys. *J Sports Sci*. 2006. Mar;**24**(3):241-51. PubMed PMID: 16368634.

54. Mitre N, Lanningham-Foster L, Foster R, Levine JA. Pedometer accuracy for children: can we recommend them for our obese population? *Pediatrics*. 2009. Jan;**123**(1):e127-31. PubMed PMID: 19117834. Pubmed Central PMCID: Source: NLM. NIHMS105326

Source: NLM. PMC2678845.

55. Brown WH, Pfeiffer KA, McIver KL, Dowda M, Almeida M, Pate RR. Assessing preschool children's physical activity: The observational system for recording physical activity in children-preschool version. *Research Quarterly for Exercise and Sport*. 2006. Jun;**77**(2):167-76. PubMed PMID: WOS:000239154400003.

56. Ridgers ND, Stratton G, McKenzie TL. Reliability and validity of the System for Observing Children's Activity and Relationships during Play (SOCARP). *J Phys Act Health*. 2010. Jan;**7**(1):17-25. PubMed PMID: 20231751.

57. Puyau MR, Adolph AL, Vohra FA, Butte NF. Validation and calibration of physical activity monitors in children. *Obes Res*. 2002. Mar;**10**(3):150-7. PubMed PMID: 11886937. Epub 2002/03/12. 10.1038/oby.2002.24.

58. Reilly JJ, Coyle J, Kelly L, Burke G, Grant S, Paton JY. An objective method for measurement of sedentary behavior in 3-to 4-year olds. *Obesity*. 2003.**11**(10):1155-8.

59. Loftin M, Sothern M, Warren B, Udall J. Comparison of VO2 peak during treadmill and cycle ergometry in severely overweight youth. *Journal of Sports Science and Medicine*. 2004. Dec;**3**(4):254-60. PubMed PMID: WOS:000228271100008.

60. Rossner SM, Neovius M, Montgomery SM, Marcus C, Norgren S. Alternative methods of insulin sensitivity assessment in obese children and adolescents. *Diabetes Care*. 2008. Apr;**31**(4):802-4. PubMed PMID: 18202247.

61. Keskin M, Kurtoglu S, Kendirci M, Atabek ME, Yazici C. Homeostasis model assessment is more reliable than the fasting glucose/insulin ratio and quantitative insulin sensitivity check index for assessing insulin resistance among obese children and adolescents. *Pediatrics*. 2005. Apr;**115**(4):e500-3. PubMed PMID: 15741351.

62. Atabek ME, Pirgon O. Assessment of insulin sensitivity from measurements in fasting state and during an oral glucose tolerance test in obese children. *J Pediatr Endocrinol*. 2007. Feb;**20**(2):187-95. PubMed PMID: 17396435.

63. Gungor N, Saad R, Janosky J, Arslanian S. Validation of surrogate estimates of insulin sensitivity and insulin secretion in children and adolescents. *J Pediatr*. 2004. Jan;**144**(1):47-55. PubMed PMID: 14722518.

64. George L, Bacha F, Lee S, Tfayli H, Andreatta E, Arslanian S. Surrogate estimates of insulin sensitivity in obese youth along the spectrum of glucose tolerance from normal to prediabetes to diabetes. *Journal of Clinical Endocrinology and Metabolism*. 2011. 01 Jul;**96 (7)**:2136-45. PubMed PMID: 2011375809. http://dx.doi.org/10.1210/jc.2010-2813.

65. Conwell LS, Trost SG, Brown WJ, Batch JA. Indexes of insulin resistance and secretion in obese children and adolescents: a validation study. *Diabetes Care*. 2004. Feb;**27**(2):314-9. PubMed PMID: 14747206.

66. Yeckel CW, Weiss R, Dziura J, Taksali SE, Dufour S, Burgert TS, Tamborlane WV, Caprio S. Validation of insulin sensitivity indices from oral glucose tolerance test parameters in obese children and adolescents. *J Clin Endocrinol Metab*. 2004. Mar;**89**(3):1096-101. PubMed PMID: 15001593.

67. Uwaifo GI, Fallon EM, Chin J, Elberg J, Parikh SJ, Yanovski JA. Indices of insulin action, disposal, and secretion derived from fasting samples and clamps in normal glucose-tolerant black and white children. *Diabetes Care*. 2002b. Nov;**25**(11):2081-7. PubMed PMID: 12401760.

68. Schwartz B, Jacobs DR, Moran A, Steinberger J, Hong CP, Sinaiko AR. Measurement of insulin sensitivity in children. *Diabetes Care*. 2008. Apr;**31**(4):783-8. PubMed PMID: WOS:000254591900030. 10.2337/dc07-1376.

69. Gunczler P, Lanes R. Relationship between different fasting-based insulin sensitivity indices in obese children and adolescents. *J Pediatr Endocrinol*. 2006. Mar;**19**(3):259-65. PubMed PMID: 16607927.

70. Waters E, Salmon L, Wake M, Hesketh K, Wright M. The Child Health Questionnaire in Australia: reliability, validity and population means. *Aust N Z J Public Health*. 2000. Apr;**24**(2):207-10. PubMed PMID: 10790944. Epub 2000/05/03.

71. Waters E, Salmon L, Wake M. The parent-form Child Health Questionnaire in Australia: comparison of reliability, validity, structure, and norms. *Journal of pediatric psychology*. 2000b.**25**(6):381-91.

72. Landgraf JM, Maunsell E, Speechley KN, Bullinger M, Campbell S, Abetz L, Ware JE. Canadian-French, German and UK versions of the Child Health Questionnaire: methodology and preliminary item scaling results. *Qual Life Res*. 1998. Jul;**7**(5):433-45. PubMed PMID: 9691723. Epub 1998/08/06.

73. Ravens-Sieberer U SS, Gosch A, Erhart M, Petersen C, Bullinger M:. Measuring subjective health in children and adolescents: results of the European KIDSCREEN/DISABKIDS Project. emotional and external eating behavior. *Psychosoc Med* 2007.**4**(8).

74. Burstrom K, Svartengren M, Egmar AC. Testing a Swedish child-friendly pilot version of the EQ-5D instrument - initial results. *Eur J Public Health*. 2011.**21**(2):178-83. PubMed PMID: 20113140552. http://dx.doi.org/10.1093/eurpub/ckq042.

75. Burstrom K, Egmar AC, Lugner A, Eriksson M, Svartengren M. A Swedish child-friendly pilot version of the EQ-5D instrument--the development process. *Eur J Public Health*. 2011b. Apr;**21**(2):171-7. PubMed PMID: 20430804. Epub 2010/05/01. ckq037 [pii]

10.1093/eurpub/ckq037.

76. Wille N, Bullinger M, Holl R, Hoffmeister U, Mann R, Goldapp C, Reinehr T, Westenhofer J, Egmond-Froehlich A, Ravens-Sieberer U. Health-related quality of life in overweight and obese youths: results of a multicenter study. *Health Qual Life Outcomes*. 2010.**8**:36. PubMed PMID: 20374656. Pubmed Central PMCID: Source: NLM. PMC2868813.

77. Ravens-Sieberer U, Wille N, Badia X, Bonsel G, Burstrom K, Cavrini G, Devlin N, Egmar AC, Gusi N, Herdman M, Jelsma J, Kind P, Olivares PR, Scalone L, Greiner W. Feasibility, reliability, and validity of the EQ-5D-Y: results from a multinational study. *Qual Life Res*. 2010. Aug;**19**(6):887-97. PubMed PMID: 20401552. Pubmed Central PMCID: 2892614. Epub 2010/04/20. 10.1007/s11136-010-9649-x.

78. Kolotkin RL, Zeller M, Modi AC, Samsa GP, Quinlan NP, Yanovski JA, Bell SK, Maahs DM, de Serna DG, Roehrig HR. Assessing weight-related quality of life in adolescents. *Obesity (Silver Spring)*. 2006. Mar;**14**(3):448-57. PubMed PMID: 16648616. Pubmed Central PMCID: Source: NLM. NIHMS41943

Source: NLM. PMC2374918.

79. Modi AC, Zeller MH. The IWQOL-Kids: Establishing minimal clinically important difference scores and test-retest reliability. *Int J Pediatr Obes*. 2011. June;**6 (2 -2)**:e94-e6. PubMed PMID: 2011369759. http://dx.doi.org/10.3109/17477166.2010.500391.

80. Erhart M, Ellert U, Kurth B-M, Ravens-Sieberer U. Measuring adolescents' HRQoL via self reports and parent proxy reports: an evaluation of the psychometric properties of both versions of the KINDL-R instrument. *Health Qual Life Outcomes*. 2009.**7**:77. PubMed PMID: 19709410. Pubmed Central PMCID: Source: NLM. PMC2749015.

81. Varni JW, Burwinkle TM, Seid M, Skarr D. The PedsQL™* 4.0 as a Pediatric Population Health Measure: Feasibility, Reliability, and Validity. *Ambulatory Pediatrics*. 2003.**3**(6):329-41.

82. Varni JW, Seid M, Kurtin PS. PedsQL (TM) 4.0: Reliability and validity of the Pediatric Quality of Life Inventory (TM) version 4.0 Generic Core Scales in healthy and patient populations. *Medical care*. 2001.**39**(8):800.

83. Hughes AR, K Farewell, Harris D, Reilly J. Quality of life in a clinical sample of obese children. *International Journal of Obesity*. 2007.**31**:39-44.

84. Modi AC, Zeller MH. Validation of a parent-proxy, obesity-specific quality-of-life measure: sizing them up. *Obesity (Silver Spring)*. 2008. Dec;**16**(12):2624-33. PubMed PMID: 18833211.

85. Zeller MH, Modi AC. Development and initial validation of an obesity-specific quality-of-life measure for children: sizing me up. *Obesity (Silver Spring)*. 2009. Jun;**17**(6):1171-7. PubMed PMID: 19265795.

86. Morales LS, Edwards TC, Flores Y, Barr L, Patrick DL. Measurement properties of a multicultural weight-specific quality-of-life instrument for children and adolescents. *Qual Life Res*. 2011.**20**(2):215-24. PubMed PMID: 20113128021. http://dx.doi.org/10.1007/s11136-010-9735-0.

87. Truby H, Paxton SJ. Development of the Children's Body Image Scale. *British Journal of Clinical Psychology*. 2002.**41 (2)**:185-203. PubMed PMID: 2002226512. http://dx.doi.org/10.1348/014466502163967.

88. Collins ME. Body figure perceptions and preferences among preadolescent children. *International Journal of Eating Disorders*. 1991.**10**(2):199-208.

89. Van Dongen-Melman J, Koot H, Verhulst F. Cross-cultural validation of Harter's self-perception profile for children in a Dutch sample. *Educational and Psychological Measurement*. 1993.**53**(3):739-53.

90. Harter S. The perceived competence scale for children. *Child development*. 1982.87-97.

91. Motl RW, Dishman RK, Saunders R, Dowda M, Felton G, Pate RR. Measuring enjoyment of physical activity in adolescent girls. *American journal of preventive medicine*. 2001.**21**(2):110-7.

92. Whitehead JR. A study of children's physical self-perceptions using an adapted physical self-perception profile questionnaire. *Pediatric exercise science*. 1995.**7**:132-.

93. Eklund RC, Whitehead JR, Welk GJ. Validity of the children and youth physical self-perception profile: a confirmatory factor analysis. *Res Q Exerc Sport*. 1997. Sep;**68**(3):249-56. PubMed PMID: 9294879. Epub 1997/09/19.

94. Hay JA. Adequacy in and predilection for physical activity in children. *Clinical Journal of Sport Medicine*. 1992.**2**(3):192.

95. Stein RJ, Bracken BA, Haddock CK, Shadish WR. Preliminary development of the Children's Physical Self-Concept Scale. *J Dev Behav Pediatr*. 1998. Feb;**19**(1):1-8. PubMed PMID: 9524299.

96. La Greca AM, Stone WL. Social anxiety scale for children-revised: Factor structure and concurrent validity. *Journal of clinical child psychology*. 1993.**22**(1):17-27.

97. La Greca AM, Dandes SK, Wick P, Shaw K, Stone WL. Development of the Social Anxiety Scale for Children: Reliability and concurrent validity. *Journal of Clinical Child Psychology*. 1988.**17**(1):84-91.

98. Mendelson BK, White DR. Relation between body-esteem and self-esteem of obese and normal children. *Percept Mot Skills*. 1982. Jun;**54**(3):899-905. PubMed PMID: 7099901.

99. Benjamin SE, Neelon B, Ball SC, Bangdiwala SI, Ammerman AS, Ward DS. Reliability and validity of a nutrition and physical activity environmental self-assessment for child care. *International Journal of Behavioral Nutrition and Physical Activity*. 2007. Jul;**4**. PubMed PMID: WOS:000249703900001. 29

10.1186/1479-5868-4-29.

100. Durant N, Kerr J, Harris SK, Saelens BE, Norman GJ, Sallis JF. Environmental and Safety Barriers to Youth Physical Activity in Neighborhood Parks and Streets: Reliability and Validity. *Pediatr Exerc Sci*. 2009. Feb;**21**(1):86-99. PubMed PMID: WOS:000263772300009.

101. Gattshall ML, Shoup JA, Marshall JA, Crane LA, Estabrooks PA. Validation of a survey instrument to assess home environments for physical activity and healthy eating in overweight children. *International Journal of Behavioral Nutrition and Physical Activity*. 2008. 11 Jan;**5**(3). PubMed PMID: 2008103009. http://dx.doi.org/10.1186/1479-5868-5-3.

102. Rosenberg DE, Sallis JF, Kerr J, Maher J, Norman GJ, Durant N, Harris SK, Saelens BE. Brief scales to assess physical activity and sedentary equipment in the home.

International Journal of Behavioral Nutrition and Physical Activity. 2010. 31 Jan;**7**(10).

PubMed PMID: 2010176362. http://dx.doi.org/10.1186/1479-5868-7-10.

Supplementary Table 1. Criteria used to allocate robustness scores for evaluation of quality

Measurement development and reporting			
The concept to be measured was clearly stated (rationale and description).		4=strongly agree (concepts are named and clearly defined) 3=agree (concepts are named and general described) 2=disagree (concepts only named, but not defined) 1=strongly disagree (concepts are not clearly named or defined)	
Was a theoretical or conceptual framework used or referenced?		4=strongly agree (theory/framework used as a basis for development) 3=agree (theory/framework named and incorporated) 2=disagree (theory/framework named but not used) 1=strongly disagree (no theory/framework described) 0=N/A= (biochemical/anthropometry, direct measures/ observations)	
Populations that the measure was intended for were adequately described.		4=strongly agree (describes at least 4 characteristics including: age, gender, race/ethnicity, and SES) 3=agree (3 characteristics reported) 2=disagree (2 characteristics reported) 1=strongly disagree (no characteristics reported)	
Were the populations that the measure was intended for involved in measurement development?		4=strongly agree (at least 3 methods of involvement including: part of study team, steering committee, pilot testing, cognitive interviews/focus groups) 3=agree (involved using at least 2 methods) 2=disagree (populations minimally involved in 1 method) 1=strongly disagree (populations not involved) 0=N/A (biochemical/anthropometry)	
Measurement evaluation			
	Sample size	**Appropriate stats**[1]	**Results/findings**
Internal consistency	≥5 participants per item	Cronbach α KR-20 Split half	α ≥0.7
Test re-test reliability	≥50	Spearman Pearson Kappa Agreement	r ≥0.4 Kappa ≥0.4 Agreement (not used to score- but reported for comparisons)
Inter-rater reliability	Study specific (depending on design)	Pearsons/ICC/rho= Kappa K= Kripendorffs alpha	r ≥0.4 Kappa >0.40
Factor analysis	≥5 participants per item	Eignevalue Factor loading % variance	Eigenvalue= >1 Factor loading= High >0.6, Low<0.4 CFA RNSEA <0.06, RNI close to 1
Criterion validity	≥50 (less for objective such as DWL (=>20)	Pearson Spearman Regression Agreement T-test (not in isolation) ANOVA	Pearsons/Spearman=>0.4 Regression coefficient = p >0.5 or r=>0.50 Agreement T-test p >0.05, T Value >1. AUC >0.7

		Sensitivity/specificity	
Convergent validity	≥100	Pearson Spearman Regression Agreement T-test (not in isolation) ANOVA Sensitivity/specificity	Pearson/Spearman=>0.4 Regression coefficient = p >0.5 or r=>0.50 Agreement T-test p >0.05, T Value >1. AUC >0.7
Construct validity	≥100	Pearson Spearman Regression Agreement T-test (not in isolation) ANOVA Sensitivity/specificity	Pearsons/Spearman=>0.4 Regression coefficient = p >0.5 or r=>0.50 Agreement T-test p >0.05, T Value >1. AUC >0.7
Responsive-ness	≥100	MCID SRM ROC AUC ES t-test	ROC AUC>0.7 ES >0.5 MCID/SRM >0.5 t-test p<0.05

[1]The protocol for consideration of statistical tests that were not listed included consideration by the team statistician (JB).