# Systematic review to identify and appraise outcome measures used to evaluate childhood obesity treatment interventions (CoOR): evidence of purpose, application, validity, reliability and sensitivity

Maria Bryant,[1]* Lee Ashton,[1] Julia Brown,[1] Susan Jebb,[2] Judy Wright,[3] Katharine Roberts[4] and Jane Nixon[1]

[1]Clinical Trials Research Unit, University of Leeds, Leeds, UK
[2]Medical Research Council (MRC) Human Nutrition Research, Cambridge, UK
[3]Institute of Health Sciences, University of Leeds, Leeds, UK
[4]National Obesity Observatory (NOO), Oxford, UK

*Corresponding author

**Declared competing interests of authors:** SJ is a current member of the Tanita Medical Advisory Board; past member of Nestlé Advisory Board (ceased 2010), Coca-Cola Advisory Board (ceased 2011) and Heinz Advisory Board (ceased 2011), and contributor to the Rosemary Conley *Diet & Fitness* magazine.

Published August 2014
DOI: 10.3310/hta18510

## Scientific summary

### Childhood obesity treatment interventions

Health Technology Assessment 2014; Vol. 18: No. 51
DOI: 10.3310/hta18510

NIHR Journals Library www.journalslibrary.nihr.ac.uk

# Scientific summary


## Background

The lack of uniformity in the outcome measures used in the evaluation of childhood obesity treatment interventions often impedes the ability to truly assess effectiveness and limits comparisons across trials. In part, this arises because of the lack of consensus on what outcomes are required and the most appropriate outcome measures to use within outcome domains.

## Objective

This study aimed to systematically review the literature in order to produce a database of outcome measures that have been used (or developed for use) in childhood obesity treatment interventions and to use expert appraisal to develop a framework of recommended outcome measures for use as a resource to guide researchers when designing childhood obesity treatment evaluations. Secondary objectives include (1) a summary of the description and measurement properties of all outcome measures identified and (2) a methodology to determine the quality of outcome measures and/or aid in the development of new outcome measures in this area.

## Methods

### *Search strategy*

Two searches were performed with the aim to identify (1) outcome measures that had already been used in existing evaluations by searching trials of childhood obesity treatment interventions and (2) methodology studies that developed and/or evaluated the outcome measures for childhood obesity research. Both searches were conducted in 11 databases and were supported with literature obtained from relevant citations (including reviews of measurement tools), conference proceedings and information from registered clinical trials in progress.

Search strategies were developed by the Information Specialist (JW), with contributions of search terms from the project team. Searches were agreed by the project team and conducted from the date of inception, with no language restrictions, from August to October 2011. Terms and keywords were selected for search 1 to identify manuscripts detailing randomised controlled trials (or pilot/feasibility studies) aimed at evaluating childhood obesity treatment interventions. Search 2 included keywords/terms pertaining to the development and/or evaluation of outcome measures.

### *Process of study selection*

Assessment of titles and abstracts was performed independently by two reviewers (MB, LA). Agreement between reviewers was tested after review of the first 130 search 1 papers and the first 50 search 2 papers. For search 1, 98% agreement was reached; for search 2, 96% agreement was reached. Disagreements were discussed to refine eligibility clarification. Papers were retained at title and abstract review if there was any degree of uncertainty by either reviewer. Full papers were then assessed against eligibility criteria, and disagreements were resolved by discussion. Measures had to have been developed specifically for childhood obesity research or evaluated in a paediatric obese population (or present results stratified by obesity) and included those in the following domains: anthropometry (primary outcome), diet, eating behaviours, physical activity (PA), sedentary time/behaviour, fitness, physiology, environment, psychological well-being and quality of life.

### Data extraction

Data were extracted from relevant search 1 papers (i.e. trials), including information concerning all included outcome measures used and corresponding citations of measurement development/evaluation papers. These cited papers were then located and added to search 2 (i.e. methodology) papers. Data pertaining to the sample, design, development, evaluation and feasibility of each outcome measurement development/evaluation paper from search 2 were then extracted on prespecified extraction forms. Disagreements were resolved by discussion.

### Quality assessment

Search 1 trials were not judged on quality/bias, as the study outcome information (i.e. intervention efficacy/effectiveness) was not relevant to the aims of this study. Quality assessment for measurement papers in search 2 was based on internal and external appraisal of the rigour in development and evaluation of each outcome measure. For internal appraisal, members of the internal project team (MB, LA) appraised each outcome measure related to evidence of development, reliability testing and validity testing using international guidelines for the development of patient-reported outcomes (e.g. Food and Drug Administration) and previous work already conducted by the chief investigator. This resulted in a database of outcome measures with a detailed description of each measure, in addition to a parallel assessment of quality (based on a scoring system). The internal project team then considered whether each measure was (1) fit for purpose (i.e. recommended for inclusion to the outcome measures framework); (2) not fit for purpose (i.e. not recommended for inclusion); or (3) uncertain (i.e. requires further consideration). This decision was based on existing evidence gathered and was reached by consensus. External appraisal was then conducted via an expert appraisal meeting, which was held (in person) with 10 national collaborators (plus five applicants). Collaborators were invited based on their experience and expertise within evaluation of childhood obesity interventions and/or measurement. Prior to the meeting they were provided with (1) a list of all included outcome measure development/evaluation papers alongside access to all papers; (2) tables describing each paper (summarised from the data extraction forms); and (3) internal appraisal documents, including scores for quality (e.g. for development, reliability and validity) and degree of certainty from the internal appraisal for each measure related to whether it should be included in the final framework. They were asked to review all measures but to focus on the outcome domain that was most closely aligned to their area of expertise (defined by the project team). The purpose of the meeting was then to agree on whether or not each measure was suitable for inclusion in the final outcome measures framework based on the evidence provided and any relevant personal experience/knowledge in using the measures.

### Methods of analysis/synthesis

This report provides a narrative summary of outcome measures, which are grouped according to outcome domain. Analysis of reliability and validity testing was considered for appraisal, but results were not pooled.

## Results

### Results of search strategy

A total of 25,486 papers were identified from both searches. Eligible search 1 papers (of existing evaluations) cited 417 additional papers linked to included outcome measures, of which only 56 were eligible methodology papers. A further 323 outcome development/evaluation methodology papers from search 2 met eligibility criteria. Combined, these 379 papers described 180 outcome measures.

### Results of quality assessment

Based on the reliability and validity evidence, eligible measures were appraised by the internal team, resulting in 29 outcomes that were considered to be fit for inclusion in the framework as a recommended tool (i.e. degree of certainty = 1); 35 outcomes deemed unfit for inclusion (i.e. degree of certainty = 2); and 121 requiring further consideration (i.e. degree of certainty = 3). External appraisal considered these findings alongside their experience and expertise, and concluded that 52 outcomes were fit for inclusion

across the 10 outcome domains (remaining 128 tools deemed unfit (degree of certainty = 2). Of these, two [body mass index (BMI) and dual-energy X-ray absorptiometry DXA] out of the 38 anthropometry measures were recommended. In secondary outcomes, recommended tools included 6 (out of 22) diet measures (all food frequency questionnaires); 12 (out of 22) eating behaviour measures; 4 (out of 24) PA measures (with no self-reported measures); 1 (out of 6) sedentary time measure; 1 (out of 13) fitness measure; 1 (out of 12) physiological measure; 10 (out of 12) health-related quality of life questionnaires; 10 (out of 17) psychological well-being measures; and 5 (out of 10) environmental measures.

### *The childhood obesity outcome measures framework*

Recommended outcome measures are presented by outcome domain alongside details relating to feasibility of implementation (e.g. number of items, costs, licensing, etc.). This framework is a tool to guide researchers but the final decision for inclusion of measures must be based on those that are (1) aligned with the targets of the intervention and (2) appropriate for use in a given population (e.g. age/ethnicity specific). This framework is recommended as an initial guide outcome measure selection. In exceptional cases when no measures meet the needs of a particular study, a detailed description of all measures meeting the eligibility criteria is provided so that researchers are able to self-select the most appropriate measure given the information available on its validity.

## Conclusions

The key findings of this study are:

1. Only 13% of trials correctly cited outcome measures used.
2. Approximately 20% of eligible primary and secondary outcome measures were recommended by experts.
3. Primary outcome measures recommended are BMI or DXA.
4. Objective measurement must be applied if available (i.e. use of activity monitors instead of self-reported PA).
5. Physiological outcomes have the potential to be primary outcomes (as they are measured with high precision and are related to adverse health outcomes) but, at present, there is insufficient evidence on what constitutes a clinically meaningful change (although it is recognised that this is also the case in existing primary outcomes).
6. Evidence of ability of measures to detect change was poor or lacking.
7. While new tools are pending, there remains no published preference-based measures for use in economic evaluations in this population. Cost-effectiveness should therefore include measures most pertinent to the targets of the intervention [e.g. costs per reduction in body mass index standard deviation score (BMI-SDS)].
8. The proposed recommended outcome measures are fit for use specifically within studies that evaluate childhood obesity treatment evaluations. They may or may not be suitable for other study designs.

### *Implications for clinical practice*

The results of the expert appraisal provide clear guidance to researchers about appropriate outcomes domains and recommended measures in each of these domains to encourage greater adoption of well-validated tools. This will make it easier to judge clinical effectiveness and enhance the comparability between different studies or treatment interventions. The review also provides details of other measures That may be appropriate for other settings with details of the extent of methodological testing already conduced to inform decision-making. Researchers wishing to use novel tools are recommended to adopt these alongside the recommended tools, wherever possible, to encourage evolution and the development of new knowledge. Details of the validity of each of the recommended outcome tools provide a knowledge trail to encourage more accurate reporting of these measures in future studies.

### *Implications for future research*

In the case of economic evaluations, primary research is urgently needed as this review did not identify a single measure that was able to calculate quality-adjusted life-years (although we are aware that some work in this area is in progress). In all other domains, a large number of outcome measures have been proposed, but in many cases robust evidence of validity is scant. There may be opportunities to make rapid progress with further testing and modifications, where necessary, of existing measures. Many outcome measures rely on self-report and more objective measures would add value, especially for dietary outcomes. There are also opportunities to consider the use of new technologies to replace pen-and-paper retrospective questionnaires to collect information on some outcomes measures. Given that a number of different types of outcome measures were identified within many outcome domains, findings from this study suggest that future research should invest in the modification (if appropriate) and evaluation of existing measures (not the development of new measures when others are available).

Research is needed to determine the ability of measures to detect change. For some (more historical) measures, such as BMI, evidence demonstrating a level of precision over multiple assessment periods may be sufficient. However, there is a lack of testing of responsiveness in many of the recommended questionnaire outcomes. Lastly, the lack of data describing the clinically meaningful change and/or appropriate cut-offs was noted as part of the expert appraisal, specifically for anthropometry, physiology and fitness outcomes.

## Funding

Funding for this study was provided by the Health Technology Assessment programme of the National Institute for Health Research.

# Health Technology Assessment

HTA/HTA TAR

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 5.116

*Health Technology Assessment* is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the ISI Science Citation Index and is assessed for inclusion in the Database of Abstracts of Reviews of Effects.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: nihredit@southampton.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

### Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

### HTA programme

The HTA programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

For more information about the HTA programme please visit the website: www.hta.ac.uk/

### This report

The research reported in this issue of the journal was funded by the HTA programme as project number 09/127/07. The contractual start date was in September 2011. The draft report began editorial review in February 2013 and was accepted for publication in July 2013. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health.



Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).