Journal of Clinical Epidemiology 115 (2019) 14–24

ORIGINAL ARTICLE

A systematic review of core outcome set development studies demonstrates difficulties in defining unique outcomes

Amber E. Young[a,*], Sara T. Brookes[b], Kerry N.L. Avery[c], Anna Davies[d], Chris Metcalfe[e], Jane M. Blazeby[c]

[a]*Bristol Centre for Surgical Research, Population Health Sciences, Bristol Medical School, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol, UK, BS8 2PS*

[b]*Cancer Research UK Clinical Trials Unit (CRCTU), Institute of Cancer and Genomic Sciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK*

[c]*Medical Research Council ConDuCT-II Hub for Trials Methodology Research, National Institute for Health Research Bristol Biomedical Research Centre, Bristol Centre for Surgical Research, Bristol Medical School: Population Health Sciences, University of Bristol, 39 Whatley Road, Bristol, BS8 2PS, UK*

[d]*Centre for Academic Child Health, University of Bristol, 1-5 Whiteladies Road, Bristol BS8 1NU, UK*

[e]*Population Health Sciences, Bristol Medical School, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol, UK, BS8 2PS*

Accepted 26 June 2019; Published online 2 July 2019

Abstract

Objectives: Core outcome set (COS) development often begins with a systematic review to identify outcomes. Reviews frequently show heterogeneity in numbers of outcomes reported across trials. Contributing to this is a lack of a uniform definition for an outcome. This study proposes a first working definition for a unique trial outcome to support reporting a quantitative assessment of outcome reporting heterogeneity (ORH).

Study Design and Setting: Eligible COS literature (development papers, protocols, and reviews) were identified using the COMET database, Ovid MEDLINE, and PubMed. Outcome numbers, definitions, timing, and grouping methodology were examined.

Results: One hundred and thirty two studies were included. 82 (88.1%) studies (excluding protocols) reported a total number of unique outcomes (median, 82; range, 12-5776; IQR, 261). Timing of assessment was reported in 32 (31.4%) studies. Methods to group similar outcomes were reported in 8 (7.8%) articles. No study defined how outcomes were agreed as different and how final numbers of unique outcomes were determined. It is proposed that a unique outcome requires original meaning and context. Thus ORH is suggested to be the reporting of multiple unique outcomes across trials related to one health care condition.

Conclusion: This review identified inconsistencies in how authors define, extract, group, and count trial outcomes. Further work is needed to refine our proposed definitions to optimize COS development and allow a quantifiable measure of ORH.

Keywords: Outcomes; Research methodology; Core outcome set; Clinical trials; Systematic review; COMET

1. Introduction

Well-designed and -conducted randomized controlled trials (RCTs) determine effectiveness through an unbiased comparison of outcomes (events or endpoints) by the intervention group [1]. The choice and selection of outcomes is critical in the RCT design, and trials may be regarded to be ''only as credible as their outcomes'' [2,3]. Trial outcomes, however, may be defined in various ways, including [4–10] ''as variables that are monitored during a study to document the impact that a given intervention or exposure has on the health of a given population'' [11] and ''as a variable measured at a specific time point to assess the efficacy or harm of an intervention'' [12]. While these definitions

Conflict of interest: All authors agree that there are no financial/personal interest or belief that could affect their objectivity in the publication of this paper.

Funding: This work was supported by a National Institute for Health Research (NIHR), United Kingdom Doctoral Research Fellowship DRF-2016-09-031. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. This work was undertaken with the support of the Medical Research Council ConDuCT-II Hub (Collaboration and innovation for Difficult and Complex randomised controlled Trials In Invasive procedures—MR/K025643/1) and Royal College of Surgeons of England Bristol Surgical Trials Centre. KNLA and JMB were supported by the NIHR Biomedical Research Centre (BRC) at the University Hospitals Bristol NHS Foundation Trust and the University of Bristol. JMB holds an NIHR senior Investigator award.

* Corresponding author. Tel.: 0117 3427910.
E-mail address: amber.young1@nhs.net (A.E. Young).

https://doi.org/10.1016/j.jclinepi.2019.06.016

What is new?

Key findings

- Proposed definition of a unique outcome: "a trial outcome is one that has original meaning and context. Outcomes with different words, phrasing, or spelling addressing the same concept and context should be categorized as one outcome."
- Proposed definition of outcome reporting heterogeneity (ORH): "the reporting of multiple unique outcomes across trials within one health care condition" relates directly to and is dependent on the definition of a unique outcome.

What this adds to what was known?

- This study analyzes current methods for identifying, extracting, defining, grouping, and counting outcomes into unique outcomes to inform core outcome set (COS) development and allow the quantitative reporting of ORH.

What is the implication and what should change now?

- COS developers will ultimately have an objective basis for accurately counting unique outcomes across health care trials. This will allow the reproducible quantitative measurement of ORH, inform trial design, and ultimately inform evidence synthesis and reduce research waste.
- Future collaborative work will be undertaken to agree and to explore the validity of the final definitions.

explain the role of an outcome in a trial, they do not define what an outcome is per se. Other authors have examined the completeness of outcome reporting using four or five levels. These include outcome, measurement, metric, aggregation of outcomes, and timing of measurement [13,14]. Other work by Page and Dwan discusses the completeness of outcome reporting. None of these articles describe how to determine what makes an outcome unique [12,15]. Without a definition for an individual, unique outcome, difficulties arise in differentiating one outcome from another. This results in the reporting of multiple, apparently different outcomes across studies and impacts negatively on evidence synthesis thereby contributing to research waste.

The lack of definition of a unique outcome may also complicate the development of a core outcome set (COS). A COS is a minimum set of outcomes that are selected, measured, and reported in trials of a specific condition [16]. These are typically developed by identifying all outcomes in the literature and combining these into groups/domains for stakeholders to prioritize using a consensus process. The first part of this process requires that researchers scrutinize the literature and extract outcomes verbatim for reasons of transparency [8]. Outcomes are then deduplicated and grouped into individual "unique" outcomes for the prioritization process. The process may identify between 20 and more than 1,000 outcomes [17–20]. This wide variation is likely to reflect the lack of application of a uniform definition of a unique outcome, in addition to real differences in numbers of different outcomes. Guidance as to when the use of different wording defines the same outcome and when it does not is lacking; for example, hospital survival and in-hospital mortality [21,22]. There is no advice as to how to group similar outcome terms into a single unique outcome. For example, some researchers may choose to include all definitions of a specific outcome under one term, with others seeking more granularity and reporting several different definitions as unique outcomes [18,23]. There is no clarity as to which process is correct for demonstrating variation in outcome reporting and for developing an outcome long list for a COS. It is also unclear whether the timing of outcome measurement affects the singularity of an outcome; for example, some researchers will count the incidence of wound healing at two specified time points as two outcomes and others as one [24,25]. Furthermore, reporting of methods for extracting and grouping outcomes is often poor. The challenges with identifying and determining what is a unique outcome means that different researchers may extract a different number of outcomes from the same data set, and the scale or presence of true variation in outcome reporting will therefore be difficult to establish. Inconsistency in methods for extracting outcomes means that the presence and scale of outcome reporting heterogeneity (ORH) is difficult to establish. ORH is a quantitative measure of the variation in outcomes reported across trials in one health care area. It is commonly reported in COS development as a number, that is, n = x different outcomes reported across n = x trials in a specific health care area. ORH will impact on the validity of the long list of outcomes used to inform the consensus processes in the development of COSs.

The aim of this article is to examine methods used to extract, and combine outcomes with the same meaning, from published research articles to inform how to establish a reproducible and measurable long list of unique outcomes. This paper suggests a first working definition of a unique outcome. A final agreed definition will ensure that COS researchers can accurately and consistently identify a quantitative assessment of the inconsistency of outcome reporting (termed here, "outcome reported heterogeneity"). We see the enclosed work as a starting point for debate between international COS researchers, with further validation before the definitions and methodology being finalized.

2. Methods

The study consisted of two phases. Phase 1 was an in-depth literature review undertaken to analyze and summarize methods for outcome extraction, grouping, defining, and counting from systematic reviews used to inform COS development. Phase 2: Findings from phase 1 combined with multidisciplinary expert opinion informed the development of a first definition of a unique outcome and methodology for the conversion of outcomes extracted verbatim from trials into unique outcomes. A first definition of ORH was developed based on the definition of a unique outcome.

2.1. Phase one literature review

2.1.1. Data source and search strategy

A structured search of the COMET (Core Outcome Measures in Effectiveness Trials) database (a repository of COS studies) was undertaken by two authors (A.E.Y. and A.D.). As optimal COS development methodology is still evolving, it was hypothesized that the most current advances in methodology are likely to be found in recent studies registered with the COMET database. This database allows the use of filters to identify the nature of the COS work (e.g., protocol, full paper, population studies, and so forth). The filters applied in this study are described in Appendix A).

2.1.2. Study selection criteria and identification

Titles from the COMET database were used to identify COS development and related published articles using Ovid MEDLINE and PubMed [26]. Papers were included if they met the inclusion criteria below and reported any details of outcome extraction, grouping, and counting from literature reviews to inform the presence of ORH and the development of a COS outcome long list.

Inclusion criteria

- A primary COS development study published between January 1, 2015 and August 20, 2018;
- A COS protocol (without an associated final COS published with the above time period); or
- A previously published literature review that was referenced in, informed, or directly related to a primary COS published within the aforementioned time period.

There was no restriction on the type of studies in terms of patient characteristics or disease area.

2.1.3. Identification of studies

Full text articles were retrieved and reviewed to determine eligibility, independently and in duplicate by two authors. Reasons for exclusion were ordered hierarchically in order of importance (Fig. 1) and applied to each full text. Discrepancies were resolved by discussion and consulting with coauthors. Reference lists of full texts were hand-searched to identify a comprehensive list of all literature reviews to support COS development.

2.1.4. Data extraction and synthesis

Two authors extracted data using a form developed and piloted by the study team. Data extracted were as follows:

1 Type of study: primary COS study, COS protocol, literature review to support a COS;
2 Methodology for extraction and grouping: whether outcomes were extracted verbatim and methodology for grouping similar outcomes into unique outcomes.
3 Outcome details: total number of unique outcomes reported (quantitative measure of ORH), presence of different wording for the same outcomes, and impact of timing of outcome measurement on numbers of unique outcomes

Results were compared between researchers, with any disagreements resolved by a senior COS researcher (J.B.). Primary COS development studies and their respective literature review(s) that directly informed the development of that COS were ‘‘paired’’ for data extraction, so that data were not duplicated.

2.1.5. Data analysis

Numerical data are presented as summary statistics. A narrative synthesis was applied to methods for extracting and grouping outcomes from trials and for managing the timing of outcome assessment [27–29]. Heterogeneity and similarity of outcome extraction methodology was noted, exploring relationships between studies.

2.2. Phase 2: development of a definition of a unique outcome and methodology for grouping verbatim outcomes into unique outcomes

A summary of the findings from phase 1 of the study were presented to and discussed by a single-center multidisciplinary group of senior researchers experienced in COS research (A.E.Y., A.D., J.M.B., and C.M.; Bristol Centre for Surgical Research). A first working definition of a unique trial outcome and methods for conversion of outcomes extracted verbatim into unique outcomes were developed, and iteratively refined through further discussion and review of the data from Phase 1. A first definition for ORH was subsequently proposed, based on the ability to define a unique outcome. These definitions are a first attempt to define a unique outcome and to use this to define ORH. They will be finalized after formal collaboration with COS developers and the COMET Initiative before validation.

The study flow chart is illustrated in Fig. 1.

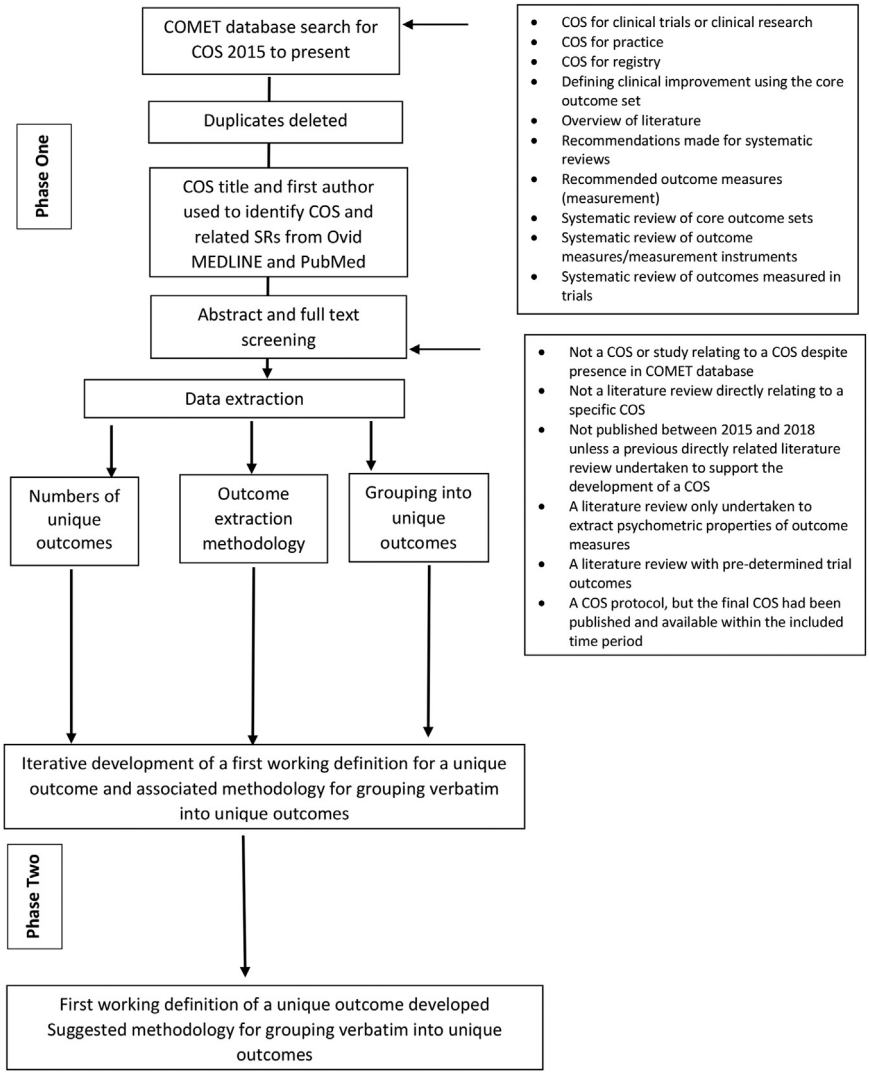

Fig. 1. Study methodology. COMET, Core Outcome Measures in Effectiveness Trials; COS, core outcome set; SRs, systematic review.

3. Results

3.1. Phase 1: literature review

The COMET database search yielded 121 titles (seven duplicates removed) for COS studies. These 121 articles were identified in OVID Medline and PubMed, and full texts were then extracted. Hand-searching identified a further 111 related literature reviews. This led to a total of 232 studies (Fig. 2). Of these, 100 articles did not reach the inclusion criteria, leaving 132 studies for data extraction.

1 Type of study: Of the 132 included studies (Appendix B), 43 (32.6%) were a final COS, 80 (60.6%) were a literature review undertaken to support a COS, and nine (6.8%) were COS protocols with details of a literature review, where the final COS was not yet published. Of the final COSs, 30 (69.8%) paired directly with a previously published literature review, leaving 13 COSs that were analyzed alone. The results described in the following are therefore taken from 102 (132-30) different COS studies. Final numbers of unique outcomes were taken from 93 studies. The 9 protocols were excluded as they would not report extracted outcome numbers, as detailed in Fig. 2.

2 Outcome details (Table 1): Thirty-two studies (31.4%) discussed the issue of the timing of outcome assessment. In 17 (53.2%), these were counted as unique outcomes (e.g., wound infection at 30 days was reported as a different outcome to wound infection at 90 days) based on time alone, whereas the remainder counted these as just one outcome. Of the 102 studies (with the nine protocol studies excluded, i.e., $n = 93$), 82 (88.1%) reported a quantitative assessment of the number of outcomes reported across the included trials (i.e., a quantitative assessment of ORH). The total number of unique outcomes reported varied from 12

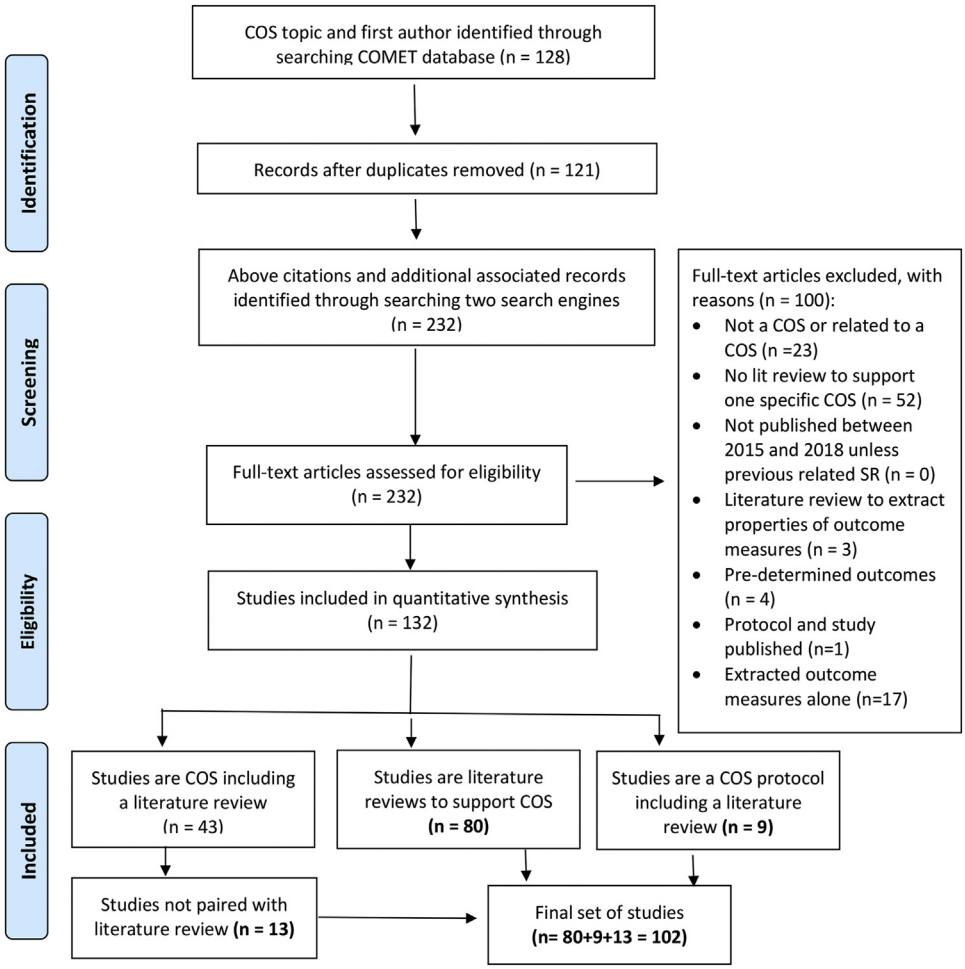

Fig. 2. PRISMA 2009 flow diagram. COMET, Core Outcome Measures in Effectiveness Trials; COS, core outcome set.

to 5,776 per review (median: 82; IQR: 261). Varying definitions for the same outcome across included trials were reported in 53 studies (52.0%).

3 Methodology for extraction and grouping (Table 2): 18 studies (17.6%) reported that extraction of outcomes was verbatim. Of these, only 18, 44% (7.8% of all 102 studies) included some text description of the authors' methodology for grouping these into unique outcomes (Table 2).

3.2. Phase 2: agreement on methodology for grouping verbatim outcomes into unique outcomes

Based on the aforementioned data, detailed discussions were held with a single-center multidisciplinary expert panel of senior COS researchers regarding the definition of how a unique outcome should be defined and how this would impact on the quantitative measure of ORH (Fig. 3). After iterative refinement, the following definition

Table 1. Outcome details (paired study data collated)

Numbers of studies reporting:	Number of studies (%)
Final numbers of unique outcomes reported across trials within systematic review (excluding protocols, $n = 9$)	82/93 (88.1)
Researchers state that outcomes were extracted verbatim from trials within systematic review	18/102 (17.6)
Researchers report different definitions for the same outcome across trials within systematic review	53/102 (52.0)
Researchers report the timing of outcome assessment	32/102 (31.4)
Researchers report that the timing of assessment impacts on number of outcomes reported	17/32 (53.1)
Methodology reported for grouping outcomes into unique outcomes	8/18 (44.4)

Table 2. Methodology for identifying outcomes as unique

COS review title	Conversion of verbatim outcomes into unique outcomes (text extracted)
A systematic review of outcomes in postoperative pain studies in paediatric and adolescent patients: towards development of a core outcome set [30].	''Outcomes were abstracted based on group consensus. We defined an outcome as the exact word-for-word terms … for any clinical end point, or physiological, metabolic, or mortality event measured by clinicians or researchers. Final outcomes were then standardised to improve the consistency of naming. For example, 'objective pain score' was changed to 'pain measurement.' After the outcomes had been standardised, we placed them in broader domains.''
Evaluating physical activity in dementia: a systematic review of outcomes to inform the development of a core outcome set [31].	''From verbatim outcomes to outcome domains: One author, … grouped verbatim outcomes with the same semantic meaning into outcome domains. For instance, the verbatim outcomes 'Functional independence,' 'Ability to develop basic activities of daily living,' and 'Functional performance' were grouped into the outcome domain 'Functional abilities and independence.'''
Outcomes mapping study for childhood vaccination communication: too few concepts were measured in too many ways [32].	''For each outcome mentioned in an included trial, we extracted into a spreadsheet all information defining the outcome, such as type…, outcome variables …, age of the subjects, and any other related details. We used the exact words of the trial authors. We did not extract data related to the timing and scale or tool used to measure the outcomes, as examination of how specific outcome variables were measured was not the subject of the research. Two researchers reviewed the extracted data. One author coded the individual outcomes according to what these measured, using the language of the trialists. These codes were discussed and confirmed. This first round of codes became the most specific level of the taxonomy. We retained a relatively large number of different groups rather than aggregating the information and potentially losing important details.''
Variability of outcome reporting in Hirschsprung's disease and gastroschisis: a systematic review [23]	''In the 35 included studies, 95 outcomes were investigated a total of 337 times. Thirty five outcomes were considered to be too similar to at least one other outcome to be meaningfully differentiated, and these outcomes were therefore mapped to one common term (e.g., continence/incontinence or frequency of stool/bowel movement frequency). Following this exercise, 74 unique outcomes were identified as having been reported. Within the included studies, 102 outcomes were investigated a total of 247 times. Within these 102 outcomes, there were 63 that were felt to be too similar to at least one other outcome to be meaningfully differentiated, and these were therefore mapped to one common term. Following this mapping process, there remained 62 unique outcomes.''
No common denominator: a review of outcome measures in IVF RCTs [33]	''We also did not record outcomes multiple times where these corresponded to repeated measurements at several time points. For each reported outcome, we extracted the numerator and denominator (for numerical variables, the denominator would be the divisor used in the calculation of a mean). Where pregnancy or live birth were reported, we extracted the corresponding definition used by the study authors. Data were extracted into two databases, one containing study-level information and another containing reported-outcome-level information. Due to the large number of outcomes identified, we reported only those appearing in more than one study. We simplified the results by combining similar numerators and denominators. For example, we combined live birth with take home baby rate and combined the denominators 'per patient with sufficient embryos' and 'per patient with sufficient blastocysts,' where 'sufficiency' could be defined on the basis of quantity or quality of embryos (or both). For this primary analysis, we did not distinguish between subtly different definitions of outcomes (e.g., clinical pregnancy may have been defined as fetal heartbeat on ultrasound at different time points in different studies). However, at the suggestion of an anonymous peer reviewer, we also present the definitions used by trial authors for pregnancy and live birth outcomes.''
Systematic review of outcome measures following chemoradiotherapy for the treatment of anal cancer (CORMAC) [34].	''Verbatim outcomes were initially reviewed by a single researcher and assigned a standardized name ('standardized outcome term') to overcome variations in wording used for the same outcome. The standardized outcome term and domain assigned to each verbatim outcome were reviewed and agreed at a meeting of the CORMAC Study Advisory Group (SAG), composed of experts in the field of anal cancer, including clinical oncologists, a colorectal surgeon, radiologist, clinical trials methodology expert, qualitative research expert, and a patient partner. There were 533 unique terms collapsed into 86 'standardized outcome terms,' representing outcomes with the same meaning but with differing wording and assigned to the appropriate outcome domain.''

(*Continued*)

Table 2. Continued

COS review title	Conversion of verbatim outcomes into unique outcomes (text extracted)
Outcome reporting in randomized controlled trials and systematic reviews of gastroschisis treatment: a systematic review [35]	''We anticipated some diversity in terminology used to report outcomes and therefore grouped similar outcomes. We identified outcomes that seemed similar or of a similar theme despite differing definitions used across studies and assigned an appropriate term to them. For instance, the outcomes 'proven catheter-related sepsis (line positive blood cultures necessitating antibiotic treatment or catheter removal)' and 'central line infections' were included in the term 'central venous catheter sepsis.'''
Developing a core outcome set for fistulising perianal Crohn's disease [36]	''Reported outcomes were extracted verbatim and listed in preparation for categorisation into domains.''

Abbreviation: RCT, randomized controlled trials.

of a unique outcome is suggested: ''a trial outcome is one that has original meaning and context.'' Outcomes with different words, phrasing, or spelling addressing the same concept and context should be categorized as one outcome. In other words, researchers should group together outcome synonyms into one unique outcome term. By the term ''original meaning,'' individual clinicians and patients would need to clearly understand what a particular outcome meant and how it was different to any other outcome. For example, the number of days in hospital has the same meaning as hospital length of stay. In terms of ''original context,'' we mean that researchers must be clear when defining the context of the outcome. For example, postoperative pain and neuropathic pain would be two different outcomes as they are different in context. The timing of outcome measurement should be clearly stated, but an outcome differing only in this aspect is not unique. Supporting evidence for this is in the fact that outcomes measured at different time points can be pooled in a meta-analysis [37]. The definition of ORH is proposed as ''the reporting of multiple unique outcomes across trials within one health care condition.'' This relates to the quantitative variation in outcomes reported across trials in one health care area and is dependent on the ability to define a unique outcome. Both definitions suggested here are purely suggestions for further discussion and validation.

4. Discussion

In this study, we aim to explore through examination of COS research, what makes one outcome different to any other outcome. Of all 102 COS studies included in this review, only eight (7.8%) reported any methodological detail about how verbatim outcomes were grouped into final unique outcomes. Despite this inability to report how the final list of different outcomes was determined, authors of 88.1% of studies still reported the final unique outcomes as a number, that is, they reported variation in outcome reporting across studies of one health care area (ORH) quantitatively.

The number of reported outcomes across studies in this review varied from 12 to 5,776 (median: 82; IQR: 261). It is unclear why there was such a wide variation in numbers of different outcomes extracted. We believe the answer lies in the authors' decisions regarding the granularity of outcomes extracted and the use of timing of measurement to define unique outcomes. Chong et al. in a systematic review on pediatric chronic kidney disease reported 5,776 different outcomes from 213 studies [38]. In one group of outcomes, 19 ''unique outcomes'' all relate to glomerular filtration rate and were measured at several time points—each of which was counted as a separate ''unique'' outcome. This resulted in 148 apparently unique outcomes in this one group. Another review used in the development of a COS on variability in the reporting of renal function endpoints in immunosuppression trials in renal transplantation reported only 345 outcomes in total (compared to 5,776) from 213 studies [39]. This review did not include outcomes measured at different times as unique. Otherwise, it is not possible to clarify this disparity as neither article reports how the number of unique outcomes were calculated.

The timing of outcome assessment was reported in just less than one-third of studies. Of these studies, more than half reported outcomes with different times of assessment as unique. Variation in wording across included trials for the same outcomes occurred in 53 (52%) of the studies (e.g., different definitions of wound healing: length of time to heal to 50% or to 70% healed, time to 90% re-epithelialization), suggesting that further grouping was necessary and that these outcomes were not unique. The lack of methodological detail and significant variation in numbers of outcomes reported across studies suggests a nonsystematic approach to outcome extraction and counting. This approach will not allow any true variation in numbers of outcomes reported across studies to be assessed accurately (quantitative measure of ORH).

Other authors have previously described, but not defined, this variation in outcome reporting or ORH [23,40,41]. They have also described how it leads to problems in evidence synthesis [23,42,43]. One reason for ORH is a lack of clarity regarding what constitutes a unique outcome, what makes one outcome different from another. Although COSs have been developed to resolve ORH, they require the extraction of outcomes verbatim from trials, and

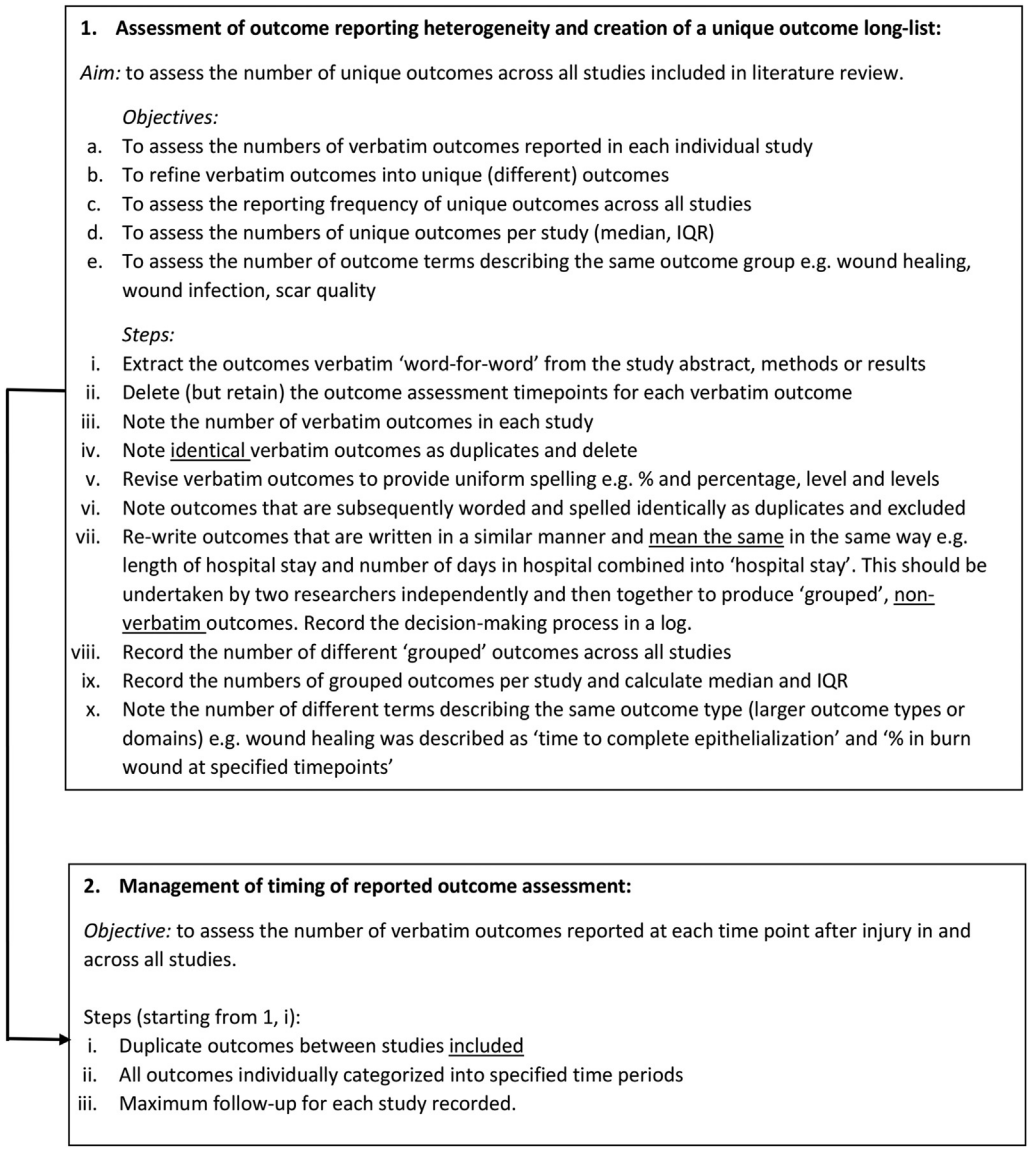

Fig. 3. Unique outcome methodology proposal.

the subsequent grouping of the same or similar outcomes into individual, unique outcomes. This process requires an understanding of what constitutes a unique outcome and which outcomes are so similar they can be combined into one term. It is important to agree the level of granularity required in outcome reporting and this may be partially responsible for the wide variation in the numbers of unique outcomes reported.

The COMET Initiative has published analyses of its COS database [44–47]. This has identified variability in aspects of COS development, namely the scope, stakeholder involvement, and consensus process [44,45,48,49]. COMET has therefore undertaken work to provide methodological guidance regarding these aspects of COS development [8,46–51]. These activities are extremely useful, although they have not focused as much on the early part of COS development. The early stages, including extraction of comprehensive lists of outcomes and identifying the magnitude of ORH, not only justify the need for a COS but directly feed into the sets themselves (Fig. 4).

The strengths of this study are that we critically and systematically analyzed 132 COS development studies, including 90 systematic reviews used to inform COSs. A limitation of the study is that we did not undertake a full systematic review but rather focused our attention on studies identified through the COMET database from September 2015, with related articles identified through two other search engines, as per the published advice on methodological systematic reviews [49,52]. To support this approach, Gargon et al. noted the comprehensive nature of the COMET database, in that 720 studies relevant to the development of COS had been included in the

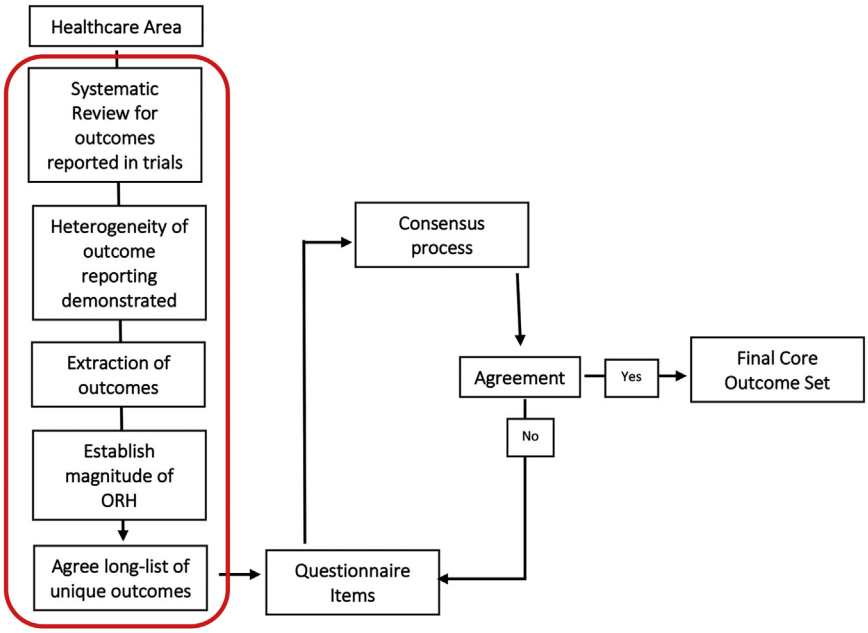

Fig. 4. Flow chart. ORH, outcome reporting heterogeneity.

COMET database at the end of December 2015 [46]. In addition, COSs registered on the COMET database are likely to be of high quality because of the COMET standards [8,46]. Finally, we further expanded our search by including directly related studies identified though two search engines and hand-searching references. We believe that it is unlikely that COS development studies reported before 2015 will describe methodology for the extraction of unique outcomes if later studies do not. In undertaking this work, we have not attempted to be comprehensive in exploring all COS development articles. We wanted to demonstrate that in a recent group of high-quality COS research collected in a prespecified manner, no one has determined objectively and with repeatability how to extract and count unique outcomes to determine a quantitative measure of ORH across articles in one health care area.

This study has shown that the process of extracting outcomes from trials included in COS literature reviews and grouping the verbatim outcomes into a list of outcomes that are individual and different (unique) from each other is complex and poorly reported. Verbatim outcome extraction is recommended by COMET for reasons of transparency [8,16]. "The first step is to group these different definitions together (extracting the wording description verbatim) under the same outcome name" [8]. Issues with this include determining when outcomes are the same even when different words are used; for example, serum albumin and albumin levels in plasma and how to incorporate the timing of outcome assessment into this process; for example, percentage wound healed at 2 weeks and percentage wound healed at 6 weeks. If authors could agree and transparently report how they extracted, grouped, and counted the outcomes reported, a true quantitative assessment of the true variation in outcome reporting (ORH) would be possible. A lack of understanding of what constitutes a unique outcome will impact on the validity of the reported presence and magnitude of ORH as demonstrated by the widely varying numbers of unique outcomes reported in the COS literature reviews included in this study. The other impact is that when outcomes presented in the later stages of COS development are not unique it makes prioritization difficult and can hinder the consensus process.

A final agreed definition for a unique outcome and methods for objectively grouping outcomes extracted verbatim into unique outcomes will provide a methodological basis for COS developers to define an outcome as unique and to determine an accurate, quantitative, and reproducible measure of ORH. The aim of this article is to start formal discussions between COS researchers and COMET to raise awareness of this methodological issue and to undertake collaborative work to refine and validate a definition for a unique outcome. The answer will allow researchers to determine a quantitative result for ORH, how many "unique" outcomes are present across trials in a specific health care issue.

CRediT authorship contribution statement

Amber E. Young: Conceptualization, Methodology, Data curation, Formal analysis, Writing - original draft, Writing - review & editing. **Sara T. Brookes:** Writing - review & editing. **Kerry Avery:** Writing - review & editing. **Anna Davies:** Writing - review & editing. **Chris Metcalfe:**

Writing - review & editing. **Jane M. Blazeby:** Writing - review & editing, Supervision.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2019.06.016.

References

[1] Ioannidis JP, Evans SJ, Gøtzsche PC, O'neill RT, Altman DG, Schulz K, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. Ann Intern Med 2004;141:781−8.

[2] Tugwell P. OMERACT conference on outcome measures in rheumatoid arthritis clinical trials: introduction. J Rheumatol 1993;20:528−30.

[3] Tugwell P, Boers M, Brooks P, Simon L, Strand V, Idzerda L. OMERACT: an international initiative to improve outcome measurement in rheumatology. Trials 2007;8(1):38.

[4] Sansoni JE. Health outcomes: an overview from an Australian perspective. Australian Health Outcomes Collaboration, Australian Health Services Research Institute, University of Wollongong, August. 2016. Available at https://ro.uow.edu.au/cgi/viewcontent.cgi?article=1823&context=ahsri.

[5] Frommer M, Rubin G, Lyle D. The NSW health outcomes program. N S W Public Health Bull 1992;3(12):135−7.

[6] Donabedian A. The role of outcomes in quality assessment and assurance. QRB Qual Rev Bull 1992;18(11):356−60.

[7] Hallas S, Nelson A, O'Meara S, Gethin G. Core outcome set for Venous leg ulceration" CoreVen": report from CoreVen meeting in Amsterdam 4 may 2017. EWMA J 2017;18(2):27−31.

[8] Williamson PR, Altman DG, Bagley H, Barnes KL, Blazeby JM, Brookes ST, et al. The COMET handbook: version 1.0. Trials 2017;18(3):280.

[9] The International Consortium for Health Outcomes Measurement (ICHOM). Available at https://www.ichom.org/.

[10] Webbe J, Brunton G, Ali S, Duffy JM, Modi N, Gale C. Developing, implementing and disseminating a core outcome set for neonatal medicine. BMJ Paediatrics Open 2017;1(1):e000048.

[11] Ferreira JC, Patino CM. Types of outcomes in clinical research. J Bras Pneumol 2017;43(1):5.

[12] Chan A-W, Krleža-Jerić K, Schmid I, Altman DG. Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. Can Med Assoc J 2004;171(7):735−40.

[13] Saldanha IJ, Dickersin K, Wang X, Li T. Outcomes in cochrane systematic reviews addressing four common eye conditions: an evaluation of completeness and comparability. PLoS One 2014;9:e109400.

[14] Zarin DA, Tse T, Williams RJ, Califf RM, Ide NC. The ClinicalTrials. gov results database—update and key issues. N Engl J Med 2011;364(9):852−60.

[15] Page M, McKenzie J, Dwan K, Kirkham J, Green S, Forbes A. Bias due to selective inclusion and reporting of outcomes and analyses in systematic reviews of randomised trials of healthcare interventions (Protocol). Cochrane Database Syst Rev 2012;5:MR000035.

[16] Williamson PR, Altman DG, Blazeby JM, Clarke M, Devane D, Gargon E, et al. Developing core outcome sets for clinical trials: issues to consider. Trials 2012;13(1):132.

[17] Deckert S, Kaiser U, Kopkow C, Trautmann F, Sabatowski R, Schmitt J. A systematic review of the outcomes reported in multimodal pain therapy for chronic pain. Eur J Pain 2016;20(1):51−63.

[18] Harman NL, James R, Wilding J, Williamson PR. SCORE-IT (Selecting Core Outcomes for Randomised Effectiveness trials in Type 2 diabetes): a systematic review of registered trials. Trials 2017;18(1):597.

[19] Andersen CR, Fitzgerald E, Delaney A, Finfer S. A systematic review of outcome measures Employed in Aneurysmal subarachnoid hemorrhage (aSAH) clinical research. Neurocrit Care 2019;30:534−41.

[20] Al Wattar BH, Placzek A, Troko J, Pirie AM, Khan KS, McCorry D, et al. Variation in the reporting of outcomes among pregnant women with epilepsy: a systematic review. Eur J Obstet Gynecol Reprod Biol 2015;195:193−9.

[21] Tokarik M, Sjoberg F, Balik M, Pafcuga I, Broz L. Fluid therapy LiDCO controlled trial-optimization of volume resuscitation of extensively burned patients through noninvasive continuous real-time hemodynamic monitoring LiDCO.[Erratum appears in J Burn Care Res. 2014 May-Jun;35(3):269]. J Burn Care Res 2013;34(5):537−42.

[22] Bechir M, Puhan MA, Fasshauer M, Schuepbach RA, Stocker R, Neff TA. Early fluid resuscitation with hydroxyethyl starch 130/0.4 (6%) in severe burn injury: a randomized, controlled, double-blind clinical trial. Crit Care 2013;17(6):R299.

[23] Allin BSR, Irvine A, Patni N, Knight M. Variability of outcome reporting in Hirschsprung's Disease and gastroschisis: a systematic review. Sci Rep 2016;6:38969.

[24] Shahzad MN, Ahmed N. Effectiveness of aloe vera gel compared with 1% silver sulphadiazine cream as burn wound dressing in second degree burns. J Pak Med Assoc 2013;63(2):225−30.

[25] Dallal MS, Safdari R, Koochak HE, Sharifi-Yazdi S, Akhoondinasab M, Pourmand M, et al. A comparison between occlusive and exposure dressing in the management of burn wound. Burns 2016;42(3):578−82.

[26] Gargon E, Williamson PR, Clarke M. Collating the knowledge base for core outcome set development: developing and appraising the search strategy for a systematic review. BMC Med Res Methodol 2015;15:26.

[27] Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, et al. Guidance on the conduct of narrative synthesis in systematic reviews. In: A Product from the ESRC Methods Programme Version 1: b92.

[28] Noyes J, Booth A, Flemming K, Garside R, Harden A, Lewin S, et al. Cochrane qualitative and implementation methods group guidance series—paper 3: methods for assessing methodological limitations, data extraction and synthesis, and confidence in synthesized qualitative findings. J Clin Epidemiol 2018;97:49−58.

[29] Lewin S, Bohren M, Rashidian A, Munthe-Kaas H, Glenton C, Colvin CJ, et al. Applying GRADE-CERQual to qualitative evidence synthesis findings—paper 2: how to make an overall CERQual assessment of confidence and create a summary of qualitative findings table. Implement Sci 2018;13:10.

[30] Ross A, Young J, Hedin R, Aran G, Demand A, Stafford A, et al. A systematic review of outcomes in postoperative pain studies in paediatric and adolescent patients: towards development of a core outcome set. Anaesthesia 2018;73(3):375−83.

[31] Gonçalves A-C, Cruz J, Marques A, Demain S, Samuel D. Evaluating physical activity in dementia: a systematic review of outcomes to inform the development of a core outcome set. Age Ageing 2017;47:34−41.

[32] Kaufman J, Ryan R, Bosch-Capblanch X, Cartier Y, Cliff J, Glenton C, et al. Outcomes mapping study for childhood vaccination communication: too few concepts were measured in too many ways. J Clin Epidemiol 2016;72:33−44.

[33] Wilkinson J, Roberts SA, Showell M, Brison DR, Vail A. No common denominator: a review of outcome measures in IVF RCTs. Hum Reprod 2016;31:2714−22.

[34] Fish R, Sanders C, Ryan N, Van der Veer S, Renehan AG, Williamson PR. Systematic review of outcome measures following chemo-radiotherapy for the treatment of anal cancer (CORMAC). Colorectal Dis 2018;20:371−82.

[35] Ross AR, Hall NJ. Outcome reporting in randomized controlled trials and systematic reviews of gastroschisis treatment: a systematic review. J Pediatr Surg 2016;51(8):1385−9.

[36] Sahnan K, Tozer PJ, Adegbola SO, Lee MJ, Heywood N, McNair AG, et al. Developing a core outcome set for fistulising perianal Crohn's disease. Gut 2018;68:226−38.
[37] Carroll R, Metcalfe C, Gunnell D. Hospital presenting self-harm and risk of fatal and non-fatal repetition: systematic review and meta-analysis. PLoS One 2014;9:e89944.
[38] Chong LS, Sautenet B, Tong A, Hanson CS, Samuel S, Zappitelli M, et al. Range and heterogeneity of outcomes in randomized trials of pediatric chronic kidney disease. J Pediatr 2017;186:110−117.e11.
[39] Knight SR, Hussain S. Variability in the reporting of renal function endpoints in immunosuppression trials in renal transplantation: time for consensus? Clin Transplant 2016;30(12):1584−90.
[40] Ma C, Hussein IM, Al-Abbar YJ, Panaccione R, Fedorak RN, Parker CE, et al. Heterogeneity in definitions of efficacy and safety endpoints for clinical trials of Crohn's disease: a systematic review for development of a core outcome set. Clin Gastroenterol Hepatol 2018;16:1407−19.
[41] Denniston AK, Holland GN, Kidess A, Nussenblatt RB, Okada AA, Rosenbaum JT, et al. Heterogeneity of primary outcome measures used in clinical trials of treatments for intermediate, posterior, and panuveitis. Orphanet J Rare Dis 2015;10(1):97.
[42] Agha RA, Fowler AJ, Pidgeon TE, Wellstead G, Orgill DP. The need for core outcome reporting in autologous fat grafting for breast reconstruction. Ann Plast Surg 2016;77(5):506−12.
[43] Hirsch M, Duffy JM, Kusznir JO, Davis CJ, Plana MN, Khan KS, et al. Variation in outcome reporting in endometriosis trials: a systematic review. Am J Obstet Gynecol 2016;214:452−64.
[44] Gorst SL, Gargon E, Clarke M, Blazeby JM, Altman DG, Williamson PR. Choosing important health outcomes for comparative effectiveness research: an updated review and user survey. PLoS One 2016;11:e0146444.
[45] Gargon E, Gurung B, Medley N, Altman DG, Blazeby JM, Clarke M, et al. Choosing important health outcomes for comparative effectiveness research: a systematic review. PLoS One 2014;9:e99111.
[46] Gargon E, Williamson PR, Altman DG, Blazeby JM, Tunis S, Clarke M. The COMET initiative database: progress and activities update (2015). Trials 2017;18(1):54.
[47] Gargon E, Williamson PR, Altman DG, Blazeby JM, Clarke M. The COMET Initiative database: progress and activities from 2011 to 2013. Trials 2014;15(1):279.
[48] Kirkham JJ, Davis K, Altman DG, Blazeby JM, Clarke M, Tunis S, et al. Core outcome set-standards for development: the COS-STAD recommendations. PLoS Med 2017;14(11):e1002447.
[49] The COMET initiative. Available at http://www.comet-initiative.org/.
[50] Kirkham JJ, Gorst S, Altman DG, Blazeby J, Clarke M, Devane D, et al. COS-STAR: a reporting guideline for studies developing core outcome sets (protocol). Trials 2015;16(1):373.
[51] Kirkham JJ, Gorst S, Altman DG, Blazeby JM, Clarke M, Devane D, et al. Core outcome set−STAndards for reporting: the COS-STAR statement. PLoS Med 2016;13(10):e1002148.
[52] Lilford RJ, Richardson A, Stevens A, Fitzpatrick R, Edwards S, Rock F, et al. Issues in methodological research: perspectives from researchers and commissioners. Health Technol Assess 2001;5(8):1−57.