Controversy and Debate Series on Core Outcome Sets. Paper 1: Improving the generalizability and credibility of core outcome sets (COS) by a large and international participation of diverse stakeholders

Astrid Chevance, Viet-Thi Tran, Philippe Ravaud

Version of Record: https://www.sciencedirect.com/science/article/pii/S0895435619302434
Manuscript_e7affda9ca287af2573af04fa8be03ea

Improving the generalizability and credibility of core outcome sets (COSs) by a large and international participation of diverse stakeholders

Astrid Chevance [1], Viet-Thi Tran [1], Philippe Ravaud [1, 2]

[1] Université de Paris, CRESS, INSERM, INRA, F-75004 Paris, France.
[2] Department of Epidemiology, Columbia University Mailman School of Public Health, 22 W 168th St, New York, NY, USA

Address and positions for each author:

Astrid Chevance: PhD Student in Epidemiology, astrid.chevance@clinicalepidemio.fr
Viet Thi Tran: Researcher thitranviet@gmail.com
Philippe Ravaud: Director of the CRESS philippe.ravaud@aphp.fr

Correspondence to:

Astrid Chevance, MD, PhD Student
Hôpital Hôtel Dieu, Centre d'Épidémiologie Clinique, Paris, France
1 place du Parvis Notre-Dame, Paris 75181, France
Tel: +33 1 42 34 89 87, fax: +33 1 42 34 87 90, email: astrid.chevance@clinicalepidemio.fr

Contributors and sources: AC is currently working towards a PhD on creating new methods for COS development. VTT is a researcher in Public Health and co-director of the PhD. He has expertise in online surveys. PR is the director of the PhD and has expertise in outcomes research. This paper introduces the methodological innovation for COS development.

Generated the idea: AC, VTT, and PR. Conceived and designed the methods: AC, VTT and PR. Wrote the first draft of the manuscript: AC. Contributed to the writing of the manuscript: AC, VTT, and PR. ICMJE criteria for authorship read and met: AC, VTT, and PR. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Declaration of interest:

AC received a PhD grant for the submitted work from the "Fondation pour la Recherche Médicale" (no. FMD20170637634).
VTT, PR reported no conflict of interest. https://www.frm.org/en

Word count: 3367

ABSTRACT: 200 words

Core outcome sets (COSs) are agreed-upon standardized sets of outcomes that should be minimally measured in all trials of a given disease. To enhance the uptake of COSs in trials, their generalizability should be increased by a change in the number and diversity of stakeholders involved in their development. Also, their credibility should be improved by methods that allow the fair capture of participants' views. This article proposes three adjustments to the development of COSs. First, instead of a qualitative study with few participants, we propose to generate the outcome domains by mapping the expectations toward treatment of a large number of stakeholders, internationally, by using an online survey with open-ended questions. Second, we propose to separate preference elicitation from the decision-making process in the selection of core outcomes. Preference elicitation would rely on an international online ranking survey, whereas the decision-making process would involve a formalized discussion among all stakeholders. Third, we propose to involve a large number of participants, including patients, in an online survey to select outcome measurement instruments. Our propositions are low-cost and scalable and help with the involvement of more stakeholders in the development of COSs, thereby increasing their generalizability, credibility and uptake in trials.

Keywords: Core Outcome Set, Patient and Public Involvement, meta-analysis, clinical trials, outcome, online survey.

What is new?

- We propose a set of methods to involve a large number of diverse stakeholders, worldwide, in the generation and selection of outcome domains to be included in core outcome sets (COSs).
- These methods should improve the generalizability and the credibility of COSs and thus their uptake in trials.
- Use of an international online survey of participants' expectations toward treatments could be more adequate than focus groups or face-to-face interviews for mapping all potential outcome domains.
- Use of an international online ranking survey to inform the selection of outcomes allows for the capture and preservation of the preferences of a large number of participants. Results from the ranking will inform the final decision-making process involving all relevant stakeholders.
- Use of online surveys to involve a large number of participants (including patients) in the instrument selection to assess their clarity and acceptability may improve the generalizability of this step.

Introduction: Generalizability and credibility of COS

Core outcome sets (COSs) are agreed-upon standardized sets of outcomes that should be minimally measured in all trials for a specific clinical area [1]. They consist of both a Core Domain Set (What domains should be measured in trials?) and a Core Outcome Measurement set (What instruments should be used to measure these domains?)[2]. COS uptake in trials contributes to reducing research waste by limiting the selective reporting of outcomes and ensuring that results can be compared and combined [3–5]. For example, meta-analyses of rheumatoid arthritis can include 87.1% of trials and 92.8% of patients because the Outcome Measures in Rheumatology (OMERACT) initiative promoted standardized outcome measures in rheumatology [6,7]. In contrast, in nephrology, for which there was no COS until 2014, meta-analyses could pool only 83.1% of trials and 44.7% of patients [8]. COS uptake in trials is the next challenge for COS research [9].

By definition, a COS is a consensus on what outcomes should be measured in trials, worldwide, for a given disease [1,2]. Therefore, the uptake of a COS in trials depends on both acknowledgement as a credible consensus and its generalizability [9].

- A consensus is credible when it uses methods that guarantee the fair participation of all stakeholders. Every stakeholder should have the same possibility of, first, expressing his/her perspective and being understood without any distortion and second, influencing the final decision. For instance, group decision-making is influenced by social and psychological determinants [10].
- A consensus is generalizable if people who did not participate in the consensus process could agree with the decision because they recognize the value of a collective decision.

To enhance the generalizability of COSs, there is a need to involve a large number of relevant stakeholders (e.g., patients, carers, clinicians, trialists, industry, and policymakers) internationally in their development. To enhance the credibility of COSs, there is a need to use methods that ensure the fair capture and preservation of the perspectives of all stakeholders [9,11]. The involvement of a large number of stakeholders from different countries and cultures in COS development has been limited [6]. Despite recent major progress, only 20% of published COSs included patients in their development; 57% of them included patients from a single country [12].

This paper proposes a reflection on COS development and consequential adjustments or alternatives to current methods of COS development [9, 11].

1. Areas for improvement in the current COS development process

The COS development process aims at identifying what should be measured (outcome domains) and how it should be measured (outcome measurement instrument). Several initiatives (OMERACT, COMET, SONG, GRAPPA, etc.) propose standardized methods for developing COSs. We chose to mainly discuss COMET and OMERACT guidance because the first has an oversight position on COS development and the second has pioneered COS research [1,2,6,7,13–16]. The common steps to develop COSs are 1) a systematic review to identify the outcome domains measured in a given area [1]; 2) the selection of the most relevant outcome domains among those identified in the review during a Delphi method involving patients and the public, clinicians, researchers, industry representatives and policy makers [1,14,17]; 3) a face-to-face meeting with a few stakeholders to consolidate the COS; and 4) guidelines for selecting measurement instruments to measure the outcome domains [7,18]. Of note, timing of the outcome assessment and outcome metrics are not

systematically investigated during the COS development because they are considered trial issues [1]. Several researchers have pointed out the limitations of methods used to develop COSs [9,17]

1.1 Limiting the risk of overlooked relevant outcome domains

By restricting outcome domain generation to a systematic review of the literature, the risk is to capture outcomes that are relevant to only trialists and researchers [1]. Involving patients in identifying outcome domains can reveal different themes that are important to them. For example, outcome domains used in trials for distal radius fracture treatment usually involved "radiographic healing", "grip" and "range of motion", but patients' important outcomes were "function", "time to recovery" and "pain" [19]. To uncover potential outcome domains overlooked by the systematic review, OMERACT and COMET recommend the use of qualitative surveys or focus groups with at least 30 participants each among patients, clinicians and other stakeholders from at least 3 continents [1,6]. Despite huge progress in including this qualitative step, among the 259 COS studies published up to 2017, only 11 (4.2%) supplemented the systematic review with a qualitative method (e.g., interviews, focus groups) [12] (unpublished data). COS developers explained the rare use of qualitative methods to generate outcome domains as being due to feasibility issues, especially for studies at an international level.

In practice, the perspectives of crucial stakeholders such as patients are currently rarely captured at an early step in COS development, even if the guidance recommends doing so. In other words, during the selection of outcomes, patients are asked to select "core outcomes" among a pool of outcomes they did not help to generate. They may be asked to choose

between outcomes A and B without the possibility of proposing outcome C. This situation could be problematic if outcomes A and B are not relevant to them but C is.

1.2 Avoiding the risk of a "token participation" of patients and the public

Despite guidance for patient and public involvement in COS development, we raise the possibility of "token participation" [1,13]. We distinguish quantitative tokenism, whereby the proportions of each group of participants are imbalanced, from qualitative tokenism, with stakeholders recruited without having fair conditions to express their voice or make it count.

For risk of quantitative tokenism, one must take into account the proportion of each group of participants to that of other stakeholders, particularly the proportion of patients to clinical experts and researchers. If the proportion of patients is low when selecting the core outcomes, their collective view may not be preserved. The combined results of the 3 systematic reviews of published COSs noted that only 14 studies involved participants in a Delphi method (with an average proportion of 31% of public participants, among them patients) [12,20–22].

Regarding the risk of qualitative tokenism, recruiting patient and public participants is not sufficient to guarantee that their views are adequately considered in the COS. How their perspectives are captured and preserved throughout the complete development of the COS is critical. Especially, sociological and psychological determinants of group decision-making can affect the results of the consensus. For instance, a common pitfall is the risk of capturing only dominant voices [23]. Power hierarchies and dynamics due to differences in social, gender, ethnic, economic or political status among participants may jeopardize the consensus process [23,24]. Other pitfalls involve "groupthink" (i.e., the group decision is the

product of the effort of each individual to conform his/her opinion to what he/she believes is the consensus), competing goals across participants because of social status, and motivation to participate in the task or "social loafing" [10]. Moreover, asymmetries in the access to information affect the ability to express an opinion. In COS development, COS literacy has been found a barrier to patient and public participation [9]: although trialists and clinicians are familiar with the topic, patients are not: a recent survey of COS developers highlighted that patients often struggled to understand COSs, their lack of COS literacy becoming a barrier for their meaningful participation. [9]

1.3 Choosing measurement instruments meaningful for patients

Recently, the OMERACT Filter 2.1 pointed out the importance of involving patients in the choice of measurement instrument [7]. Involving patients in choosing relevant outcome measurement instruments is not systematically done. For instance, an instrument could measure an important domain for patients but with questions or items that may be unintelligible to them or that only partially express their views. This situation could lead to the use of irrelevant tools and uncovered needs. In addition to the expertise required for evaluating the metrological properties of instruments, generally performed by researchers, patients need to be involved to determine whether the measurement instruments fit their perspectives.

1.4 Enhancing COS uptake and implementation in trials

COS uptake and implementation in trials is considered the "biggest challenge" by several COS developers. Both OMERACT and COMET recommend anticipating this issue from the beginning of the COS development, by involving regulation authorities [17]. Indeed, after the US Food and Drug Administration and the European Medicines Agency recommended the

use of the COS for rheumatoid arthritis in their guidelines, uptake in pharmacological studies increased from about 40% to 81% [25–28]. However, currently, methods to involve regulation authorities while preserving the perspectives of less powerful groups, such as patients, remains unclear.

2. Propositions for methodological adjustments/alternatives to the current COS development process to improve generalizability and credibility

We propose some adjustments and alternatives to the current methodology of COS development to improve their generalizability and credibility. Figure 1 is derived from the COMET Handbook figure and illustrates our methodological propositions [1].

2.1 Improving outcome domain generation by including all relevant stakeholders

The first step of a COS development is to gather a long list of potential outcome domains to be included in the COS. The recommended method involves combining a systematic review of outcomes used in the literature with focus groups or face-to-face interviews to identify overlooked outcome domains, which are in fact rarely performed.

Instead, we propose the use of an online survey with open-ended questions on stakeholders' expectations for the treatment of a given condition. Online surveys allow for recruiting a large number of stakeholders internationally. For instance, such a survey identified the different aspects of the burden of treatment of 1053 participants across 34 countries [29]. Moreover, the lack of interactions "in real life" allow for a reduction in social desirability bias and self-censoring [30]. Above all, online surveys with open-ended questions may be more adequate than focus groups and face-to-face interviews to map the diversity of outcome domains. Indeed, focus groups and face-to-face interviews are used to understand or

elaborate a theory of human experiences, which is not required to list potential outcome domains [31]. Also, going in-depth with a few individuals will not diversify their perspectives, and investigating only a small sample of participants may lead to overlooking some outcome domains and produce data not relevant for COS development.

To generate outcome domains, the online survey should:

1) Rely on a pilot study using classical qualitative methods with few participants to determine the wording and content of the open-ended questions to be asked in the online survey in order to capture the perspective of participants.
2) Ask participants about expectations for the treatment rather than outcomes that should be measured in trials. This formulation requires no literacy about COSs, outcomes or trials. Rather it focuses on participants' direct interest; for instance, what is difficult to live with in terms of the condition, what they want to see improved in daily living, and what they expect from a treatment.
3) Not use preexisting ontologies or taxonomies for analysis of the answers. Rather, we suggest performing a thematic analysis with themes revealed from the verbatim responses of participants, thus preserving their views [32].
4) Pursue the recruitment of participants until data saturation, which is critical when the objective is to avoid overlooking outcome domains. For online surveys, data saturation can be assessed with an objective method [33].

The results of the online survey should be a fine-grained map of outcome domains that use the wording of participants reflecting the diversity of their views. This map can be compared and combined with the outcomes identified during a systematic review of the outcomes used in the literature, as recommended by both OMERACT and COMET.

2.2 Capturing and preserving all stakeholders' view during outcome selection

To select domains to be included in the COS among those identified during the online survey and the systematic review, we propose an alternative to the Delphi process recommended by OMERACT and COMET. The Delphi technique "is a forecasting technique that elicits, refines and draws upon the collective opinion and expertise of a panel of experts" [34]. The Delphi method is used as consensus technique but has limitations [35,36]. Two points should be taken in account when using the Delphi method for COS development.

First, the Delphi method was not designed for mixing experts (e.g., researchers or clinicians) and non-experts (e.g., patients or the public) [34]. Classically, in the Delphi method, an expert is a professional who has remarkable skills/knowledge in a given topic in a given professional field. Some may consider that patients are "experts" of the lived-experience of the disease; however, the lived-experience of a disease is epistemologically not comparable to the professional knowledge of researchers or clinicians. With their experiences, patients can give a testimony or elicit preferences, whereas experts have opinions derived from their professional knowledge. In a Delphi method, participants are asked to make compromises about what they consider important across rounds. In the case of COS development, this is acceptable for health/research professionals but poses ethical questions for patients: if a patient says that the most important outcome for him/her is "quality of life", forcing the patient to change their perspective is questionable.

Second, using a Delphi method for decision-making is a conundrum because the definition of a consensus is the "opinion stability between the rounds" [36]. To account for this problem, we propose to separate the production of the scientific knowledge about the participant's preferences from the decision-making of what should be included in the COS.

Therefore, we propose a two-step process for selecting outcomes in the COS:

1) The production of the scientific knowledge about the participant's preferences involves using an online ranking survey inspired from the Q methodology. This survey will involve a large number of participants representing all relevant stakeholder groups (patients, clinicians etc.). Each participant selects the 10 outcome domains they find the most important and then orders these 10 outcomes in four ranks (1 as the most important and 4 as the least important). This method limits the cognitive burden of participants in that studies of human cognition showed that ranking more than 7±2 items is difficult [37]. Several rankings according to socio-demographic or health characteristics will be provided to refine the information and help in the final decision-making process.

2) Because the selection of the core outcomes is a decision-making process, it requires a formalized discussion between all relevant stakeholders. Patients, cares, clinicians, and researchers but also regulatory authorities and the trialists who will use the COS in trials should be involved in this discussion. During this meeting, all stakeholders will be able to speak and be heard, for instance with mediation techniques, which is crucial for the credibility of the process and the COS [6,38].

2.3 Involving patients in selecting outcome measurement instruments

The last step of a COS development aims at selecting potential measurement instruments for each outcome domain of the COS. Recent improvement in the methods underlined the importance of taking in account patients' views to assess the content validity, the feasibility and patient burden (e.g., length, completion time, cost, copyright) [7,18]. To evaluate

whether the instrument is a match for the target domain, similar to the OMERACT Filter 2.1, we propose to account for the perspective of "*people, particular those with the lived experience of the disease and domains, to see if the instrument captures the breadth and the depth of the experience*" [7]. For instance, the OMERACT Filter 2.1 proposes a survey template for patients that could be adapted in an online survey [39]. Here, also involving a large number of participants will allow for more generalizability.

At the end of this step, the final product of the COS development is a list of approximately 7±2 core outcome domains, as recommended by OMERACT, matched to corresponding measurement tools for which metric properties, feasibility and patient relevance are established by a large number of participants and thus more likely generalizable and credible [13,37].

Discussion

In this article, we propose some adjustments and alternatives to the current COS development process to enhance generalizability and credibility:

1) Instead of focus groups or face-to-face interviews for generating outcome domains, an international online survey involving a large number of stakeholders allows for mapping potential outcome domains generalizable across different contexts.

2) Regarding the selection of outcomes, a preference-ranking survey of a large number of participants will inform a formalized discussion between relevant stakeholders on what should be core. Separating the elicitation of preferences from the decision-making process reinforces the credibility of the COS.

The involvement of a large number of diverse stakeholders from different countries at each step should enhance the generalizability of the COS. Moreover, using methods that capture and preserve the views of participants (e.g., separating the production of knowledge and decision-making, using tailored wording, several languages, mediation techniques, etc.) will enhance the credibility of the COS by lowering the risk of token participation.

Proof of concept of these methods is ongoing with the development of a COS for depression. PROCEED is an online study that was able to rapidly involve thousands of participants from multiple countries for generating outcome domains for depression (1000 after 1 week, 2000 after 2 months of advertisements on social media, professional mailing lists and specialized media and 3000 after another 4 months by only a snowball effect). Three questionnaires about expectations toward treatment in 3 languages were disseminated to patients, carers (e.g., family and friends) and clinicians (e.g., psychiatrists, psychologists, family doctors, nurses). About 80 potential outcome domains for trials were identified. Moreover, because

depression implies cognitive impairments, psychomotor retardation, pessimism and social withdrawal (i.e., symptoms that can be a barrier to patient participation, particularly answering open-ended questions), this approach is likely effective for other diseases.

Our work on COS development allowed us to have reflective considerations about the epistemological (how the knowledge is built) and ethical (kind of relationship of the people involved and their individual and social consequences) implications of COSs. As mentioned previously, developing a COS involves making a decision, and decision-making processes are determined by underlying values, which are rarely clarified. A way to determine underlying values is to raise the question of COS utility. COSs have proven scientific utility because they enhance knowledge by standardizing outcomes and thus allow for the comparison of study results. Beyond scientific utility (knowledge production), COSs also have social utility. As a thought exercise, we can imagine a COS, composed of standardized and valid instruments but without any clinical relevance, or worse, without any relevance for patients. Because questions of social utility and ethics have a normative purpose, they should be democratically debated.

If we consider that the ultimate social utility of COSs is to improve patients' conditions, participatory action research (PAR), which aims at transforming practice by solving socially meaningful research questions, could be a possible epistemological and ethical framework for COS development [40,41]. For COS development, the practice will be COS uptake in trials and the research question will be the development of the COS. The ethics of PAR rely on 2 underpinning values: the idea that truth consists of the alliance of several perspectives (e.g., that deciding what outcomes are core is a shared decision between all stakeholders) and that people have the right to participate in processes that could potentially change their life

(e.g., patients who will receive the treatment should participate in the COS development) [42]. The epistemology of PAR relies first on the assertion that participants are the best source of knowledge concerning their experience and second that knowledge is an active process elaborated with the collaboration of all stakeholders. In PAR, researchers are facilitators helping to formalize the experiential knowledge and the professional expertise of the study participants [23].

Further development of the ethics of COS development is required: the elicitation of the values of COS development should help in finding adequate methods and defining guidance for decision-making.

Conclusion

COSs are effective tools to limit waste of research by improving the comparability and combination of trial results. Therefore, COS uptake in trials is the outcome of COS development. This article proposes some adjustments and alternatives to the current process of COS development to enable more generalizability and credibility and open reflections on the ethics of COS development and their methodological translations.

Acknowledgments

The authors acknowledge Sarah Gorst for sharing unpublished data.

Funding

This work was supported by the Fondation pour la Recherche Médicale, Paris, France *(under the number FMD20170637634)*

Bibliography

1 The COMET Handbook: version 1.0. - PubMed - NCBI. https://www.ncbi.nlm.nih.gov/pubmed/28681707 (accessed 17 Oct 2017).

2 Boers M, Beaton DE, Shea BJ, *et al.* OMERACT Filter 2.1: Elaboration of the Conceptual Framework for Outcome Measurement in Health Intervention Studies. *J Rheumatol* Published Online First: 15 February 2019. doi:10.3899/jrheum.181096

3 Chalmers I, Bracken MB, Djulbegovic B, *et al.* How to increase value and reduce waste when research priorities are set. *The Lancet* 2014;**383**:156–65. doi:10.1016/S0140-6736(13)62229-1

4 Williamson PR, Altman DG, Blazeby JM, *et al.* Developing core outcome sets for clinical trials: issues to consider. *Trials* 2012;**13**:132. doi:10.1186/1745-6215-13-132

5 Page MJ, McKenzie JE, Kirkham J, *et al.* Bias due to selective inclusion and reporting of outcomes and analyses in systematic reviews of randomised trials of healthcare interventions. *Cochrane Database Syst Rev* 2014;:MR000035. doi:10.1002/14651858.MR000035.pub2

6 Maxwell LJ, Beaton DE, Shea BJ, *et al.* Core Domain Set Selection According to OMERACT Filter 2.1: The OMERACT Methodology. *J Rheumatol* Published Online First: 15 February 2019. doi:10.3899/jrheum.181097

7 Beaton DE, Maxwell LJ, Shea BJ, *et al.* Instrument Selection Using the OMERACT Filter 2.1: The OMERACT Methodology. *J Rheumatol* Published Online First: 1 February 2019. doi:10.3899/jrheum.181218

8 Sautenet B, Contentin L, Bigot A, *et al.* Strong heterogeneity of outcome reporting in systematic reviews. *J Clin Epidemiol* 2016;**75**:93–9. doi:10.1016/j.jclinepi.2016.01.008

9 Gargon E, Williamson PR, Young B. Improving core outcome set development: qualitative interviews with developers provided pointers to inform guidance. *J Clin Epidemiol* 2017;**86**:140–52. doi:10.1016/j.jclinepi.2017.04.024

10 Bang D, Frith CD. Making better decisions in groups. *R Soc Open Sci* 2017;**4**. doi:10.1098/rsos.170193

11 Eun J. Consensus Building through Participatory Decision-Making. *Gest Manag Public* 2016;**me 5 / n° 2**:5–20. doi:10.3917/gmp.052.0005

12 Davis K, Gorst SL, Harman N, *et al.* Choosing important health outcomes for comparative effectiveness research: An updated systematic review and involvement of low and middle income countries. *PLOS ONE* 2018;**13**:e0190695. doi:10.1371/journal.pone.0190695

13 Handbook O. Handbook. OMERACT Handb. https://omeracthandbook.org/handbook (accessed 12 Mar 2019).

14 Kirkham JJ, Davis K, Altman DG, *et al.* Core Outcome Set-STAndards for Development: The COS-STAD recommendations. *PLoS Med* 2017;**14**. doi:10.1371/journal.pmed.1002447

15 Kirkham JJ, Gorst S, Altman DG, *et al.* COS-STAR: a reporting guideline for studies developing core outcome sets (protocol). *Trials* 2015;**16**:373. doi:10.1186/s13063-015-0913-9

16 Humphrey-Murto S, Crew R, Shea B, *et al.* Consensus Building in OMERACT: Recommendations for Use of the Delphi for Core Outcome Set Development. *J Rheumatol* Published Online First: 15 February 2019. doi:10.3899/jrheum.181094

17 Tunis SR, Maxwell LJ, Graham ID, *et al.* Engaging Stakeholders and Promoting Uptake of OMERACT Core Outcome Instrument Sets. *J Rheumatol* 2017;**44**:1551–9. doi:10.3899/jrheum.161273

18 Prinsen CAC, Vohra S, Rose MR, *et al.* How to select outcome measurement instruments for outcomes included in a 'Core Outcome Set' - a practical guideline. *Trials* 2016;**17**:449. doi:10.1186/s13063-016-1555-2

19 Goldhahn J, Beaton D, Ladd A, *et al.* Recommendation for measuring clinical outcome in distal radius fractures: a core set of domains for standardized reporting in clinical practice and research. *Arch Orthop Trauma Surg* 2014;**134**:197–205. doi:10.1007/s00402-013-1767-9

20 Gorst SL, Gargon E, Clarke M, *et al.* Choosing Important Health Outcomes for Comparative Effectiveness Research: An Updated Review and Identification of Gaps. *PloS One* 2016;**11**:e0168403. doi:10.1371/journal.pone.0168403

21 Gorst SL, Gargon E, Clarke M, *et al.* Choosing Important Health Outcomes for Comparative Effectiveness Research: An Updated Review and User Survey. *PloS One* 2016;**11**:e0146444. doi:10.1371/journal.pone.0146444

22 Gargon E, Gurung B, Medley N, *et al.* Choosing Important Health Outcomes for Comparative Effectiveness Research: A Systematic Review. *PLOS ONE* 2014;**9**:e99111. doi:10.1371/journal.pone.0099111

23 Corbie-Smith G, Wynn M, Richmond A, *et al.* Stakeholder-driven, consensus development methods to design an ethical framework and guidelines for engaged research. *PloS One* 2018;**13**:e0199451. doi:10.1371/journal.pone.0199451

24 Domecq JP, Prutsky G, Elraiyah T, *et al.* Patient engagement in research: a systematic review. *BMC Health Serv Res* 2014;**14**:89. doi:10.1186/1472-6963-14-89

25 Kirkham JJ, Boers M, Tugwell P, *et al.* Outcome measures in rheumatoid arthritis randomised trials over the last 50 years. *Trials* 2013;**14**:324. doi:10.1186/1745-6215-14-324

26 Kirkham JJ, Clarke M, Williamson PR. A methodological approach for assessing the uptake of core outcome sets using ClinicalTrials.gov: findings from a review of randomised controlled trials of rheumatoid arthritis. *BMJ* 2017;**357**:j2262. doi:10.1136/bmj.j2262

27 US Department of Health and Human Services, Food and Drug Administration. Guidance for Industry Clinical Development Programs for Drugs, Devices, and Biological Products for the Treatment of Rheumatoid Arthritis (RA). 1999.[http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM071579.pdf]

28 The European Agency for the Evaluation of Medicinal Products, Unit for the Evaluation of Medicinal Products for Human Use. Guideline on Clinical Investigation of Medicinal Products other than NSAIDs for Treatment of Rheumatoid Arthritis. 2003.http://www.emea.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003439.pdf

29 Tran V-T, Barnes C, Montori VM, *et al.* Taxonomy of the burden of treatment: a multi-country web-based qualitative study of patients with chronic conditions. *BMC Med* 2015;**13**:115. doi:10.1186/s12916-015-0356-x

30 Kreuter F, Presser S, Tourangeau R. Social Desirability Bias in CATI, IVR, and Web SurveysThe Effects of Mode and Question Sensitivity. *Public Opin Q* 2008;**72**:847–65. doi:10.1093/poq/nfn063

31 Denzin NK, Lincoln YS, editors. *The SAGE Handbook of Qualitative Research*. Fourth Edition edition. Thousand Oaks: : SAGE Publications, Inc 2011.

32 Braun V, Clarke V. What can "thematic analysis" offer health and wellbeing researchers? *Int J Qual Stud Health Well-Being* 2014;**9**. doi:10.3402/qhw.v9.26152

33 Tran V-T, Porcher R, Tran V-C, *et al.* Predicting data saturation in qualitative surveys with mathematical models from ecological research. *J Clin Epidemiol* Published Online First: 24 October 2016. doi:10.1016/j.jclinepi.2016.10.001

34 Gupta UG, Clarke RE. Theory and applications of the Delphi technique: A bibliography (1975–1994). *Technol Forecast Soc Change* 1996;**53**:185–211. doi:10.1016/S0040-1625(96)00094-7

35 Humphrey-Murto S, Wit M de. The Delphi method—more research please. *J Clin Epidemiol* 2019;**106**:136–9. doi:10.1016/j.jclinepi.2018.10.011

36 Donohoe HM, Needham RD. Moving best practice forward: Delphi characteristics, advantages, potential problems, and solutions. *Int J Tour Res* 2009;**11**:415–37. doi:10.1002/jtr.709

37 Miller GA. The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychol Rev* 1956;**63**:81–97.

38 Maturo A, Sciarra E, Ventre AGS. Counselling: Decision Making, Consensus, and Mediation. *Procedia - Soc Behav Sci* 2010;**5**:1770–6. doi:10.1016/j.sbspro.2010.07.362

39 Handbook O. Workbooks & Resources. OMERACT Handb. https://omeracthandbook.org/workbooks-%26-resources (accessed 30 Sep 2019).

40 Baum F, MacDougall C, Smith D. Participatory action research. *J Epidemiol Community Health* 2006;**60**:854–7. doi:10.1136/jech.2004.028662

41 Bergold J, Thomas S. Participatory Research Methods: A Methodological Approach in Motion. *Forum Qual Sozialforschung Forum Qual Soc Res* 2012;**13**.http://www.qualitative-research.net/index.php/fqs/article/view/1801 (accessed 13 Sep 2018).

42 MacDonald C. Understanding Participatory Action Research: A Qualitative Research Methodology Option. *Can J Action Res* 2012;**13**:p34-50.

Fig.1 Current method of COS development and proposition of adjustments/alternatives to enhance generalizability and credibility. (adapted from COMET Handbook)