Reviewer’s report

Title: Developing core outcome sets for clinical trials: issues to consider

Version: 1 Date: 12 May 2012

Reviewer: Joy C MacDermid

Reviewer’s report:

This is an important issue; and reflects an emerging issue. A number of groups are starting to realize the benefits for agreeing on common outcome measures across studies or clinical contexts. There is a benefit to having a a structure; and framework for such exercises. The paper has valuable contributions; but in my view treats some areas of the issues superficially. There is a potential danger in not discussing the pros and cons of some aspects of this process in more detail in that some groups may not be sufficiently rigorous in their methods and conceptualization on coming forward with recommended core measures. Thus, some additional discussion of the potential dangers of preliminary establishment of core measures should be considered along with the potential benefits to be had. Overall, the paper is useful in that it generates some discussion on this topic. I am not sure that the checklist is ready for prime time yet as it seems like it could use further development; or at least more information about how the item should be interpreted or presented. Overall, it is a timely important issue.

Specific comments

I do not necessarily see the connection or the statement below and GRADE. The fact that GRADE recommends up to seven important outcome measures be considered (more realistically suggests that that would be a maximum), does not really directly lead to establishing core measures for trials. I agree there is a rationale for doing so; and the authors have made the case well-but not sure that GRADE is it. I would suggest removing that sentence.

The development of a COS for trials overlaps with the work being done by the GRADE (Grading of Recommendations Assessment, Development and Evaluation) group (http://www.gradeworkinggroup.org), which recommends that up to seven patientimportant outcomes are listed in the Summary of Findings tables in systematic reviews

The paper discusses the importance of patient involvement in establishing outcome measures that are core to future clinical research. However, there is extensive experience with establishing guidelines and recommendations in a number of current areas of clinical practice that while the importance of patient perspective is often mentioned it is commonly operationalized poorly.
Researchers who lead consensus exercises are often quite comfortable with reviewing/synthesizing the literature to determine what current practices exist in use of outcome measures and how those might relate to conceptual frameworks that they use. They are less comfortable in embracing qualitative methods on an equal basis with quantitative methods. I would suggest they are even less comfortable with having patients as a driving force in setting agendas for outcome measures (or other aspects of clinical research agendas). Having a single (or small number) of patient representative on a outcome measure panel may not empower patients to establish their agendas and to move forward in a direction that is patient centric in terms of priorities. Ultimately the researchers may feel that they have listened to patients but the process may not actually be very patient centric unless extraordinary efforts are made in this regard. For example, when conducting consensus exercises that are based on voting if patients do not have the same number of votes as researchers than their voice may not come through in the consensus process.

Another issue to be considered is how the important concepts are identified. Researchers like conceptual frameworks and to fit things into those frameworks. As important as conceptual frameworks are in research they may drive the process rather than explain it. Furthermore, having patients endorse a menu of different types of outcome measures or concepts concepts may not bring forward their concerns in the same way that starting from scratch to determine those important concepts might direct the process. A little bit more about the importance of being conscious about the order and priority about looking at qualitative versus quantitative information, how both are used to generate conceptual items of importance, and the relative empowerment of patients in the process could be better elaborated on.

The meaning of the following sentence taken from the text is not very clear. Furthermore, while the first part is accurate that COSMIN does provide a detailed method of looking at the reliability and validity of measurement tools it should be acknowledged that it is a complex appraisal process; and the reliability of the process itself has been variable across studies.

Resources exist for determining the validity and reliability of a particular measurement instrument through the work of the COSMIN (COnsensus-based Standards for the selection of health Measurement Instruments, www.cosmin.nl) group whilst feasibility is a further consideration.

The following recommendation taken from the document seems quite arbitrary. It would seem that these type of recommendations should have an empirical basis. A number of rounds may be held in which responses are summarised and fed back to individuals, allowing them to change their score in light of the group’s opinion. Consensus regarding whether an outcome should be in the core set could be defined as 70% or more respondents scoring it 7-9 and less than 15% scoring as 1-3. Consensus that an outcome should not be included in the
core set could be defined as 70% or more scoring as 1-3 and less than 15% as 7-9. All other score distributions would be taken to indicate disagreement.

It might be challenging in some areas to establish an outcome measure as being critical versus important or of limited importance because this could be variable across patients. This would be particularly true for interventions or conditions where we know there are higher rates of issues around patient preference/trade-offs etc. Again this does not suggest we should not try to do so but some discussion of the potential that falls is formed

The document does not consider potential downsides of establishing core sets if they are done before there is sufficient evidence on the particular outcome measures that might be used. I think there are ways to handle this. I know some groups who have established “core measurement” recommendations have focused on the concepts and stated that there is no clear ideal measure for that concept. Others have suggested several outcome measures that might be used for that concept. Others have suggested a single measure to try to promote uniform practice-usually when there is at least moderate quality evidence for that tool. However, there are a large number of clinical assessment measures and patient reported outcome measures that have had limited investigation that ultimately might be wonderful instruments with better investigation of the their clinical measurement properties. I think that establishing core outcome measures needs to carefully consider the pros and cons of the process in terms of promoting more comprehensive and uniform assessment of concepts; but not recommending specific instruments that might stifle the development and innovation in the clinical measurement area. Some discussion of this as being part of the process seems warranted.

The checklist—comments by item

3. The fit for PRISMA checklist is not intuitive—I could see lots of circumstances where it does not apply and I’m not sure why it is being promoted for this particular aspect. Conversely, the role for a conceptual framework such as The International Classification of Functioning Disability and Health or some other framework of what constitutes an outcome seems to be absent.

5. It seems very unlikely that a study protocol would exist for a consensus exercise at least the way they are currently happening. This seems like a carryover from what is included in other checklists; but is less likely to be applicable here.

6. Eligibility is an important item—however more detail about justifying the role for different stakeholders seems warranted.

7. This is also an important item to know what information is provided to the participants ahead of time and how it was summarized are presented; and at what time point in the process. It seems to me that items six and seven are critical indicators and should be detailed to a greater extent than they are in the
manuscript (and discussed in greater depth in the discussion).

9. This item is very unclear. It is dangerous to use outcomes scoring when talking about outcome consensus because it generates confusion. I am not clear if this item is about looking at the scoring algorithms of instruments; or how the consensus exercises would be scored (I think the latter but you get my point about confusion of concepts).

12. Findings of such exercises could fall under different domains concepts to be measured, potential measures, considerations for modifying the core sets for practice versus research, or epidemiological versus clinical studies etc.-i.e. different contexts. Groups may also identify subgroups of patients who may not be appropriate for the core set, may require additional considerations etc. Furthermore, such consensus exercises often identify gaps in knowledge, areas where there is an acknowledged need for new measures or testing of existing measures; and concerns about how measures are interpreted or communicated. It seems like there should be categories defined here.

17. Conflicts of interest should specify that being a developer of an instrument is a conflict. It is not necessary that these people would be excluded from development panels but they should declare their status as outcome measure developers as this may influence their opinions.

In other words--- it seems that there is need for greater clarity in the checklist items and how they would be interpreted/presented.

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Declaration of competing interests:

I am an outcome measure developer and have participated in consensus exercises to establish core outcome measures. All of these have been conducted without financial reimbursement/incentives