Reviewer’s report

Title: COS-STAR: A reporting guideline for studies developing core outcome sets (Protocol)

Version: 3
Date: 20 July 2015

Reviewer: George Tomlinson

Reviewer’s report:

Since this is a protocol, there is little background on core outcomes sets and their use, and a focus principally on the methods of the planned project. The protocol lays out a straightforward plan of action, mainly without justification for the separate steps, and clearly defines the various participant groups and their interactions. It is clearly written and the authors are centrally involved in the COMET core outcomes initiative.

My first comment below assumes that the project is still amenable to change. The second and third comments are about the manuscript describing the protocol.

Major comments:

1. For the Delphi exercise, the plan is to approach approximately 200 COS developers, 220 journal editors, 8 members of the PoPPIE group and what looks like thousands of PIs from open phases III/IV trials on ClinicalTrials.gov. Presumably, there will be a much higher percentage of respondents from the first three groups but there is a possibility that the opinions of these three more informed groups will be washed out what could be a majority of ‘trialists,’ even if only 10-20% respond Presumably this is why there will be a stage 3 review giving priority to the critical items from the Delphi exercise. This is the stage at which the steering committee, expert panel and a few other stakeholders will make a final decision on the items to be included in the reporting guidelines. If items are considered to have been missed or wrongly included at the threshold of 70% with a score of 7-9, the consensus meeting has the chance to make a correction. My comment is this - do you really need to approach all the PIs of eligible trials? A much smaller random sample or a stratified sample from a few different clinical areas could be more efficient and provide just as much useful information. A smaller sample might allow you to provide a little more inducement to participate (through personalized emails, for example) from this likely less-engaged group. A smaller sample of PIs also gives them a more appropriate weight in the ranking exercise and may leave the consensus panel with an easier job at the end.

2. The manuscript might be strengthened by some justification of the choices of the various participants of the Delphi exercise and the numbers in each group.

3. The manuscript would be strengthened by a description of the methods
whereby the consensus group will use the results of the Delphi exercise. The earlier stages are well-described but the process in stage 3 is somewhat vague. As it currently reads, a pragmatist might wonder about the value of the earlier stages when the final COS-STAR guidelines are decided by an expert panel and a few additional stakeholders. I understand that some details will depend on the findings of the Delphi exercise, but I am curious about things like how much weight will be given to the results of that exercise in relation to the opinions of the consensus panel members, if there is a maximum size for the number of items in the reporting guidelines, or if there are previous examples of decisions at similar consensus meetings that could be used to give a sense of how the decisions will be made here.

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests.