Author's response to reviews

Title: Instruments to measure patient experience of health care quality in hospitals: A protocol for systematic review

Authors:

Michelle Beattie (michelle.beattie@stir.ac.uk)
William Lauder (william.lauder@stir.ac.uk)
Iain Atherton (iain.atherton@stir.ac.uk)
Douglas Murphy (D.Y.Murphy@Dundee.ac.uk)

Version: 3
Date: 27 November 2013

Author's response to reviews: see over
Dear Professor Moher and colleagues

Instruments to measure patient experience of health care quality in hospitals: A protocol for systematic review

Many thanks for considering the above paper for publication in the Systematic Reviews Journal.

The reviewers’ comments have been very helpful and the manuscript has had the necessary changes made. We believe the subsequent revisions have improved the overall clarity of the protocol. The reviewers’ comments are noted in italics and detailed responses are identified using the ** symbol.

1. **P8, Study Objectives. It seems that objectives 1, 2, 3 and 5 are largely concerned with describing the literature, whereas 4 and 6 are more analytical/interpretive. It is the methods relating to study objectives 4 and 6 that need to be a bit clearer (see point IV below).**

**This is a helpful point. Objectives 1-3 are descriptive, whereas objectives 4, 5 and 6 are analytical/interpretive. Further explanation of how these objectives will be achieved are given under point 4, particularly how the Utility Index and COSMIN checklists will be operationalised and integrated into the narrative synthesis.**

2. **P8, Study Method. The first sentence here is tricky to follow. Perhaps separate into two sentences.**

**Separating the first sentence into two sentences has made the point easier to read.**
3. **P9, Inclusion Criteria.** The authors state that they will double screen a random 10% sample of retrieved papers, with disagreements resolved by consensus. Presumably this is to test the level of inter-reviewer agreement for study selection, but what if this turns out to be low? Will they then double screen the remaining 90% of studies or simply accept that error and/or bias might have crept in at this stage? What level of agreement for the 10% sample would be sufficiently reassuring?

**Many thanks for bringing this important point to our attention for further consideration.** We will ascertain the level of inter-reviewer agreement by calculating Cohen’s kappa statistic. As the instruments included may be used for high stakes purposes we will be aiming for a high level of inter-reviewer agreement ($k>0.8$) in our 10% sample. Where agreement falls below this threshold we will ensure the remaining 90% of titles and abstracts are double checked. This will enable us to be explicit about the level of agreement between two reviewers for this step. Where the threshold is not met with two reviewers we will consider the feasibility implications of increasing the number of reviewers. If resource implications preclude increasing reviewers we would acknowledge this limitation and determine a sufficient number of reviewers needed (D-studies) in our recommendations. Explanation of this has been inserted on Page 9, under Inclusion Criteria.

4. **P11-13 Assessment of quality to data analysis.** The authors describe the content of the van der Vleuten utility index and the COSMIN checklists, but exactly how these will be implemented is not clear. For example, the first sentence in this section states “Identification of the instruments intended use will enable us to judge the relative importance to be placed on the five components of the van der Vleuten’s utility index” and on the next page “…therefore categorising instruments according to use will enable us to judge the balance of utility metrics.” This raises the question of how this weighting will be operationalized and the further question of exactly how it will be incorporated into the synthesis. The same applies to the COSMIN evaluation.

The authors state that they plan to undertake a narrative synthesis following the ESRC guidance; because of the lack of standardisation of narrative synthesis, this guidance highlights the importance of being as specific as possible at the planning stage to avoid spurious comparisons, data dredging etc. While it may not be possible for the reader to describe every aspect of the planned synthesis, it will benefit the reader (and review) if at least the role of utility and methodological quality assessments in the synthesis are given a priori.

**The suggestion for clarity around how the Utility Index and COSMIN checklists will be applied and integrated into the narrative synthesis has been helpful.** In fact, after further consideration, we have reversed the order in which we will apply these tools. We had initially planned to conduct the utility matrix, which would be determined on the instruments purpose. However, having retrieved a sample of papers on these instruments, it is now apparent that not all papers make explicit the instruments primary purpose. We will now apply the COSMIN checklist before the utility matrix. This will ensure we can determine the aspects of utility that have been tested and their subsequent results. From this, we can populate the utility matrix and determine the potential purpose of the instrument. The decision making around the purpose will be made explicit in the narrative discussion. Grouping the instruments according to purpose will also enable us to provide a narrative synthesis which will lead to recommendations for policy, practice and research (as per study objective six).
Explanation as to how the utility matrix will be implemented is explained on page 12. Figure two has been added to display the criteria to be determined for each aspect of utility. Details have been given on page 12 and 14 of how the COSMIN checklist will be implemented. Figure 3 has also been added to augment explanation of the COSMIN checklist. The incorporation of these tools into the narrative synthesis has been explained under data analysis on page 14 and 15.

5. **P1, 1 Defining Quality.** The word “quality” is used to describe both health care quality and the methodological rigour. The authors should use terminology that disambiguates these two concepts.

**The two concepts have been clarified. “Quality” in relation to methodology has been referred to as rigour or scrutiny throughout. “Quality” in relation to health care has remained.

6. **It is unclear how or where statistical analysis will be used.**

**The COSMIN team are currently exploring the possibility of numerical grading to establish levels of quality for psychometric instrument studies and results. This work remains under development; therefore we will be unable to attribute numbers associated with definitive levels of quality. References 20 and 32 have been added which explain the COSMIN position.

**Cohen’s kappa statistic will be calculated to ascertain inter-reviewer agreement during application of the inclusion criteria to titles and abstracts as previously explained.

**We will not be conducting a meta-analysis of results as this would be inappropriate for the study question.

7. **The repeated use of the phrase “the patients’ experience” can jar a little – would simply “patient experience” be an adequate substitute? A proofread might help catch remaining typos (e.g. p.9, “secondary reference” should be pluralised; p.11, final sentence under ‘Data extraction’ should begin “Where consensus…” and the occasional inconsistent use of apostrophes.

The patients’ experience has been replaced with patient experience throughout. The protocol has been proofread and necessary changes made.

I hope you will find the revisions satisfactory and look forward to hearing from you.

Yours sincerely

Michelle Beattie
Lecturer