Reviewer’s report

Title: Did we describe what you meant? Findings from an empirical validation study for a systematic review of reasons

Version: 1 Date: 5 June 2014

Reviewer: Jonathan Ives

Reviewer’s report:

I should note that I know two of the authors. I have worked (taught) with DS on two occasions, and I expect to be hosting MM as a visiting academic to my institution in 2015. I do not feel this represents and conflict of interest that prevents me from carrying out a fair review.

This is a well written paper, on a topic of relevance to the journal and likely to be of interest to its readership, though perhaps a niche.

The paper reports an attempt to ‘validate’ a novel methodology developed by two of its authors. This methodology has been published in respected peer reviewed journals.

The paper is generally of good quality, and the idea if sound and worthwhile, but I have some suggestions for essential changes

Major revisions

The discussion of why the response rate was so low is (pp12-14), I think, overly defensive and focuses on the supposition that authors are prejudiced against, and ignorant of, systematic reviews of reasons, without good justification. It makes a controversial discussion point, certainly, but there is nothing in the data to suggest that this is the case, and there is no reason to suppose this is a better explanation for the low response rate than people being busy. The latter is declared to be ‘unlikely’ on page 13, but no reason is given for why this is unlikely. The authors are entitled to present their suppositions, but I think that (a) it should be made very clear that this supposition is not based on any evidence and (b) making this claim will likely alienate and annoy many readers. Particularly, the word ‘ignorant’ in English is more often than not used in a pejorative sense, and may well cause unintended offence. I suggest this discussion is toned down, and reworded. As it stands, the language used does not come across as scholarly, and has an angry and almost ‘bitter’ tone, which I do not think is intended.

In general, I feel it is not clear whether the process of validation by author checking is aimed at checking the ‘reason mentions’ against what the authors actually wrote, what you understand them to have meant, or what they themselves understand themselves to have meant. I think that foregrounding this distinction, and making it clear what you are interested in, will help make sense
of the parts of the discussion (for example, on page 11 where discussion point 1 is discussed) that talk about to what extent the validation has been successful. This distinction is quite central to understanding how it was determined that that an error had or had not been made, and in-particular who is authoritative on the matter.

On a related point, I thought that the general problems with any validation methodology based on member checking should be considered, in-particular question of to what extent can the author/participant be considered to have special insight and what extent the researcher can override that insight. These are part of an ongoing debate about rigour in qualitative research, and that debate is, I think very relevant to the work being done in this paper.

Minor revisions

P4. You discuss the importance of a systematic review of reasons as a way to avoid bias and ensure completeness, and stress the importance of validating coding in a systematic review of reasons to ensure the systematic review was performed well, but some might argue that robust peer review already does this. The whole point of sending it out to expert reviewers is supposed to ensure that positions are correctly presented and no key arguments are missed out. Can you offer some counter to this claim, to explain why peer review is not sufficient to validate the results of a review of reasons (i.e. why go to all this trouble)? It would be very useful to have a clear and strong argument, early on, for why conducting a systematic review of reasons is preferable to not doing one, and how this differs from simply good scholarly practice when examining the literature (which is what ethicists/philosophers are more likely to do).

P5. Line 91. the phrase ‘with the according different coding’ is unclear. Line 96, the sentence ‘This coding was taken over for the questionnaire’ is unclear’.

P6. Line 109. the phrase ‘according coding’ is unclear.

P6. Line 126. It is not entirely clear, to my mind, why some references were rejected for the survey. The table does not really help me to understand. It is not clear what ‘not made into questionnaire’ means, and after reading the explanation I am still not sure why some were explicitly rejected in the pilot study.

P6/7. Line 129-130. You mention that the pilot and the final questionnaire were not significantly different so that could be analysed together, but so imply that there were differences. Can you let us know what these differences were, or what kind of differences they were (some example perhaps).

P7. You use the phrase ‘positive response’ to refer to a questionnaire response that was completed and returned and was usable. This is confusing, as ‘positive’ has multiple meaning in English, and in the context of their paper it is too easy to confuse his with questionnaire that gave positive answers to the questions (i.e. said the coding was correct). I suggest you use the terms ‘complete/incomplete’ to refer to questionnaires instead of ‘positive/negative’. This needs changing throughout the paper.
Some examples are given of the neutral author validating claims of incorrectness. Can you indicate whether this exhaustive list, or just selected examples? I assume the latter?

This is very long sentence, and consequently very hard to follow. Can it be broken up?

I am not convinced that the claim about the questions listed being relevant to overarching questions of quality in bioethics is obviously true. I think that this connection needs to be established, rather than taken for granted. For example, it is not obvious that the question about incorrect coding relevant to quality in bioethics. Similarly, why should someone not convinced by the need for a systematic review of reasons be at all concerned about why 70% of contacted authors did not take part in the validation? These questions seem quite specific, and not clearly related to general questions about quality in Bioethics.

Can subheadings be used to structure the discussion, rather than ‘Regarding (1)’ etc?

This is another very long sentence that needs breaking up.

You state that you have several good reasons to believe that your methodology was appropriate and can be recommended for future attempts, but you do not expand on this by outlining those reasons – which really needs to be done.

Level of interest: An article whose findings are important to those with closely related research interests

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