Reviewer's report

Title: A systematic review of Impact of Routine Collection of Patient Reported Outcome Measures on Patients, Providers and Health Organisations in An Oncologic Setting

Version: 3 Date: 15 April 2013

Reviewer: Joanne Greenhalgh

Reviewer's report:

Thank you to the authors for their work in revising the manuscript. They have addressed most of my comments sufficiently and raised some interesting issues for debate in the process.

Major points:

I still think the manuscript would benefit from further attention to my fifth comment about how the authors produced the findings in the last paragraph in the results section. In my reading, this para is trying to determine the contexts in which the impact of PROMs feedback on different outcomes can be maximised. However, I still do not feel it is altogether clear how these conclusions were reached.

First, the authors indicate that they have drawn these inferences from the 200+ full text articles they initially reviewed, but also state that these articles are not included in the review ‘proper’. If we are to judge this paper according to the quality standards of traditional systematic reviews, it is confusing and somewhat misleading to present ‘results’ or ‘findings’ or ‘inferences’ from papers that are not, according to the authors’ inclusion and exclusion criteria, included in the review itself.

Second, referring to the 27 studies included in the review, the authors indicate that they reviewed mediation effects through conducting path analysis and a multiple stage regression approach and assessed moderating effects through explicitly testing the interaction effect/moderating effect and through subgroup analysis. A summary of the strength of the findings on different outcomes for each study is provided in Appendix four. However, it is not clear how this data was interrogated or analysed to come up with the findings in the last para of the results. Tables do not speak for themselves and it would be useful to have a description in the methods section of the paper to explain how the authors moved from Appendix 4 to the findings in the last para of their results. Currently, little detail is given in the methods section of the manuscript about how the results were synthesised, other than ‘No attempt was made to quantitatively synthesise the results as data were too heterogeneous to support pooling’. It would be helpful to have further information in this section to describe what kind of synthesis was conducted and how the study findings were interrogated to produce the findings outlined in the last paragraph of the results section.
More minor points:

The feedback below is more for debate. I welcome the authors’ interest and willingness to debate methodological issues in their responses.

In response to point 1, I am not altogether convinced by the argument of adopting the same methodology to answer research questions 1-3 ‘for consistency’. The research questions are asking different things and it is also valid to argue that they should be answered using different methodologies.

Furthermore, a ‘causal mechanism’ is different from an endpoint. An intervention such as PROMs feedback has a very long implementation chain and depends on the achievement of multiple intermediate outcomes to achieve the ‘final’ outcome of improving the patient’s health status. The authors have quite rightly used existing models to identify these multiple intermediate outcomes and then examined the extent to they are affected by PROMs feedback, thus answering their first research question. They have done a very credible systematic review and have provided some useful insights into whether PROMs feedback works in cancer settings for different outcomes. However, this analysis is not designed to shed any light on the process through which the achievement of one intermediate outcome may or may not lead to the achievement of another. That is, it does not tell us anything about the different reactions that clinicians or patients may have to PROMs data that might explain why, for example, PROMs feedback improves communication but does not necessarily change how the patient is managed. The PROMs feedback itself does not have ‘causal powers’ to produce outcomes. PROMs feedback offers resources to clinicians and patients and it is their choices and decisions to act on or utilise these resources (or not) that give rise to changes in outcomes. This is what I would define as a ‘mechanism’ (drawn from Ray Pawson’s work) and I do not think the current review has addressed these causal mechanisms.

Further comments

The authors raise some important and interesting criticisms about realist synthesis. A common criticism of the method is the lack of guidelines for conducting an RS, raising questions about its repeatability; Wong and colleagues have gone some way to addressing this criticism in their recent publication of standards for Realist Synthesis from the RAMESES project (Wong et al, 2013, BMC Medicine, 11:21) – though obviously this published after the current review.

A further source of confusion in RS is how to weight evidence from different study designs – which speaks to the issue of inclusion and exclusion criteria and quality appraisal. In RS, decisions about inclusion and quality appraisal depend on the study’s role in testing the theory. RS is focused on explaining ‘what works, for whom, in what circumstances and why’ through theory development, testing and refinement. Different components of the theory are tested using different study designs. For example, hypotheses about the optimal contexts for the intervention are tested in comparative outcome data (for example, from trials);
claims about the reactions of particular groups of subjects (ie mechanisms) are tested using qualitative data; implementation ideas are tested in process research, and so on. So decisions about whether to include a study in the review is based on a determination of whether this study is relevant to testing a particular component of the theory, whether the design is the most useful study design to test this component of the theory and secondly, given the design of the study, whether the study is a high quality example of that study design. Quality appraisal is thus done on a case-by-case basis as appropriate to the method utilised in the original study. This is very different from traditional systematic review, where the focus is on answering the question ‘did it work, or not?’. Here it is much simpler to construct a hierarchy of evidence with the RCT at the top.

Finally, RS is not focused on deriving quantitative endpoints for cost effectiveness analysis and it would be inappropriate to use the methodology if that was the focus of a review.

**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 have no conflicts of interest, other than that already stated (as a proponent of realist synthesis)