Author's response to reviews

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

Authors:

Jack Chen (jackchen@unsw.edu.au)
Lixin Ou (lixin.ou@unsw.edu.au)
Stephanie Hollis (s.hollis@unsw.edu.au)

Version: 4 Date: 7 May 2013

Author's response to reviews: see over
We thank the reviewer for the thoughtful comments and provide our responses as follows.

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.*”

We would like to clarify that most the 200+ full text articles did not meet the inclusion criteria and may cover a completely different issue (such as the methodology in developing clinical and meaningful minimum changes from PROMs) and so our conclusions and summation of results are only based on the 27 studies that we presented in the detailed analysis. In a systematic review covering a topic with very broad settings, definitions and research designs, it is often expected that the reviewers will extract a large number of papers to make the decision if the publication should be included in the review. We have added the following statement in the paper to make this clearer.

“The results and conclusions drawn were based on the 27 studies included in the analyses despite the large number of full-text articles extracted.” (Page 11).

“*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.*”

We added the following paragraph to the methods section:
“In order to answer review questions 2 & 3 for the studies included, all possible explicit mediation effects were reviewed through examining if a path-analysis or a mediation-analysis by multiple, staged regression approach was presented in the paper. To examine potential moderating effect, each study was examined to determine if it explicitly tested the interaction effect/moderating effect, or inexplicitly conducted subgroup analysis. Significant possible mediating or moderating effect results were indicated as part of review endpoints in Appendix 4. Inferences made and the discussion were based on these results.” (Pages 8-9).

We also made more explicit referencing for the last paragraph (page 17, paragraph 2) to make the source of discussion more related to the referred results table.

1. More minor points:

We would like to thank the reviewer in providing further views on a RS which are both insightful and thought-provoking. We have provided some of our thoughts below, being those who have a keen interest on a RS but with no any real expertise in conducting one. Some questions raised here are more or less a wish list from the newcomers and these discussions are better read with our lack of expertise on a RS in mind.

“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.”

The reviewer’s discussion was very interesting with regards to the potential ‘mechanism’ that the current review methodology was unable to discover. There is potential room for realist synthesis to shed more light on these potential ‘mechanisms’ as we have acknowledged in the discussion. The current review is to provide a synthesis, based on a conventional systematic review methodology with some hypothetical critical causal links between the collection of PROs and selected outcomes as the endpoints (from a well-developed and argued framework). The review results are predetermined and limited. There are different levels and layers of ‘mechanisms’ or ‘causal theories’
which are potentially moderated by different settings. To creditably uncover such ‘mechanisms’ is a difficult task and often beyond what a single review can achieve. We fully acknowledge the limitations of the current review (in fact, any review regardless of what type of jargon or methodology is used).

2. 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.”

It is our view that there is a great need for both theorists and practitioners of realist review to develop widely accepted guideline(s) and test its suitability. The value, validity, practicality and suitability of these guideline(s) for different stakeholders who have a different agenda and purpose in conducting a review should also need to be carefully articulated.

“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.”

Again, we felt that many conceptual points raised and statements made here would benefit from some clearly written, thoroughly thought-through and practically manageable guidelines. For example, the definition of ‘mechanisms’ could be entirely different to different stakeholders. We agree with the reviewer’s view that the tenet of a RS which aims to explain “what works, for whom, in what circumstances and why” provides great potential and much excitement and hopes for researchers, social scientists and decision-makers. However, we couldn’t entirely agree with the view that traditional systematic review only
focuses on the question ‘did it work, or not?’. We felt that this does not summarise well what a conventional/traditional systematic review is and perhaps is somewhat a misperception by some RS practitioners about what a conventional systematic review should be. In our view, the tenet of a conventional systematic review may also be understood as to recognize “what works, for whom, in what circumstances and why”, but for some pre-determined, selected endpoints (mostly, if not entirely, based on quantitative studies). It is less ambitious compared to a RS but with very good reasons. To fulfill these tasks, some systematic developed methodology and frameworks in how to explore the 'heterogeneity' (i.e. what, who, where, why and how) are well-established in a systematic review approach, using graphic plots, statistical tests, and meta-regression techniques, etc. The quantitative methods in combining different information from different types and sources of data have been developed (but admittedly, there is still much work to be done). These methods have been tested, debated and accepted with clear understanding of its limitations and strengths and new methods are still emerging. It is our hope that there will be similar well-developed tools and frameworks in a RS approach (not just slogans and claimed conceptual strengths) to enable its wide adoptions. It is true that some systematic reviews conveniently forgot the importance of exploring the ‘heterogeneity’ and resorted to bravely presenting a simple summary outcome from a meta-analysis which often provided misleading results and should never have been done in the first place. However, this is not how a conventional systematic review should be done, but a sign of failure to realise the intricacy of complex intervention, and a failure to understand the imperativeness in exploring the heterogeneity. This is where we would hope a RS could shed more light on the topic.

There are many more areas where concrete guidance in conducting a RS is needed. For example, it was emphasised that a RS would adopt the suitable ‘best design’ study to answer different questions, which is conceptually very sound. However, in a messy real world, it is often less clear how to implement such a principle. It may be worth noting that there are often not enough good quality studies to answer even one single question and the promise of exploring “what works, for whom, in what circumstances and why” and use the ‘best designed’ studies, may become an impossible mission. For example, it has long been realised that the impact of introducing PROs on patient outcomes should be ‘best’ tested based on cluster randomised controlled trials. However, there has been no such study ever conducted. We are left with many different designs with large varying qualities. What should be the inclusion and exclusion criteria for the ‘best’ study design based on a RS perspective? What would be the impact of different decisions on final review conclusions? Even for questions 2 & 3 in the current review for exploring potential mediating/moderating effects (small part of possible ‘mechanisms’), if we were to adopt a RS perspective, what types of studies should be best suited to answer these questions? Should only qualitative studies be included or should both quantitative and qualitative studies be included? If both types of studies are permissible, what content should be examined from quantitative studies in order to ascertain such an effect from a RS perspective? What types of results should we look into in order to identify possible mediating/moderating effects between implementation PROs and patient outcomes from a qualitative study, such as a focus-group
discussion? Suppose we have a universally agreed way to extract such information from a qualitative study, how do we make sense of the possible situation that different studies may generate totally contradictory conclusions? How do we assess the quality and weight of the results from both qualitative and quantitative studies and manage to research some sort of conclusions based on these results? What is a RS perspective if quite contradictory results emerge between different types of qualitative studies, between different types of quantitative studies, or between qualitative studies and quantitative studies?

Having a keen interest in RS, we strive for clear guidance on how to assess the quality and validity of a RS and how a RS could be used in guiding decision-making. Many such frameworks have been established for conventional systematic reviews. For any review methodology, there is always room for erroneous results. A clear discussion of the possible pitfalls and strengths in conducting a RS would greatly benefit all stakeholders and help in providing the rigor, reliability, validity, and transparency which are among the hallmarks of any good review methodology.