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: 3 Date: 9 March 2013

Author's response to reviews: see over
Dear Editor,

We would like to thank the reviewers for their helpful comments and provide our responses below.

Reviewer: Daniela C Goncalves

“This is a potentially useful manuscript, which summarises the literature on the routine use of PROMs in oncologic patients, ascertaining their impact on patients, providers, and health organisations. The authors went to great lengths to tabulate and assess the included studies, and provided a theoretical approach as to better integrate their results. There were however some questions raised while reading the manuscript, which I would like the authors to address.”

Thank for the positive comments and for raising important questions.

Introduction

1. “The introduction is considerably short and does not seem to properly introduce the study conducted. For instance …”

We have provided more detailed background and rationale as follows:

“There are growing interests from clinicians, researchers, industry and policy-makers in routinely collecting PROs to facilitate timely, patient-centred and evidence-based care. For example, the National Health Service (NHS) of the UK has been implementing a world-leading initiative of the routine collection of PROMs that firstly included a few selected elective surgeries (e.g., unilateral hip replacements, unilateral knee replacements, groin hernia surgery or varicose vein surgery) but are soon expanding to many other conditions such as mastectomy and breast cancer, among others. In the USA, the Patient-Reported Outcomes Measurement Information System (PROMIS), a National Institutes of Health funded initiative starting in 2004, is providing a publicly available Web-based resource that can be used to measure key health symptoms and HRQOL. The traditional paper-based PROs instruments are limited by its lack of flexibility, language and literacy requirement, possible inappropriateness towards minority groups, lack of timeliness (in generating instantaneous clinical meaningful interpretations) and inability to adopt the state-of-the-art measurement science such as Item Response Theory (IRT) and Computer Adapted Test (CAT) technique.” (Page 4)

Methods:

1. “The rationale for the current review is not clear to me, in the sense that the authors mention that they have identified “over 200 existing reviews on the same or similar topics (…) identified in a broad search covering PROs and quality of life measures in cancer patients between January 2000 and October 2011”. This is a considerably high number, and it would be useful to know how the authors identified these reviews and what their main findings are. It would also be useful to know, probably in the discussion, what the current review adds to the previously identified 200 reviews.”
We have used the similar search strategy as outlined in the method sections to search for review articles but without limitation of cancer patients. Most of the reviews retrieved are narrative and expert summation of pertinent literature in various topics which are not on cancer patients and not directly related to our review aims. We have checked titles and abstracts of each review to identify those systematic reviews with similar purposes or methodology that may be helpful to our study. No reviews were found having the same purpose, methodology and inclusion and exclusion criteria like the current study. We chose three recent systematic reviews (after 2005) that were useful as the base for our review. The results and conclusions of these reviews are presented in Appendix 1. Given that none of these reviews have adopted the same search strategy, theoretical framework and literature as included in the current review, their result and conclusions are understandably different from the current review findings. The search for reviews was not restricted to cancer patients so we have deleted ‘in cancer patients’ and apologise for the typographical error and confusion caused (Page 6).

2. The authors then state that “Three reviews were identified as the baseline reviews for this project”; I would like to know what was the rationale for choosing these three reviews, especially when considering that two of them do not specifically address the use of PROMs in cancer settings.

Please see above. We chose the three recent systematic reviews to assist in forming our search strategy and outcome frameworks, selecting only the most recent ones and including those not necessarily on cancer patients. We felt that to summarise the results of 200+ review articles, where most were not in a cancer setting and on different topics, fell beyond the scope of current review. However, we provided a summary table on the three baseline reviews that contributed more to the concept of the methodology of the current review (Page 7).

3. “The authors mention backward and forward search for all “over 200 in total” reviews identified, plus the search strategy developed; however, from the PRISMA flow diagram, it is not clear how many references were identified through each method, amongst other details. For instance, where all backward and forward references assessed for the “over 200 in total” reviews?”

We would like to clarify that we did not include all 200 reviews in the backward and forward search but the three baseline reviews (after the assessment of 200 reviews) as the start point.

4. The first inclusion criterion is that “substantial content in presenting empirical evidence on the impact of routinely collected PROs on at least one of the outcomes listed on table 2”; however, in all papers included, including those with an observational design, feedback was provided to the healthcare professional. This becomes even more relevant when it is one of the aspects of the results highlighted by the authors. Was the provision of feedback an inclusion criterion?

We did not include the ‘provision of feedback’ as an inclusion criterion. All studies included somehow had such a component (some of them inexplicitly with no N provided).
Results:

1. RCTs, CBAs, survey, and clinical audits provide different levels of evidence regarding an intervention; although the authors acknowledge this, even classifying the studies according to the strength of their design, it does not seem to be mentioned when presenting the results, i.e., they present the level of evidence regardless of study design. This is particularly relevant when considering that one third of all the included studies are observational. Related with this, the authors mention at the top of their results section that the present review identified more studies than previous reviews – would the broad inclusion criteria explain this?

We agree with the reviewer’s observation that the increase in number of studies identified is mostly due to the broad inclusion criteria (i.e. observational studies) and made it clear in our previous draft.

2. Complex interventions have a myriad of components, which often are different from study to study; the impact of routine collection of PROMs can only be ascertained when these are specifically taken into account. The authors properly acknowledge the need for “a comprehensive theoretical model and framework”, and cite Greenhalg and colleagues’ contribution to the topic, but then fail to address the detailed aspects highlighted, namely the intervention’s design and implementation, the interaction between proximal and distal outcomes, and the outcome criteria considered by the studies included. For instance, previous systematic reviews identified already some aspects underlying the successful implementation of PROMs in clinical practice, including specific training and the timing and structure of feedback, amongst others [2, 3]. In this sense, the current manuscript seems to fall short of identifying the components underlying the successful use of PROMs.

We agree with the reviewer that the current review by no means has identified and examined all possible important theories/mechanisms that could possibly explain such complex intervention and its underlying ‘causal theories’ (to borrow a realist synthesis term). We also agree with another reviewer comment that our systematic review approach adopted in this current review may be limited in summarising many possible complex theories. Another approach such as realist synthesis may be helpful in shedding light. We made following addition to the discussion as the limitation of the current review:

“Fourth, our study follows a systematic review approach with inclusion of both experimental trials and quantitative observational studies. However, we did not include qualitative studies in our review which may provide addition insights into the questions raised. This is particularly relevant with respect to question 2 and 3 as there was very little quantitative evidence from the included studies. It is important to note that despite our efforts in formulating the review endpoints based on solid and well-established causal and theoretical frameworks which provides insights into not only if but also how the introduction of PROs affects the patient outcomes, the causal mechanisms and process endpoints included in the current review are by no means to be exhaustive and there may be other important causal mechanisms that could be benefited from a realist review approach.” (Page 18).
Discussion

1. Could the authors please explain the sentence “Overall, there is reasonable evidence in favouring the hypothesis that implementing a routine collected PROs system brings positive changes to patient management in the settings where a patient management plan is an integral part of routine collected PROs.”

We rewrote the sentence as: “...is integrated with a routine collection of PROs” (Page 17).

2. The discussion would benefit from a tighter integration of the results obtained by the previous review with those obtained by previous reviews. In this way, the authors summarise the findings for previous work, but do not seem to integrate both sets of results.

As we clarified above, no review has had the same aims, study scope, review endpoints, methodology and framework like our study, we have preferred to present the three other reviews’ results in Table A (Appendix 1) for those who may be interested in their findings.

3. It could also be informative to refer to the role of PROMs within the National Health Service in the UK, and how this world-leading initiative might provide information for the routine collection of PROMs. Information has already been collected for a few selected elective surgeries, but other conditions are also being tackled, namely mastectomy and breast cancer, amongst others.

We added the following discussion in the Introduction section.

“For example, the National Health Service (NHS) of the UK has been implementing a world-leading initiative of the routine collection of PROMs that firstly included a few selected elective surgeries (e.g., unilateral hip replacements, unilateral knee replacements, groin hernia surgery or varicose vein surgery) but are soon expanding to many other conditions such as mastectomy and breast cancer, amongst others. In the USA, the Patient-Reported Outcomes Measurement Information System (PROMIS), a National Institutes of Health funded initiative starting in 2004, is providing a publicly available Web-based resource that can be used to measure key health symptoms and HRQOL.” (Page 4).

Minor essential revisions:

1. How come there are two different search strategies? Under which conditions were they used, taking into account only one database was searched?

The search items are identical, with the second one restricted to title only and the first one restricted to title, abstract and key words. The second one serves a more thorough examination of title and abstract and if necessary, the full-text article.

2. “Patient satisfaction” and “unmet need” address a considerably different
construct, when compared with the previous terms used in the same search string; I would like to know the authors’ rationale for including those (and not others), namely when patient satisfaction and needs are more related with PREMs, rather than PROMs.

There is no universal agreement in terms of what exact definitions for “PROMs” are. There are narrowly defined ones and broad defined ones such as FDA’s definition as outlined in the outset of the paper. FDA’s definition is any report coming directly from patients about a health condition and its treatment. From such definition, we adopted a broad view about the measures of PROs. In that sense, some (if not all) of the Patient Reported Experience Measures (PERMs) such as poor pain management, severe side effects, or ‘unsatisfaction with the care provided’ are a subset of PROs. That said, we acknowledge that there may be a need for further consolidation of different terms and jargons used in this area.

3. What were the simplified terms used for grey (also spelled as gray in the manuscript) literature?

We have used “Patient reported outcome(s)”, “PROMs”, “PROs”, “Quality of life”, “QOL” as the search terms for grey literature. (Page 7)

4. How were “leading researchers and experts” identified?

We have discussed with the leading experts from CINSW regarding the potential experts and leading researchers. We also examined the recent review articles and editorials. We further compared the number of articles published in the related field and citations generated after 2000 (Page 7).

5. Could the authors please provide a definition of “composite PRO”?

In current paper, we define ‘composite PRO’ as those PROs which are often based on a well-developed instrument and aimed for measuring a substantial aspect of patient conditions or treatment. We exclude any PROs with only one or very few items (1-3 items). (Page 8)

6. Was all data extracted independently by two reviewers? If so, what was the agreement rate?

All data extracted independently by two reviewers with an original agreement rate of 86%.

Results:

If the date cut off point for inclusion was >1999, shouldn’t Trowbridge et al (1997) study have been excluded?

We added the following: “As Trowbridge et al. (1997) was the only article in the 1990s included in two of the previous reviews, it was listed in the summation tables for the purpose of comparison. (Page 11)
Discretionary revisions

Some sentences seem to be missing a verb or idea (e.g., last sentence in the results section of the Abstract; sentence in the Introduction that starts with “Some rationales”)

The sentence beginning with ”Some rationales” reads fine. We have gone through the paper again to make sure of its readability.

Methods

1. Shouldn’t step 2 be prior to step 1?

We believe that it is in the correct order as Step 2 reviews the overall 200 reviews in total from Step 1. However, we have added more detail to Step 1 to make it clearer about how the search for reviews and original articles were conducted, respectively. (Page 6, 7)

2. The table authors refer to as 2 is listed as 1.

We have made the suggested correction. (Page 8,9)

3. The papers published by Velikova and colleagues report on the same study as the 2004 paper.

We were aware of that and clarified it in the previous draft as “Two recently published studies were the continuation of an earlier study published by Velikova et al. (2004)” . (Paragraph 2, Page 12)

Reviewer: Joanne Greenhalgh

Minor essential revisions:

The reference given to Figure 1 is incorrect - it should be Greenhalgh et al 2005 not Greenhalgh 2009.

The suggested correction has been made.

Discretionary revisions

This is a traditional systematic review of the evidence of the effectiveness of PROs with two differences: (1) the authors limited their review to studies of PROs in adult cancer patients and (2) the authors included other quantitative study designs, not just RCTs. The authors also consider a wide range of potential impacts of PROs. However, the underpinning methodology is that of a systematic review. In my view, this methodology is most useful for answering the author's first research question, but less useful in answering questions 2 and 3.
Systematic reviews can usually tell us whether something works or not, but they are much less useful in telling us who the intervention works for or why and how it works. So while I feel the authors have done a systematic review well in and of itself, I feel that it only really addresses one of their research questions and less so the other two. To fully answer those, I feel a different review methodology is needed and my preferred methodology would be a realist synthesis. However - this is my opinion and matter for debate, but important to state my position so that the authors and journal readers can see where I am coming from when I make the following comments.

I feel the manuscript could be improved if the authors provided clarification or further discussion of the following:

(1) The authors list their research questions in the last para of the introduction. The review methodology chosen is most useful in answering Question 1 but less useful and less successful in answering Question 2 and 3. Did they consider using other review methodologies to answer their research questions? Why did they choose a traditional systematic review methodology.

We appreciate the reviewer’s comments. The reason that we have adopted a traditional systematic review instead of a realist review approach is that: 1) an up-to-date systematic review with a transparent and systematic methodology can be used to compare with the previous related systematic reviews and serve for the baseline for the future assessment of the evidence-base in the related area; 2) despite the promise offered by a realist review for the research questions raised here, the lack of well-accepted exact guidelines on how to weight evidence from different types of studies are still a challenge for the topics covered in the current review; 3) we have reviewed a much broader literature and adopted the “best possible theories” or ‘causal mechanisms’ as our systematic review endpoints. We believe that such an in-depth review of the possible causal mechanisms is exactly in line with the tenets of a realist review but with added rigor of a systematic review; 4) we have adopted the same methodology for Q2 and Q3 for the sake of consistency; 5) we agree with the reviewers’ view that a totally different review approach such as a realist review may shed new light on Q2 and Q3.

(2) The authors state their inclusion criteria in the section on ‘inclusion and exclusion’ criteria. The first criteria appears to be very broad and could also potentially include qualitative studies evaluating clinicians or patients views of the impact of PROS in practice. However, from the studies included it appears to have been interpreted rather more narrowly as quantitative studies evaluating the impact of PROs on a range of outcomes. Could the authors clarify this? If this is the case, it would suggest that they might have had more difficulty in using the included studies to answer their second research question about potential mechanisms, which are most often to be found in qualitative studies.

We would like to confirm that we have excluded qualitative studies from our review (page 9). We agree with the reviewer that there was little evidence from the quantitative studies that we have included in the current review for the Q2 and Q3. We also agree with the reviewer’s view that including qualitative studies may shed more light on these.
(3) Why did they limit their review to adult cancer patients? Some very interesting work on the use of PROs in children has been conducted in the Netherlands and could have usefully added to the review.

We did not include children in the current review as we chose not to cover all things for every possible population given the very complex nature of assessing the impact of PROs. The types, settings and related PROs and its instrument for child cancers could be quite different from those in adult cancers. We felt that child cancers require a specific study and therefore excluded them from the current review.

(4) The grading used for Domain 2 (described in para 3 of the section 'data extraction and quality assessment) matches very closely a traditional hierarchy of evidence for evaluating the effectiveness of interventions and prioritises quantitative studies. As such, it fits most closely with answering the first research question. However, it is much less useful when evaluating qualitative studies on the mechanisms through which PROs might change clinical practice. Can the authors comment on this.

We agree with the reviewer’s comments. Despite the fact that including only quantitative studies in a systematic review, (with a structured approach in assessing the quality of the study and its potential weight in deriving the final conclusions), may provide us with some consistent, repeatable ways in assessing the generalisability and adoptability of conclusions (drawn from a review of a complex intervention), it is inevitable that it will be: 1) more focused on limited settings (through avoiding heterogeneity by design such as excluding child cancer patients in the current review); and 2) less dynamic and insightful from the exclusion of qualitative studies and lack of ‘theory-generating’ possibility offered through a realist review approach. In our humble view, even a realist review approach were adopted, it also needs more standardised conceptual and operational guidelines and frameworks to overcome some ambiguities in the methodology and its execution in order to avoid possible propensity of reviewer-dependence (or lack of repeatability), and inability to derive quantitative end-points for potential cost-effectiveness analysis, etc.

(5) Overall, the authors have been able to answer the first of their research questions but have been less successful in answering Question 2 and 3. It would be helpful to provide more detail of how their synthesis of the data enabled them to explore the mechanisms through which PROs change clinical practice and what factors influence this. Indeed - how did they synthesise the data? What led them to make the statements made the last para of the results?

We have drawn these inferences from all the references that we have come across (over 200+ full-text articles and reports plus many other web sources). Also, for the 27 studies included, we also reviewed all possible mediation effects (such as a path-analysis and mediation-analysis through multiple, staged regression approach) or moderating effect (through explicitly testing the interaction effect/moderating effect, or inexplicitly through subgroup analysis). We indicated these results (moderating and subgroup effects) in Appendix 4 in the original manuscript. Overall, there is a lack of quantitative studies explicitly
addressing Q2 and Q3 so our inferences are less certain and fall to a more-or-less narrative review fashion. We would be interested in seeing if a different approach such as realist review may shed more light.

In summary: we acknowledge the reviewer’s insight regarding potential value of using an alternative approach to include qualitative studies in reviewing questions 2 & 3. We made the following addition to the Discussion to acknowledge this as a limitation of the current review and encourage future studies to explore this issue further:

“Fourth, our study follows a systematic review approach with inclusion of both experimental trials and quantitative observational studies. However, we did not include qualitative studies in our review which may provide addition insights into the questions raised. This is particularly relevant with respect to questions 2 and 3 as there was very little quantitative evidence from the included studies. It is important to note that despite our efforts in formulating the review endpoints based on solid and well-established causal and theoretical frameworks which provides insights into not only if but also how the introduction of PROs affects the patient outcomes, the causal mechanisms and process endpoints included in the current review are by no means to be exhaustive and there may be other important causal mechanisms that could be benefited from a realist review approach.” (Page 18)