Author's response to reviews

Title: Measuring team factors thought to influence the success of Quality Improvement in primary care: a systematic review of instruments

Authors:

Sue E Brennan (sue.brennan@monash.edu)
Marije Bosch (marije.bosch@monash.edu)
Heather Buchan (Heather.Buchan@health.gov.au)
Sally E Green (sally.green@monash.edu)

Version: 3 Date: 7 November 2012

Author's response to reviews: see over
7 November 2012

Dr. Paul Wilson
The Implementation Science Editorial Team
BioMed Central

Dear Dr Wilson,

On behalf of my co-authors, Marije Bosch, Heather Buchan and Sally Green, I am pleased to submit a revised version of our manuscript “Measuring team factors thought to influence the success of Quality Improvement in primary care: a systematic review of instruments”.

We thank the reviewers for their very helpful review of our manuscript. Our revised manuscript incorporates the reviewers’ suggestions (highlighted using track changes). We provide additional responses to the reviewers’ comments below with cross referencing to changes in the manuscript.

We look forward to hearing from you.

Yours sincerely,

Sue Brennan
Reviewer 1: Miranda Laurant

General comment regarding reporting of the measurement review in two companion manuscripts
Practically, it was not possible to report the entire set of 161 instruments in a single paper. Separating the team measures is further justified because there is a distinct body of theory and empirical research focussed on work groups and teams. Given the coherency of this literature, and the measures arising from it, we felt that this provided a natural split for reporting the review. While the relationship between levels is important, the focus of our review is on describing instruments. The relationships between levels are considered in the paper describing the conceptual framework.

General comment regarding guidance on choosing instruments
We have not specified which instruments are best in any given circumstance. Each study will have unique requirements that depend on the hypotheses to be tested, the setting and conditions in which the instrument is to be used, and acceptable respondent burden given total demands on participants. The tables and text are intended to guide researchers in identifying instruments most suitable for their circumstances (e.g., by enabling comparison of content across instruments; highlighting comprehensive measures by construct; providing information on purpose and context for which the instrument was developed, the number of items, and available evidence on their measurement properties).

We agree that the length and acceptability of self-report measures is an important consideration, with the potential to introduce bias through non-response and missing values.

In response to the reviewer’s comment, we have added a paragraph to the discussion to:

- identify practical issues that need to be considered when selecting instruments, and
- highlight issues related to response rate and instrument length (page 28, paragraph 1).

Minor essential revisions

1) Page 6: There is no reference to the initial framework. It would be very helpful to understand on which work this initial framework is based.

We have added a brief description of the basis for the framework (page 5, paragraph 4)

2) On page 6 the term “measurement review” is used, but it is not clear what is meant. Perhaps you used this term as you were looking at measurement instruments which have been used to measure influencing team-level factors. The term ‘measurement’ is redundant, you are performing a review, without meta-analysis.

We used the term ‘measurement review’ as shorthand for ‘review of measurement instruments’ (p6) to indicate the type of systematic review we were undertaking, and to distinguish it from our review of theory (referred to in relation to development of our framework) and from other types of systematic reviews (e.g. a review of observational studies examining the association between context and CQI outcomes).

We think the distinction is important, so have made the following changes to clarify our description:
- Replaced the term ‘measurement review’ with ‘review of instruments’ (page 5, paragraphs 3; page 6 paragraph 2)

3) Page 7: although figure 4 reports the inclusion criteria for stage 2, it would also be very helpful if criteria for full text screening (stage 1) is mentioned. The inclusion criteria stage 1 and inclusion criteria stage 2 are preferable not only reported in figure 4, but also in main text as this is important information to understand the results.
The inclusion criteria applied when screening abstracts and full text papers were the criteria for inclusion of instruments for the first stage of data extraction and analysis (described as Stage 2 of the review).

We have made the following changes to clarify this and address a similar comment from Reviewer 2 (comment 8):
- We have reworded the figure slightly to refer to ‘initial inclusion criteria’ rather than ‘inclusion criteria for Stage 2’ (Figure 3).
- We have added a sentence regarding the inclusion criteria to the main text (page 8, end of paragraph 3)

3) Page 11: Please include the number of identified papers (n=6296), and refer to figure 4.

We have added a sentence reporting the number of non-duplicate references, and referred to Figure 4 (page 12, paragraph 2)

4) Page 12: it would be very helpful if the number of papers and instruments is reported. E.g. Teamwork context was measured in xx papers, including xx unique instruments. Similar for other domains, and perhaps also for the different constructs, eg. Comprehensive measures of enabling conditions was measured in xx papers.

We have added the number of instruments for each of the domains (page 13, paragraph 2), but have not done so for individual constructs. This information is available in Tables 3 to 5 and we were concerned that including it in the text would add substantially to the length of manuscript, without adding information that aids the selection of an instrument. The number of studies is summarised for the review as a whole in Figure 4 (in line with PRISMA reporting guidelines). For individual instruments, we reference all reports in relevant contexts (Additional files 4, 5 and 6), which is important for understanding the body of evidence supporting the measurement properties of an instrument. We do not summarise number of papers by construct because the unit of analysis for this review was the instrument (rather than the paper or study) – each new instrument added to our understanding of a construct, but secondary reports of an instrument typically added few, if any, additional insights.

5) Page 28/29: Limitation of the study: Only 1 person included the papers and extracted the data. Only 10% of all papers included in stage 4 were extracted by a second person (page 8). This is a severe limitation of study design. In the discussion (page 29) you concluded that “this is first application of our taxonomy, refinement is likely. .... Alternative categorizations are possible”. Given this limitation (e.g. why not extraction by at least two persons, consensus in broader group) how should I value the outcome of this study. Is it very likely that constructs are different categorized or that categories within constructs should be further refined? Perhaps you can strengthen your conclusion on the value of the taxonomy, refer to 10% double extracted papers, was there good inter-rater score, or did both vary within extraction, and what were (and perhaps still are) the main ‘discussions’ regarding which domain, construct and/or categories.

We thank the reviewer for their suggestions about how to strengthen our conclusions, and list the changes made to do so at the end of the response to this comment. While it is true that one person screened papers, and that there was double data extraction for 10% of the final set of papers (Stage 4), we do not consider that this is a severe limitation in a review of this type for the reasons that follow.

The following factors (discussed in order of various phases of the review process) influenced our decision to use one screener and data extractor for the majority of studies:

**Screening**
- Instruments with strong evidence of their measurement properties derive this evidence from multiple studies. This lessens the chance that an important instrument will be missed in screening. Reference list checking – particularly of systematic reviews – provided an additional check to ensure we avoided missing potentially important measures.
Data extraction

Our data extraction methods were applied to 30 papers across our two measurement reviews. The majority of extracted data was qualitative, there were few discrepancies between the two data extractors across this data set, and these were mainly due to initial differences in interpretation of the data extraction guidance, which were resolved through discussion and refining the guidance.

Minor discrepancies in data extraction are unlikely to have important implications for researchers selecting instruments. Selection of instruments for a particular purpose involves weighing up the relative strengths of each instrument based on the combination of the information provided in the review (e.g., content, number of items, assessments of a different of measurement properties based on multiple tests, and the extent to which this evidence derives from relevant contexts). Much of this information is qualitative. This is a very different circumstance from a review of effectiveness, where errors in data extraction influence estimation of the overall effect estimate and can lead to erroneous conclusions about the effect of an intervention.

Categorisation of content:

A second person independently categorized the content of a sub-sample of ten instruments (12%), including all instruments where there was any uncertainty over categorisation. The categorisation was discussed to facilitate revision of the taxonomy and confirm final categorisation. There were few discrepancies in our categorisation. These were at the lowest level of categorisation, and resulted in addition of factors rather than re-categorisation to an existing category, revision of higher level categories, or changes to the structure of the taxonomy.

We have made the following changes to the manuscript to address this comment:

- Added a statement that checks of the data extraction method were performed in both reviews, and supported use of a single data extractor (methods page 9, paragraph 2; limitations page 30, paragraph 2).
- Added description about the process used to independently check categorisation and refine the taxonomy (page 11, paragraph 1; previously only reported under ‘limitations’) and the nature of resulting changes to the taxonomy (end of page 12)

6) Page 25: Limitation of the study: Only self-report instruments are included in review, while at this page you report that actual behavior can best be measured by observation. To my opinion this is something you could have figured out at the start of the study. I don’t read anything about reasons to limit the review to self-reported instruments, in other words why are other methods, such as observations and interviews, excluded from this review. Furthermore, this restriction should be reported as limitation of this study. Observations, interviews might also be valid to measure team level factors thought to influence CQI-outcomes and could contribute to the taxonomy and guide researchers in their choice of appropriate instruments.

Our review was limited to self-report instruments at the outset (Figure 3 inclusion criteria) – we should have stated this in the aims. We agree that other measurement methods are important, but they require substantial resources, making them expensive and limiting their feasibility for use in large evaluations. They are therefore outside the scope of this review.

We have made the following changes to address this comment:

- Specified in the aims that the review was limited to ‘self-report’ instruments (page 6, paragraph 2)
- Included a rationale for limiting the review to self-report instruments (page 6, end of Paragraph 2)
- We have not included this as limitation of the review. The focus on self-report instruments reflects the original purpose and scope of the review, rather than a limitation of the review methods.
Lay-out/textual issues

1) Page 6: Abbreviation of InQuIRe misses a underline “Quality”
This has been corrected.

2) Page 8: The InQuIRe framework (figure 1) provided.... In previous sentences figure 1 was referred to as initial framework. Please be clear is figure 1 the final framework, or initial framework.
We have added ‘initial version’ to the figure caption.

3) Page 11: Observational designs were most commonly reported in other papers (n=214). Delete ‘other’.
We have deleted ‘other’.

4) Page 24, line 6: Please rephrase “In combination with the already limited evidence for the measurement properties of some instruments, this means that additional testing of instruments in relevant context is required”. And please also report what kind of testing you mean, do you mean to test the instruments validity and reliability or do you mean that instruments should be tested whether or not these constructs the instruments represent/test do influence CQI outcomes.
We have rephrased this sentence and specified that we mean by testing of the instrument’s measurement properties (page 25, paragraph 1).

5) Page 24, last sentence: this sentence seems out of place here.
We have restructured the end of this paragraph to make the links between sentences clearer (page 27, paragraph 1)

6) Table 1, page 44: Example content validity seems missing
Thank you for identifying this omission. We have added the missing text. (Table 2, page 48)

7) Table 3: for readability, consider # or other symbol instead of X (bold X) and X*. Reader a much better and quicker overview of the table. Similar for tables 4 and 5
We have made this change.

8) Table 6: for readability there should be a clear distinctive line between ‘other healthcare’ and ‘non-healthcare’
We have increased the bolding to try to make the line more visible. The bolding appears to be poorly reproduced in the automatically generated pdf, so we will check to ensure this is clear in the final version.

9) Figure 2: is this figure correct? “Team composition & structure”“ is a category with different elements like Size, Tenure, etc. but Competencies of team members, again with different elements as sublevel, seems to be an element (font size), but is also called category. Please check.
Figure 2 is correct. ‘Category’ is a term we used for any group of constructs, irrespective of level. We do this to avoid using multiple labels (e.g. category, subcategories, dimensions, elements).

We have made some changes to the description of ‘construct’ and ‘category’ in Figure 2 to try to make this clearer.
10) Additional file 2: only report search PsychInfo, please also include search terms used Medline and HaPI.

We have added the search strategy for PsycINFO and HaPI to Additional file 2.

11) Additional file 4 and 6: add (Stage 4)

We have added ‘(Stage 4)’ to the list of additional files and to the table titles in the Additional files.

Reviewer 2: Sara J Singer

Major compulsory revisions

1. The authors should be aware of, make reference to, and distinguish this review from another excellent recent review of teamwork instruments: Valentine, M. A., Nembhard, I. M., & Edmondson, A. C. (2011). Measuring Teamwork in Health Care Settings: A Review of Survey Instruments. Harvard Business School Working Paper #11-116. I believe that both make useful contributions. However, since the Valentine et al. article is already in the public domain, this paper should acknowledge it.

We have added a section in the background to reference this recent review and an earlier review of teamwork instruments (the latter, previously identified in our results section only), and describe how our review differs (page 7, paragraph 1).

2. Abstract

While I recognize that it is fairly standard to begin paper by identifying the gap in the literature you will fill through your paper, it seems a little disingenuous to say, “Teamwork is an important element of Continuous Quality Improvement (CQI), yet we know little about how team-level factors influence CQI outcomes.” In addition, the gap this paper fills is about measuring team-level factors, not about their relationship to outcomes. I suggest you emphasize the importance of this link and establish the need for measurement tools in order to evaluate it.

We thank the reviewer for highlighting that the abstract background didn’t appropriately reflect the purpose of the paper. We have revised the abstract background accordingly.

Background

3. I understand from your title and methods that you narrowed the scope of this study by focusing on CQI in primary care. Your introduction should explain why this focus was appropriate. As written, I don’t believe you addressed primary care at all.

We have added background to explain our focus on primary care (page 7, paragraph 1).

4. p.28 Limitations. The first sentence should be modified to reflect the preceding article by Valentine et al.

We have modified this sentence (page 29, paragraph 3)

5. p.28 bottom. You should also add a limitation acknowledging that you did not pursue further instruments from articles that did not include the survey items—an apparently large number of papers.

We have added this to the limitation of the review (page 30, paragraph 2)

Conclusions
6. This section should again emphasize the importance of teamwork for CQI in primary care in particular.
We have edited the conclusion to emphasise this point (page 31).

**Discretionary Revisions**

**Background**
7. p.4 The second paragraph provides a more compelling beginning to this paper than the first. I’m not sure you need the first paragraph at all. If you feel otherwise, perhaps you could weave some of its content into the second paragraph.
We have removed the paragraph and inserted the main references elsewhere.

**Methods**
8. p.8 top. “Reference lists of identified systematic reviews were screened. Snowballing techniques (citation searches, reference lists) were used to trace the development and use of instruments included in Stage 4.”

Since you haven’t introduced Stage 4 at this point, please provide a little context for this reference. Also, you write, “The full text of potentially relevant studies was retrieved and screened for inclusion by one author (SB).” On what basis were articles screened for inclusion in Stage 1?

We have changed ‘Stage 4’ to ‘all stages of the review’ (page 9, paragraph 2).

Both reviewers asked for clarification of Stage 1 inclusion criteria (see also, comment 3 from Reviewer 1). Stage 1 was the searching and initial screening stage, which determined inclusion of instruments in stage 2 (the first stage of data extraction and analysis).

To make this clearer:
- We have reworded the text in Figure 3 to refer to ‘initial inclusion criteria’ rather than ‘inclusion criteria for Stage 2’ (Figure 3).
- We have shifted the section on data extraction so that is under the Stage 2 heading (page 9, paragraph 2)

9. p.8 Data Extraction. I realize that you refer the reader to Table 1 to see the list of data extracted, but it would be helpful if you characterized here what you considered “data.”

We have added a sentence to characterize the data (page 9, paragraph 2).
We have done the same for data extracted at Stage four (page 11, end of paragraph 2).

10. p.9 Stage 3. You write: “For constructs not adequately covered by suitable instruments, we included instruments with potential for adaptation.” Please explain what you mean by “potential for adaptation.” For example, does this mean you borrowed instruments from other industries or ones that did not meet your initial screening criteria? Do you convey which are the areas in need of more survey development? This would be very helpful.

We pre-specified in our inclusion criteria that, for constructs where there were few suitable instruments, we would include instruments that could be adapted (Figure 3). In the absence of suitable instruments, we think the first option should be to adapt and test existing high quality instruments, rather than develop new ones. In the results text, we identify instruments that were included on this basis, and indicate the nature of development work required for adaptation (e.g. page 16, paragraph 2). We also summarise the main adaptations required on a per instrument basis in Additional file 5 (in which we tabulate all instruments requiring adaptation).
Changes made to address this comment:

- We have cross referenced from the paragraph to Figure 3, which reports sources and inclusion criteria for each stage.
- We have added a sentence to provide an example of what we mean by ‘potential for adaptation’ (page 10, paragraph 3).

11. p.11 bottom. *It may be hard to imagine how teamwork instruments might not be relevant to a primary care setting. Please expand.*

We have added an example to illustrate (page 10, paragraph 1).

**Results**

12. *Missing for me from each section is a brief statement about why each of these domains and constructs matter to CQI in primary care. Perhaps if this were done well up front I would not feel its absence so acutely, but as written I feel like these results are insufficiently grounded to the objective of your effort.*

We have added a paragraph to the background providing our rationale for focusing on CQI teams in primary care (page 7, paragraph 1 - addressing major essential revision 3). This paragraph includes a brief overview of why certain factors may be particularly important in primary care. We have also made changes to the paragraph describing the framework (page 7, paragraph 2) to focus more on how the included domains and constructs relate to CQI teams. We have not provided further description for each construct in the results section, primarily because we think the additions to the background prompted by the reviewer’s comments provide an efficient way to address this issue without adding substantially to the length of manuscript. It is our intention to provide a rationale for each domain and construct in a separate paper focused on the development and content of the framework.

13. p.21 *Instrument characteristics, development and measurement properties.*

   *This section would benefit from a brief statement of why you reviewed the development and measurement properties of these instruments—presumably to assist the reader in choosing the most valid and reliable instruments.*

   We have added this statement (page 22, paragraph 3)

**Discussion**

14. p.27 *Ensuring conceptual clarity. An example of two similar items measuring emergent states and team processes that would constitute an intermingling would be helpful.*

   We have added an example (page 28, paragraph 3).