Author's response to reviews

Title: Measuring organisational and individual 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 Green (sally.green@monash.edu)

Version: 2 Date: 10 June 2012

Author's response to reviews: see over
21 May 2012

Dr. Anne Sales
The Implementation Science Editorial Team
BioMed Central

Dear Dr Sales,

On behalf of my co-authors, Marije Bosch, Heather Buchan and Sally Green, I am pleased to submit an amended version of our manuscript “Measuring organisational and individual factors thought to influence the success of Quality Improvement in primary care: a systematic review of instruments” for your consideration in the journal Implementation Science.

We thank the reviewers for their thoughtful review of our manuscript. Our amended manuscript incorporates the reviewers’ suggestions (highlighted using track changes), which we believe have improved the paper. 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 (on behalf of Marije Bosch, Heather Buchan and Sally Green)
General comments in response to concerns about the conceptual framework

This measurement review is one output from a larger project, which aims to develop a conceptual framework and identify a suite of measures for evaluating CQI in primary care. In response to the reviewers’ comments, we have made changes to the background to reflect the broader aims and sequencing of our work (p6). We have separated out those aims that relate directly to the measurement review from those that inform our final framework which will be reported in a separate paper (p6). We think this clarifies the purpose of the manuscript and provides greater clarity about the stage of development of the framework presented in the background.

We do agree that the measurement review would benefit from more explicit reporting of the origins of the conceptual framework, and have used the reviewers’ suggestions to guide our revisions (annotated below in relation to each comment). The development of the framework is a substantive piece of work and draws heavily on the work reported in the measurement review (this paper and the companion paper on team level factors (in preparation)), and so we plan to report it separately.

In response to a suggestion from reviewer 2, we have named the framework. We use the name (InQuIRe) throughout the paper. We have also re-numbered the stages of the review. We now refer to Stage 1-4 in the paper (previously Stages 0-3).

Reviewer 1

Major compulsory revisions
1. In order to support the organization of the conceptual framework, more review of the theories, described in the first paragraph of the Scope of the Review section, is needed in the Background section or prior to the first paragraph in the Scope of the Review.

We have added the following to address this comment (and related comments from Reviewer 2):
- Background: provided description and additional referencing of the main sources used to develop our conceptual framework (references are major examples only because of the extensive referencing required to report included studies) (p7, paragraph 2)
- Background: indicated the models of CQI and practice change that were most influential in informing our framework. (p7, paragraph 2)
- Background: clarified that the reason for development of the framework was the absence of an existing integrated framework for evaluating CQI in primary care (p7, paragraph 1)
- Background: to clarify the scope of literature reviewed in developing our framework, we have stated the basis for considering non-CQI literature and how this literature was used (p7, paragraph 2)

Minor essential revisions
1. In the first paragraph of the Background, third sentence, it is unclear to what “the methods” are referring. Perhaps the sentence should read “these methods?”
We have made this change.

2. In Stage 2, after the sentence, “Construct definitions and labels assigned to scales guided but did not dictate categorisation because labels were highly varied and often not a good indicator of instrument content” add an example of this discrepancy.
We have added an example (page 13, paragraph 2).

3. In Appraisal of Evidence Supporting Measurement Properties, first paragraph, the sentence, “The COSMIN criteria were intended for studies reporting instruments for the measurement of patient reported outcomes, however we were unable to identify equivalent appraisal criteria for organisational measures” should be “…outcomes; however, we…”
We have made this change.

4. In Identification of Unique Instruments, it is unclear what the authors mean by an “index paper” and how it was used.
We have made changes to clarify what we mean by index paper and how it was used. (p16, end of 2nd paragraph)

5. Define acronyms before solely listing them as acronyms (e.g., TQM); throughout manuscript.
We have made this change, and checked for other instances in the manuscript.

6. First paragraph under Measurement of Organisational Context in the Discussion section, in the last sentence should be ‘used’ and ‘be’ should be removed.
We have made this change.

Discretionary Revisions
1. In the third paragraph of the Background, “CQI methods” and “intervention” seem to be used interchangeably (i.e., “Complex interventions such as CQI are not easily replicable…”). If CQI is considered an intervention in an of itself, then this should be explicitly stated, otherwise it is confusing to the reader who might consider CQI its own field of study related to, but distinct from, specific interventions such as a new treatment technique among primary care physicians.
We have included an operational definition of CQI under our description of the content of the framework (p8, end of paragraph 2). This definition specifies the methods that are the core elements of a ‘CQI intervention’.

2. Because of the complexity of the review, phases, and development of the framework, consider using examples throughout the text to help the reader visualize the application of the methodology for a specific measure, particularly in the Taxonomy Development sections and Stage 2 sections.
We have made the following changes to address this comment:
- Methods: provided additional description about the methods used to development our taxonomy to clarify (p 12-13)
- Results: provided examples for the decisions we made at stage 3 (previously stage 2) (p 17, paragraph 2)
3. Under Development of Taxonomy and Categorisation of Instrument Content, second paragraph, the authors reported excluding measures based on certain criteria, two of which were if the measure was “too long” or the “response format [was] unsuitable.” It is unclear why or how these judgments were made, or moreover, how it was determined that an instrument characteristic such as length may impact CQI assessments. For example, at what point is a measure ‘too long’ or ‘short enough.’ The authors should consider explaining/justifying their exclusions using the research available.

We thank both reviewers for raising this concern. It has alerted us to the fact that our reporting of our inclusion criteria for this stage of the review was misleading in the text. Decisions to include instruments in stage 2 were based on item content (originally reported only in Figure 2 (now Figure 3)). All instruments with item content suitable for primary care were included (feasibility, length etc were not a consideration). For constructs that were inadequately covered by suitable instruments, we included instruments if they could be adapted for primary care. It was this set of instruments where we considered feasibility in making our decision about the potential use in primary care. Prompted by the reviewer’s concerns, we went back to our list of excluded instruments to identify the breakdown of reasons for exclusion. We have documented these in the results section, with examples, to provide greater clarity.

To address the reviewers’ concerns we have made the following changes:
- Results: added examples to illustrate our application of the exclusion criteria (p17, paragraph 2)
- Results: specified the number of instruments excluded based on each of the criteria to allay concerns that a large number of instruments were excluded based on ‘length’. (p17, paragraph 2)

Reviewer 2

Major compulsory revisions

1a) First, it appears that this framework was both proposed for and modified by the current study. If this is not the case, please clarify.

This is the case. The development and final content of the framework will be reported in a separate paper. The taxonomy (tables 3-5) reports the modifications to the framework arising from the analysis in this measurement review. The final version of the InQuIRe framework will draw together the analysis from this paper and the companion paper reporting team measures (in preparation), and our initial synthesis of CQI theory (to develop the version of InQuIRe used to set the scope of the measurement review).

We have made the following changes to make our approach and rationale clearer:
- Background: we have described the relationship between the taxonomy and framework, and specified that the modifications reflected in the taxonomy will be incorporated in final InQuIRe framework (page 6, paragraph 1).
- Methods: we have provided a rationale for refining the initial taxonomy based on the content analysis in the measurement review (page 12, paragraph 1).
- Methods: we have amended the description of the taxonomy development to clarify the methods underpinning our approach (page 12-14; see response to comment 2)

1b) Second, how were models of QI, teamwork theory, and preliminary findings merged to create a framework that depicts temporal relations between factors when the “known relationships between variables” and the “magnitude of effect of these variables... is limited” (see first paragraph entitled “Scope of the review-conceptual framework”)?

Our framework is a synthesis of CQI theory, developed to support the design of evaluations and studies testing these theories, not a synthesis of empirical evidence intended to inform practice. The temporal relations depicted in the framework are intended to reflect reported theories about the relation between antecedents, CQI process and outcomes.

We have made the following changes to the manuscript to address the reviewers’ question:
- Background: we have emphasised that our framework (as presented in this paper) reflects a preliminary synthesis of theory, intended to support the design of evaluations and studies testing these theories, not a synthesis of empirical evidence intended to inform practice (p7 paragraph 1 and 2).
- Background: included a description of the methods use to develop the framework (p7 paragraph 2)
- Background: included a paragraph describing the basis for depiction of the temporal relations, and our rationale for choosing this approach (p8 Paragraph 1).
- Background: stated our intention to report the methods and content in full in a separate paper (p6, paragraph 1)

1c) Third, why were previously established frameworks such as the Consolidated Framework for Implementation Research (CFIR; Damschroder et al., 2009) not considered? Similarly, the framework includes factors previously identify by Proctor and colleagues (2010) in their Implementation Outcomes paper (e.g., effectiveness, safety, equity, timeliness, efficiency, cost) or originally articulated by the Institute of Medicine, but the authors do not even refer to these works. At a minimum, it is this reviewer’s opinion that the presentation of this new framework would be strengthened if existing theoretically or empirically derived frameworks were critiqued and an explanation provided for why they were not used.

We agree with the reviewers’ comments that the description provided is limited. We intend to report the rationale, methods and content of the final framework in detail in a separate paper. We will critique existing frameworks in that paper.

While the frameworks listed by the reviewer are important, they are not specific to CQI. The primary source for our conceptual framework was CQI theory. Secondary sources - such as models of organisational change and diffusion of innovation, were used to ensure we have not missed obviously important factors that might pertain to CQI in primary care and to aid definition of constructs.

Our review was in progress when the sources listed were published, so they were not used as secondary sources for the version of framework for the measurement review. However, we are confident that neither the CFIR model nor Proctor’s taxonomy would alter the scope of our measurement review (also see response
to 1d). In developing our framework, we considered the factors in Greenhalgh’s model with which the CFIR model has considerable overlap. Proctor’s outcome taxonomy presents a valuable synthesis of frameworks we have cited in the discussion in relation to intervention fidelity (e.g. the RE-AIM framework). However, it is generic so doesn’t draw out outcomes specific to CQI (for example, dimensions of CQI intervention fidelity).

We have made the following changes to address the reviewers concerns and questions:
- Background: provided description and additional referencing of the main sources used to develop our framework (references are major examples only) (p7, paragraph 2)
- Background: stated the purpose for which we used non-CQI literature to help clarify the scope of literature reviewed in developing our framework (p7, paragraph 2)
- Background: indicated the more prominent models of CQI and practice change used to inform our framework (p7, paragraph 2), and stated that there is currently no integrated theoretical framework for evaluating CQI in primary care (p7, paragraph 1).
- Background: included a footnote in our conceptual framework referencing the IOM as a source of quality dimensions listed under distal outcomes (figure 1).

1d) Fourth, although the conceptual framework provides decent coverage of potential factors implicated in CQI, why were others omitted? For instance, provider attitudes, planning, and compatibility have been identified by other frameworks as factors influencing implementation and likely CQI efforts. More description of the decision regarding how potential factors were and were not included would be helpful.

We have added a description of our approach to developing the framework. Importantly, we suspect that some of the specific decisions we made about inclusion or labelling of factors in our initial framework were unlikely to have an important impact on the findings of the measurement review. This is because omission of potentially important factors from the initial framework did not mean measures of these factors were excluded from the review. We knew at the outset that we were likely to identify new factors, conceptualisations and labelling of factors because the literature on CQI is vast. This drove our decision to use a broad search and to use the findings of the review to develop and refine our framework and taxonomy. It also determined our inclusion criteria for stage 2 (previously stage 1), which was to include measures of any factor identified by authors as a potential determinant of QI outcomes, irrespective of whether it was pre-specified in our framework (figure 3, stage 2 inclusion criteria).

Changes made to address the comment:
- Background: we have provided more detailed description of the basis for developing our framework as described under 1c (p7, paragraph 2)
- Methods: we have added a sentence to indicate the inclusive approach we took to considering factors in the review (p12, paragraph 2)

2a. There is a cyclical process used to identify the taxonomy, which is then used to categorize the instrument content. It might be more appropriate to identify the taxonomy using previous research, review articles, theory, etc and then categorize the instruments within the taxonomy.
Reflecting on the reviewer’s comments, we think our description of the taxonomy development was unclear in the original manuscript. In particular, we may not have been clear that the content of our framework provided the content and structure for the first version of our taxonomy. We have revised our description of the methods used to develop the taxonomy to clarify.

We agree that an alternative approach would have been to finalise the taxonomy at the outset of the review. It was our view however that the taxonomy should be based on the framework, but should also reflect the conceptualisation of constructs measured in QI studies. To achieve this, we based our methods on the framework approach for analysing qualitative data [1]. We felt that incorporating the conceptualisations that emerged from the analysis in the measurement review was particularly important because our attempts to develop a comprehensive framework from CQI theory and review articles revealed very little formal theory development. We suspected, and confirmed, that the studies included in the measurement review would contribute additional insight into the theories held by researchers about how CQI worked and the factors that influenced effectiveness. Drawing together the literatures from our theoretical and measurement reviews has enabled us to piece together a comprehensive picture of prevailing theories of CQI, reflected in the factors represented in the taxonomy (tables 3-5).

We have made the following changes to address the reviewer’s comment:
- Background: stated that the broader aim of work was to use the findings from the measurement review to refine our conceptual framework with the goal of ensuring comprehensiveness (p6 paragraph 1)
- Methods: edited our description of the taxonomy development to clarify the steps and methods used, and provide a rationale for our approach (p. 12-13)

2 b. no rationale is provided for why 84 instruments were used to construct the taxonomy to organize the 41 instruments. A brief statement regarding the rationale would be helpful. This is another example, similar to the above-mentioned points, where the author’s ambitious efforts were not adequately described in the manuscript.

We have made the following changes to address this:
- Methods – taxonomy development (stage 2): included the following text “At this stage, we were aiming to capture the breadth of constructs relevant to evaluating CQI. Hence we included all measures of potentially relevant constructs irrespective of whether item content was suitable for primary care.” (p12, paragraph 2)
- Results: we have included text to describe our reasons for including instruments at the next stage (previously reported only in figure 2 (now figure 3)), with examples to illustrate exclusion decisions. Inclusion was based on item content being suitable for primary care, or potential to be adapted. (p17, paragraph 2)

3. In their conclusion, the authors circle back to their aim of providing “guidance to support decisions around which factors to include in evaluations of CQI”. This, however, was not a main (or even small) component of the discussion. Given the plethora of information presented and communicated in this review, the
manuscript may actually be strengthened by removing this aim. It is not clear from the information presented in the manuscript which factors should be included in evaluation of CQI other than those included in their proposed conceptual framework which has the limitations outlined above.

We agree that the primary purpose of this paper is to report on the measurement of factors thought to influence the success of CQI, hence we have removed this aim from the specific aims for this paper, and indicated that it was an aim of our broader project.

We have also altered our conclusion as suggested. We believe however that the review does provide guidance to support decisions around which factors to include in evaluations of CQI. The taxonomy provides a synthesis of the explicit and implicit theories held by researchers about how CQI works and the factors thought to influence its effectiveness. It maps back to the framework, enabling identification of how different factors could be incorporated in a longitudinal evaluation. We think it does provide guidance for researchers on the factors that could be included in evaluations and have included a paragraph in the discussion to address this.

We deliberately avoided privileging some factors over others. The absence of empirical evidence about the relative importance of different factors means there is no compelling basis on which to recommend some factors ahead of others, particularly in the context of primary care. We also believe that researchers will need to make decisions about which factors are salient in the context in which their research is conducted.

We have made the following changes to address this point:
- Background: removing aim “to identify factors” from specific aims for the measurement review (p6, paragraph 2)
- Background: changes to reflect the broader aims and sequencing of our work (p6, paragraph 1)
- Discussion: addition of a paragraph about how the analysis informed the taxonomy presented in the review and, in turn, the framework. Clarification of how the taxonomy provides a guide to factors that could be included in an evaluation. (p33, paragraph 1)

4. The authors at times are vague in their report of methods and results perhaps given the immense task of describing such an elaborate process within the confines of one report. However, there are places in which more information would make the manuscript stronger. For instance, the authors contest that their category development is set within the “broader literature” but they do not provide specific examples of the review articles or conceptual papers used (see paragraph 2 under Stage 1 Taxonomy Development).

In response to the reviewer’s comment, we have restructured the results reported under “coverage by content domain”. Each section now begins with a description of the final domain and includes references to the broader literature that informed our decisions around the final content of the taxonomy. We focus on major examples that illustrate the main categories in the taxonomy. It is our intention to report our analysis of literature used to inform the final framework in a separate paper (as indicated above).

We have made changes to the manuscript to try to increase the clarity of reporting of methods and results as follows:
- Terminology: we have added a description and figure illustrating the terminology in the review. We have also included a glossary (page 9, paragraph 2, figure 2, and additional file 1).
- Methods: we have reported the steps used to develop the taxonomy, so that the text better reflects information provided in figure 2 (addressing comment 2 above)
- Methods: we have reported our rationale for refining the taxonomy (addressing comment 2 above)
- Methods: we have included additional information about the forms of construct validity assessed, and our rationale for use of and additions to COSMIN (addressing comment 5 below)
- Methods: we have shifted our description of extracted data (stages 2 and 4) to tables, for ease of reading and to shorten the text (tables 1 and 2)
- Results: we included clarification and examples of the inclusion decisions made at stage 2 (now stage 3, p17)
- Results: we have restructured the section reporting coverage of instruments (Content and coverage of domains of the InQuIRe framework, pages 18-24). Each of the three domains begins with a description of the domain based on the final taxonomy. Instruments are then reported under subheadings reflecting categories within the domain.

5. It is unclear why the authors used the COSMIN and made modifications (e.g., added criteria to assess level of analysis issues) but did not include evaluation of other, potentially useful measures of validity (e.g., concurrent, convergent, divergent, predictive).

Several factors influenced our decision to use the COSMIN appraisal checklist. To address the reviewer’s question, we have added a brief description of our rationale to the methods section, and more detailed explanation in additional file 2. In brief, our decision was based on i) the strong evidence for the content validity of the checklist, ii) consistency of the content with the Joint Committee Standards for Educational and Psychological Testing [2], indicating the relevance of the checklist’s content beyond the measurement of health outcomes, and iii) the absence of alternative appraisal tools for organisational measures.

We have added a rationale for modifying the COSMIN tool to Additional file 2 (p3). In relation to the forms of validity evidence listed by the reviewer, these are encompassed by the hypothesis testing domain. In this domain, construct validity is assessed based on accumulated evidence of the “degree to which the scores of an instrument are consistent with hypotheses”. This includes testing hypotheses about whether scores converge with related measures, diverge from measures of other constructs, discriminate between groups, and predict relevant outcomes. COSMIN avoids labelling these forms of construct validity, mainly because of the confusion that has arisen from the discrepant use of terminology.

In summary, we have made the following changes in response to the reviewer’s comments:
- Methods: we have specified that data extracted about construct validity includes convergent, discriminant and predictive validity (reflecting our original definition provided in Additional file 1, Table 1) (table 2)
6. The authors state that 68 instruments were excluded from further review because the instrument was not “feasible”. This is a difficult criterion to use particularly when their definition does not operationalize a cutoff. To be clear, when authors state the instrument is “too long” some example of what this means would be helpful to better understand why so many instruments were excluded. Having one author subjectively rate the feasibility of a measure, without a clearly defined and articulated cutoff, runs the risk of reifying opinion rather than communicating objective, empirical support.

Please see response to Reviewer 1, Discretionary revision 3.

7. The authors stress the importance of defining concepts in their discussion, yet I do not see definitions provided for the factors included in their framework. At the minimum it seems important to include definitions for the taxonomy referred to in the report.

We agree that definitions are important. We have restructured the results reported under ‘Content and coverage of domains of the InQuIRe framework’ to include a section describing the content of each domain (for example, ‘Description of the CQI use and implementation domain’ on p19). We include explicit definitions for constructs where there were clearly divergent conceptualisations in the papers we reviewed (for example the delineation between ‘CQI methods’ and climate for CQI is unclear in the literature). In our taxonomy (Tables 3-5), we aimed to provide a level of detail that would enable readers to identify dimensions of constructs. We also used descriptive labels for constructs, aiming for sufficient detail to facilitate understanding of what the construct encompassed. We acknowledge this is not a complete substitute for explicit definitions, although we do think the level of detail reported in the taxonomy provides a guide to the content of available measures (our primary purpose for reporting taxonomy in this review). It is our intention to report definitions for all constructs in our paper reporting the InQuIRe framework. That paper will reflect the taxonomy in this paper, and the one in the companion paper reporting our review of measures of team level factors. Combining the two will provide greater clarity and ease of use for those attempting to use the framework than if we were to report the detailed content of the framework in separate publications.

Changes in response to the reviewer’s comments:
- Results: restructuring to include a section describing the content of each domain (page 19, 21, and 23)
- Results: addition of further definition for each domain (page 19, 21, and 23)

8. Although the authors articulate some limitations, they do not include the application of their newly developed framework (which has not been empirically evaluated) as a limitation. And, if they choose to retain as an aim to provide guidance about which factors to include in evaluations of CQI, the limits of their framework should be reiterated in this particular section.

We have deleted the aim to identify factors.

Minor essential revisions
- To provide the reader with greater clarity, it is recommended that the authors either define or provide examples of CQI and how this is different from implementation of an innovation if indeed it is
conceptualized as such.
We have added this definition (page 8).

2. The authors write in the third paragraph of the Background section that “measuring these factors as intermediate outcomes permits investigation of the mechanism by which CQI works”. It is perhaps more appropriate to state that this “permits investigation of at least one mechanism”.
We have made this change.

3. The authors might replace the phrase “consistent interpretation” used in “Stage 1 Data extraction” with the term inter-rater reliability. In addition, inter-rater reliability kappa should be reported.
Our use of “consistent interpretation” rather than statistical consistency or agreement was deliberate. The vast majority of data we extracted was text (e.g. descriptions of methods used to assess instruments, construct definitions). We did not calculate inter-rater reliability for this data because to do so would have required dichotomising or assigning a score to the outcomes from our data extraction (to calculate Kappa or an ICC). We felt there was no clear basis on which to do this.

Discretionary Revisions
1. It might be helpful to name the proposed conceptual framework
We have adopted the reviewer’s suggestion. To increase clarity, we have used the name InQuIRe for our framework throughout.

2. Tables 1 & 2 might better facilitate consumption if instead of author name/year the instrument name or acronym is included.
Few of the included instruments had names assigned by their developers, and some had multiple names (sometimes attributed by the same authors in different papers). We felt that naming instruments according to their source was the best way to avoid adding to an already confusing literature.

3. In the Method section under Stage 3 Data Extraction, the authors state that they extracted information about the development of the instrument, but they do not indicate the method of extraction.
We extracted all data into a data extraction form in a spreadsheet. During the qualitative analysis, we tabulated information extracted about each instrument’s content to enable comparison across measures and prepared summaries to facilitate iterative development of the taxonomy. We did not use NVivo.

Minor issues not for publication
1. Incomplete sentence in abstract: We have made this change.
2. Long sentences: we have reviewed the manuscript for instances of this and re-written to simplify. This includes a revision to the sentence provided by the reviewer as an example.
3. **Capitalisation of first letter of each word**: No changes made. Our understanding of the IS typology was that we should “capitalize only the first word, and proper nouns”. We will of course make these changes if we have misinterpreted the guidance.

4. **Inconsistent citation format**: We have checked the manuscript and corrected instances where we refer to instruments with [citation number] alone.

5. **Fix the structure of sentence** “Frameworks for specifying and defining these dimensions exist for health behaviour change interventions; encompassing intervention intensity (e.g. duration, frequency), quality of delivery and adherence to protocols [86-87].” We have reworded this sentence.

6. **Use of “relationship” to refer to the “relation” of two constructs**: If we are in error then we are very happy to correct this mistake. Gliner’s text on research methods identifies these terms as synonyms. We checked the Cambridge dictionary of statistics (Everitt, B. S. and A. Skrondal (2010). Cambridge, Cambridge University Press) and the Encyclopedia of biostatistics (Armitage P, Colton T, 2nd ed. Hoboken, N.J.: John Wiley & Sons, 2005.) for a ruling but couldn’t find one. Relationship was more commonly used in all three sources (e.g. “Association: a general term used to describe the relationship between two variables.”)

7. **Typographical error**: “use” corrected to “used”

8. **Typographical error**: “instrument” corrected to “instruments”

9. **TQM acronym to be defined**: first instance of TQM changed to “total quality management (TQM)”

References
