**Author's response to reviews**

**Title:** Quality of Prenatal Care Questionnaire: Instrument development and testing

**Authors:**

Maureen I Heaman (maureen.heaman@umanitoba.ca)
Wendy A Sword (sword@mcmaster.ca)
Noori Akhtar-Danesh (daneshn@mcmaster.ca)
Amanda Bradford (bradfor@mcmaster.ca)
Suzanne Tough (suzanne.tough@albertahealthservices.ca)
Patricia A Janssen (patti.janssen@ubc.ca)
David C Young (david.young@iwk.nshealth.ca)
Dawn A Kingston (dawn.kingston@ualberta.ca)
Eileen K Hutton (huttone@mcmaster.ca)
Michael E Helewa (helewam@cc.umanitoba.ca)

**Version:** 2  
**Date:** 27 February 2014

**Author's response to reviews:** see over
February 27, 2014

Dr. Mechthild M. Gross, Section Editor
Ms. Janelyn Ann Cruz, Editorial Office
BMC Pregnancy and Childbirth

Dear Dr. Gross and Ms. Cruz:

On behalf of our research team, I am re-submitting our revised manuscript entitled, *Quality of Prenatal Care Questionnaire: Instrument development and testing*, for consideration for publication in *BMC Pregnancy and Childbirth* journal as a research article. We would like to thank the reviewers for their thorough reviews and helpful suggestions which have resulted in an improved manuscript. Our responses to the reviewers’ comments are described below.

**Reviewer 1: Katrien Beeckman**

Thank you for your comment that “This is an elaborated, very interesting, clearly described, and transparent instrument development with testing at different levels.”

Changes made in response to major compulsory revisions:

- In response to your question about time to complete the QPCQ and easiness of use, we added the following information to the manuscript: *During the pretesting phase, the mean length of time for women to complete the 111-item version of the QPCQ ranged from 10 to 23 minutes, with a mean of 16 minutes. Women indicated that the QPCQ was easy to complete.* We did not conduct any formal assessment of the time it took women to complete the final 46-item version of the QPCQ. Because the final version of the QPCQ contains less than half the items of the version used in the pilot test, we are assuming it would take less time to complete (approximately 8-10 minutes), and that readers would deduce this as well.

- We have added a statement to the second paragraph of the Discussion section to clarify that the subscales can be examined separately: *In addition to the total QPCQ score, the score for each of the subscales can be examined separately.* The Conclusion section also contains the following statement: *The QPCQ is a new self-report instrument that measures overall quality of prenatal care, and quality of care for six factors or subscales.*

- We have clarified the labelling for the five response categories of the 5-point Likert scale as follows: *When constructing the QPCQ, the research team decided that each item would be rated using a Likert scale with five response categories consisting of “Strongly Disagree” (1), “Disagree” (2), “Neither Agree Nor Disagree” (3), “Agree” (4) and “Strongly Agree” (5). All points on the scale were labeled to prevent the tendency for respondents to endorse labeled points more often when only some are labeled.* Because we purposefully chose items for the QPCQ that we believed were universally applicable to women receiving prenatal care (in a Canadian context), we did not provide a “not applicable” response option. As with any questionnaire, there will always be some respondents who will not answer some of the questions, and researchers will need to follow standard methods of dealing with non-response to items.
- We added the following statement to clarify what was evaluated in the 15 prenatal care guidelines and why: *Items generated from the guideline review reflected components of prenatal care rated as having a high certainty of net benefit (i.e., “A” grade evidence)*.

- A statement has been added about how to interpret the QPCQ results: *The QPCQ is a norm-referenced measure, in which an individual’s score takes on meaning when compared with the scores of others (e.g., in the same sample). Higher scores on the QPCQ and its subscales reflect a higher rating of quality of prenatal care. Because the QPCQ is a norm-referenced measure, and not a criterion-referenced measure, there is no established cutoff score or external criterion against which the quality of care is judged.*

Changes made in response to minor essential revisions:

- A statement has been added to the text and as a footnote to Table 2 that missing data were excluded from analyses.

- The questions we developed were judged to be relevant in a Canadian context. We were not intending to develop items that were globally applicable at this stage of instrument development. The following statement in the Limitations section emphasizes this point: *The QPCQ was developed in the context of the Canadian health care system, so its applicability to health care systems, prenatal care provision, or populations that are substantively different will need to be assessed prior to widespread use. As an aside, we are currently testing the QPCQ in Australia, and made some changes to the instrument to make it more applicable to the Australian health care setting (e.g., the term “prenatal care” was changed to “antenatal care”), in consultation with health care providers in that country.*

**Reviewer 2: Susanne Grylka-Baeschlin**

Thank you for rating this manuscript as “an article of outstanding merit and interest in its field.”

Based on your suggestions, we have extensively re-structured the manuscript. Each of your points is addressed below.

1. In the Background section, we have provided the background information first, followed by the purpose and aims of the study.

2. One of the redundant sentences in the Background section has been removed. We also removed the “Measurement of prenatal care quality” heading and restructured the Background section as requested.

3. The fonts of titles of sections and sub-sections have been revised to conform to BMC guidelines. Some of the titles of sub-sections have been removed from the manuscript to avoid confusion.

4. The paragraph on Design has been incorporated at the beginning of the Methods section, and is no longer labelled “Design”.
5. The content on ethical approval is now part of the Methods section.

6. The section on Methods and Results has been separated into two distinct sections.

7. We have paraphrased sentences instead of citing definitions from textbooks, although the textbooks (many of which are classic resources on measurement theory and instrument development) have been retained as references for the statements.

8. We have now classified Cronbach’s alpha in a consistent manner as “acceptable,” and provided the following description in the manuscript: A Cronbach’s alpha coefficient of at least 0.70 is considered acceptable, while 0.80 or greater is desirable.

9. We have followed the *BMC Pregnancy and Childbirth* Instructions for Authors, which state that figure legends should be included in the main manuscript text file at the end of the document, rather than being a part of the figure file.

10, 11, & 12. The errors in the formatting of tables have been corrected.

13. We corrected the last line in the figure.

**Reviewer 3: Elisabeth Svensson**

We appreciate your thorough review, and for taking time to point out the consequences of our decisions during the different phases. We acknowledge that there are different viewpoints related to the approaches to be taken in instrument development. As explained in the manuscript, our approach to instrument development was guided by the methodological frameworks for developing measurement scales described by Streiner and Norman (reference 46) and Pett, Lackey and Sullivan (reference 47) and used an applied research approach. In particular, the Streiner and Norman resource is widely used in Canada, and one of the textbook authors holds a position at McMaster University (Department of Clinical Epidemiology and Biostatistics) where several members of our research team also hold positions. Our study proposal underwent rigorous review for scientific validity by a Canadian Institutes of Health Research (CIHR) peer review committee. The initial proposal was revised and re-submitted based on reviewers’ comments, and funded for $516,248 CAD over 4 years, attesting to the fact that CIHR evaluated our study as worthy of funding in a very competitive environment. Each of the five phases of the study builds upon previous phases, and decisions made in the initial phases cannot be undone. Therefore we are limited in our ability to change the approaches used.

We agree that we could explain more fully the different stages of the tool development, and have provided more explanations. Our comments pertain to three major issues as listed below which will be addressed in the same order:

1. Use of parametric statistical methods such as mean (SD), Pearson correlation, and factor analysis (exploratory and confirmatory) for Likert scale variables instead of non-parametric techniques.

2. Use of exploratory factor analysis in Phase 3 “instead of their own skills”.

3. Use of inappropriate measures to test for reliability.
Issue #1: Use of parametric methods to analyze Likert scale variables

-Reviewer’s Comment: “According to the purpose explanatory factor analysis and psychometric testing were chosen. These approaches have been very popular in some applied research fields, probably because they are computer intensive and take “care of everything”. These approaches do not take care of the type of data, the knowledge of the researcher, the qualitative responses of interest, but transform data to quantified, normally distributed figures – far away from the real information collected by hundreds of pregnant women and others. The authors have made this choice of approach and some other important choices during the different phases and I will point at some of their consequences”.

Our response: We believe that the reviewer would agree with the notion that the “Likert scale is the most commonly used scaling technique in psychosocial and health care research” [Making Sense of Factor Analysis: The use factor analysis for instrument development in health care research. Pett, Lackey, Sullivan, Sage Publications, 2003] and use of parametric methods for Likert scales has become more of the rule than the exception. At the moment, there are two schools of thought: one group believes that Likert scale measures are drawn from single Likert response format items where each item is an ordinal measure, not an interval measure, and therefore, use of parametric techniques should be avoided. The other group argues that each Likert scale is intended to measure some underlying continuous variable; therefore, in essence it is continuous, not categorical or ordinal, although it is measured using a 5- or 7-point scale. Unfortunately, what is usually overlooked is that a Likert scale measure is different from a Likert response format item. A Likert scale measure is usually made up of several Likert items (using either mean or sum of the items). As a Likert scale measure is an underlying continuous concept, parametric techniques can be used to summarize or (factor) analyze it. Although there is not much room for discussion in this note, we agree with the second group and for a thorough discussion on this issue, we would like to highlight a recent excellent paper published by Carifio & Perla (Ten Common Misunderstandings, Misconceptions, Persistent Myths and Urban Legends about Likert Scales and Likert Response Formats and their Antidotes. Journal of Social Sciences 3(3): 106-116, 2007).

Issue #2: Use of exploratory factor analysis in Phase 3 of the study “instead of their own skills”

Reviewer’s Comment: “The authors chose exploratory factor analysis instead of their own skills, experiences and gained measurement knowledge. Consequently they rely on mathematical transformations of the self-rated levels of (dis)agreement to statements”.

Our response: Based on the points mentioned above, we believe it is mathematically sound to use a factor analysis for Likert scale data. However, we do agree with reviewer that use of exploratory factor analysis should not be the first step for tool development and researchers should use their own knowledge and skills. Indeed, Glass et al. indicated that most of the parametric approaches are quite robust to the violation of normal distribution (Consequences of failure to meet assumptions underlying the analyses of variance and covariance. Review of Educational Research, 42: 237-288, 1972). However, as we pointed out in the manuscript, since not much is known about quality of prenatal care, we used a mixture of our previous knowledge
and exploratory factor analysis to come up with some sensible factors. We indeed tried different factor solutions and, finally as the content experts, chose the solution that made more sense to us. This has been described more fully in the manuscript as follows: *Use of exploratory factor analysis extracted 5-, 6- and 7-factor solutions. The researchers examined the 3 solutions, and selected the 6-factor solution because the items were judged to be the most relevant and grouped into factors in the most meaningful way based on our clinical knowledge and experience.*

Issue #3: Use of inappropriate measures to test for Reliability

Reviewer’s Comment: “The first sentence is wrong—VALIDITY of an instrument refers to the extent to which an instrument measures what it is intended to measure. Reliability refers to the extent to which repeated measurements yield the same result. In this case of self-rated instrument the aim is to evaluate the intra-rater reliability, or test-retest reliability that refers to agreement between repeated ratings of the same item. The Internal consistency reliability and Cronbach’s alpha is a completely inappropriate approach.”

Our response: We appreciate the reviewer’s comment; however, it seems in the sentence, “Reliability of an instrument is the degree of consistency with which it measures the attribute it is intended to measure,” the word “consistency” was overlooked. We respectfully disagree that internal consistency reliability and Cronbach’s alpha is an inappropriate measure for reliability testing as evidenced and suggested by respected experts in instrument development (e.g., references 47, 57, 59).

Reviewer’s Comment: “The test-retest reliability refers to agreement, and not to correlation. It is a well-known fact that a high level of relationship (correlation) does not necessarily mean high level of agreement! Repeated assessments could result in a systematic change in rating for various reasons – still highly correlated but not reliable.”

Our response: We would like to thank the reviewer for her constructive comments and we agree that Pearson correlation is not appropriate for test-retest reliability. As mentioned in the manuscript, we used the “intra-class correlation coefficient (ICC)” to evaluate test-retest reliability. The ICC is a test of level of agreement.

Other revisions based on this reviewer’s comments:

-As mentioned in our response to the first reviewer, we have now explained how the five response categories of the Likert scale were labelled.
-Table 2: We added the unit of birth weight (grams). However, we retained percentage decimals (see explanation above for Issue #1).
-We expanded our explanation regarding the Flesch-Kincaid test and how to interpret the result: *We used the Flesch-Kincaid Grade Level test, available in Microsoft Word, to assess the readability of the 46-item QPCQ. This test rates text on a U. S. school grade level, which is similar to the Canadian grade level system. The QPCQ had a Flesch-Kincaid grade level score of 8.7, which means that women with a grade 9 education can read and understand the items in the QPCQ.*
- We added an explanation to the text regarding how the mean scores in Table 4 were calculated: Each subscale mean score was calculated by first reversing the scores of any reverse scored items in the subscale, then summing the scores for the items of the subscale and dividing the sum by the number of items. As noted in response to the first reviewer, an explanation of how to interpret QPCQ scores has been added to the manuscript.

- Hypothesis testing is an accepted method of evaluating construct validity. Quality and satisfaction are different theoretical concepts, so we were not looking for agreement between the two measures. We were examining whether women’s perception of quality of care is associated with their satisfaction with care. We respectfully contend that this reviewer’s conclusion that “the QPCQ is not necessary” reflects a misinterpretation of the information presented.

- We agree that convergent validity is a sub-concept of construct validity and have revised the paper accordingly.

Other comments for the Editor:

The ethics statement has been updated to include the name of each ethics committee that approved our study.

We have received a Creative Common license for the QPCQ. The following statement has been added to the Acknowledgements section (if this is not the appropriate section, please advise me as to where this statement should be located in the manuscript):

![Creative Commons License](cc_license.png)

*The QPCQ is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. © Copyright 2013. W. Sword, M. Heaman, and the QPCQ Research Team. McMaster University. All rights reserved.*

Thank you for your consideration of this manuscript. We look forward to hearing the results of the review.

Sincerely,

Maureen Heaman, RN, PhD
Professor, Faculty of Nursing
Associate Professor, Departments of Community Health Sciences and Obstetrics & Gynecology, Faculty of Medicine
Scientist, Manitoba Institute of Child Health
Research Scientist, Manitoba Centre for Health Policy
University of Manitoba