Author's response to reviews

Title: Validation of the Disease Burden Morbidity Assessment by Self-Report in a French-speaking population

Authors:

Marie-Eve Poitras (Marie-Eve.Poitras@USherbrooke.ca)
Martin Fortin (Martin.Fortin@USherbrooke.ca)
Catherine Hudon (Catherine.Hudon@USherbrooke.ca)
Jeannie Haggerty (Jeannie.Haggerty@mcgill.ca)
José Almirall (Jose.Almirall@USherbrooke.ca)

Version: 2 Date: 20 December 2011

Author's response to reviews: see over
December 20, 2011

Editorial Office
BMC Health Service Research

On behalf of my colleagues and myself, I would like to thank you and the reviewers for reviewing our manuscript “Validation of the Disease Burden Morbidity Assessment by Self-Report in a French-speaking population (MS 4871060485729588), resubmitted for publication to BMC Health Service Research.

Below, you will find a detailed response addressing each reviewer’s comments.

Reviewer 1

Comment:
Please provide reference(s) to research/literature to support the statement: "However, impact of daily living is best evaluated by the patient."
Response:
The sentence has been extended and a reference was added.

Comment:
2. In the 8th sentence, the word sensibility should be "sensitivity".
Response:
The change was made.

Comment:
3. What other measurement properties of the DBMA have not been reported? Have those properties not yet been assessed or as written, not yet reported or published?
Response:
We specify now the properties that have not been reported or published yet.

Reviewer 2

Comment:
1. Introduction:
The article would improve if the authors managed to fit the research better in the existing context. What patients are we talking about here (first sentence), all patients? Could it also be relevant for a general population? What other instruments are relevant here? For example, how about the Charlson comorbidity index? The article seems to be based on
the work of Bayliss, Ellis and Steiner (2005), who demand replication of their work in other settings. That is relevant.

**Response:**
We made changes in the Introduction explaining that the instrument is relevant for studies involving patients with chronic diseases, either in primary care settings or the general population, and we explain the limitations of other existing instruments.

**Comment:**
Why and how was the additional item for depression/anxiety added to the DBMA? Does it consists of 21 or 22 conditions?

**Response:**
The DBMA we used consists of 22 items. We explain now in the text why we added depression.

**Comment:**
Setting: Family medicine Unit. Is the DBMA particularly relevant here? Why? Is it a unit with many chronic diseases? Likely to exist also after two weeks (test-retest reliability).

**Response:**
The Family medicine Unit where we work is a primary care setting where, as many primary care settings in the world (if not all of them), patients with multimorbidity are the rule rather than the exception. Chronic conditions in these patients are not likely to change in a period of time of two weeks. The DBMA is relevant for other primary care settings and also for studies evaluating disease burden in the general population.

**Comment:**
Concomitant validity: The choice of the CIRS is ok but in terms of the ICF framework of the WHO they are measuring different health/disease impacts. I think the DBMA is aiming at measuring impact on Activity domain where as the CIRS is measuring impact on Bodily Function. Hence, they might be correlated but certainly not perfectly.

**Response:**
We agree with the reviewer, both instruments are based on different constructs. We discuss the differences in the article. The choice of the CIRS was based on the fact that it is an instrument measuring multimorbidity while taking into account condition severity. The DBMA also evaluates disease severity from a different perspective.

**Comment:**
Analyses: Analyses of sensitivity and specificity should be stated more clearly. Some of the text mentioned at p.11 (Discussion) could be moved here.

**Response:**
Some of the text mentioned in the Discussion was moved to this section and the analysis of sensitivity and specificity was stated more clearly.

3. Results:

**Comment:**
I miss a description of the DBMA-Fv itself. What were the average number of chronic conditions mentioned, which conditions were most frequent (I guess Table 2), and which conditions had most impact (on average)?

**Response:**
The reviewer is right, such a description was missing. We added it to the Results because it is not the same information reported in Table 2.

**Comment:**
The description of missing values (p 12) should be moved to the Results section on validation study (not sure about this heading).

**Response:**
We corrected the Results so that all information used in the Discussion regarding the missing values was previously reported in the Results. The heading Validation Study was deleted because, indeed, all results refer to the validation study.

**Comment:**
As is know, three different samples are the basis of the analyses. T1, T2, and the complete data. I think the authors could reduce this by working with two samples: the T2 when conducting sensitivity analyses, and the complete sample (66) when testing test-retest reliability and concomitant validity. The T1 sample is not that interesting in itself although it is useful to provide characteristics of those within complete questionnaires, especially because this was quite large I think. Any more information about reasons would be very helpful because it could contribute to the feasibility of the questionnaire.
The multivariate analyses are not reported. I would also report characteristics of the 66 with complete questionnaires at T1 and T2 in Table 1. Place of birth is not informative – leave out.

**Response:**
The results of the multivariate analysis are now reported. Place of birth was removed from Table 1. We analyzed the characteristics of the 66 patients with complete questionnaires and indeed it did not add any interesting information to Table 1 and decided to leave it out. We also analyzed the concomitant validity using only questionnaires with complete data at T1 and T2 and obtained similar results.

**Comment:**
Table 2: replace sensibility with sensitivity. I suggest the authors use those with complete questionnaires at T2 – N is the same for all diseases.

**Response:**
Sensibility was replaced with sensitivity. We decided to keep all questionnaires at T2 because their completeness did not have any effect on sensitivity and specificity calculations per disease.

**Comment:**
4. Discussion
Confounded with results (see previous comments) and includes some repetitions. As with the introduction, the discussion should be more embedded in a context. Of relevance, patient groups, previous research. Ideally, it should conclude with a recommendation for
when to use it, how and further research. The DBMA-Fv might be important for but IS not a quality of life assessment.

**Response:**
We made changes both in the Results and Discussion sections trying to correct these aspects.

---

**Reviewers 3**

**Comment:**
1. Translation. In the field of patient-reported outcomes, forward-backwards translation has become the standard method for translating questionnaires. Perhaps this method was not required here given the nature of the questions, which comprise a list of conditions. However, this should be discussed.

**Response:**
A statement addressing this issue was included in the methods section under the heading “Translation”.

**Comment:**
2. Why was the additional item of anxiety/depression included? Was this justified according to the results?

**Response:**
This comment was also made by reviewer 2. We now explain in the text why we added depression.

**Comment:**
3. Cognitive interviews. Please explain how the convenience sample of patients was selected. Was this also in primary care?

**Response:**
We now explain in the Methods that two of the authors recruited the convenience sample from their consulting patients.

**Comment:**
4. Cognitive interviews. What was the process of making changes to the questionnaire? How many interviewers were there and what was their professional background? Did the researchers meet up after each interview to discuss potential changes? Please be more explicit.

**Response:**
We expanded that section in Methods and explain more explicitly the process for the interview and changes to the questionnaire.

**Comment:**
5. Methods, validation study. Were reminders used in the postal survey?

**Response:**
Yes. We added a section explaining this point in Methods.
Comment:
6. Methods, validation study. In what order were the DBMA-FV based on medical records and the CIRS completed by the research nurse, or were they completed simultaneously?
Response:
First, the nurse completed the DBMA-Fv based on the medical record for each participant in the study. Afterwards, she completed the CIRS for each participant.

Comment:
7. Methods, validation study. One research nurse was used to extract data from patient records and intra- and inter-reliability were not assessed. The limitations of this approach as a means of assessing validity should be discussed.
Response:
We already addressed this issue in a previous publication (Hudon C, Fortin M, Vanasse A: Cumulative Illness Rating Scale was a reliable and valid index in a family practice context. J Clin Epidemiol 2005, 58:603-608.). In the study, we analyzed the intra- and inter-rater reliability of the method and concluded that trained nurses can safely score the CIRS from chart review.

Comment:
8. Data analysis. The criterion for incomplete questionnaires is stringent. Does this follow the developers’ recommendations? Is this recommended in practice or are there grounds for taking this approach in the current evaluation of the DBMA? This should also be taken up in the Discussion.
Response:
The developers of the DBMA reported a rather low response rate (28%) but they did not address the issue of incomplete questionnaires. Indeed, the number of incomplete questionnaires in our study was important at T1 (19.6%) but not at T2 (6%). Being stringent in this criterion increases the non response error in our study but it is negligible at T2. We used data from T2 in our analyses.

Comment:
9. Results. The dropout rate due to missing data is very high for T1 compared to T2. Is this because patients lacked time for completion? If so this maybe adequate grounds for using the T2 data for validation purposes. However, there may have been a learning effect, with patients having had time to think about the questions between administrations and hence provide data of higher quality that is possibly more reliable and validat T2.
Response:
We agree. For these reasons we used T2 data for validation purposes.

Comment:
10. Results. What was the range of missing data? Most patients completed all items but did one or two patients miss out several? Were some items more frequently missed than others? This is mentioned in the Discussion but should also be reported in the Results. Descriptive data for the DBMA items would have
been informative including missing data and mean scores (sd) for the activity limitations. How many patients included additional conditions and what were they? How many patients changed areas at T2?

Response:
We expanded the description of the DBMA in our results. We explain that, at T2, there was only one unanswered question in five patients. We also included a description of conditions most frequently self-reported, those conditions with more impact on patients’ daily activities, and conditions added to the list by the patients.

Comment:
11. Was age the only variable that was statistically significant in the regression model? Please clarify and give the level of significance. Older people have greater comorbidity and hence face a more demanding task when completing the DBMA. How does this relate to the evidence more generally for completion rates And age? This should be taken up in the discussion.

Response:
We added to our results the significance level of all variables included in the multivariate logistic regression model and took up this point in the Discussion.

Comment:
12. The administration of the test-retest study took place in two different settings. The potential implications of this should be taken up in the discussion.

Response:
We agree. We now address this point in the discussion.

Comment:
13. Were there differences between the groups who returned (n=85) and did not return (n=12) questionnaires at T2?

Response:
We did the calculations for the reviewer and found some differences. Comparing the group who returned the questionnaire vs. the group who did not, we found that the group who did not was younger (mean age + SD: 50.9 + 15.4 yr vs. 38.9 + 16.6 yr) and had more female subjects (% of female subjects: 64.7% vs. 83.3%) but had a similar degree of multimorbidity (mean CIRS score + SD: 7.7 + 4.2 vs. 6.6 + 5.7). As the group who returned the questionnaire is 7 times larger than the group who did not, we consider that the former is more representative of the primary care setting than the latter and these differences do not affect the validation results.

Comment:
14. The test-retest results relate to what might be a select group. This should be taken up in the discussion.

Response:
We took up this point in the discussion of the limitations of the study.
Comment:
15. Did the source of unreliability relate mostly to the activity limitations ratings or the medical conditions? In the Discussion it is stated that both components of the measure could be used (page 11, para 2) and hence this issue should be taken up.

Response:
We realize that the mentioned paragraph is confusing as we refer to different things. It is misleading. We now present the content in two paragraphs. In the first paragraph we discuss the expected type of correlation between two instruments based on different constructs. In the second paragraph we comment on characteristics of the DBMA and mention that it could also be used as a simple count of chronic conditions. This comment stems from the fact that many studies on multimorbidity are not comparable because of differences in the list of chronic conditions considered. Use of the same list of chronic conditions would allow better comparisons between studies.

Comment:
16. The inclusion of additional areas by patients might increase content validity to the individual patient but are there criteria for determining if an additional area is valid?

Response:
The DBMA is a relatively new instrument and there is no general agreement regarding this aspect. We considered valid additions those chronic conditions needing medical follow-up.

Comment:
17. Discussion. ‘Excellent’ reliability might be taken to imply that reliability could not be improved upon. I would prefer ‘satisfactory’ or ‘high’ which is more appropriate with a test-retest estimate under 0.90.

Response:
We made the change in the text.

Comment:
18. ‘Sensibility’ is used in the Introduction and Table 2 - ‘sensitivity’.

Response:
We made the change in the text.

Discretionary revisions

Comment:
1. Introduction. It is stated that complete measurement properties have not been reported. The other measurement properties that are relevant for such a measure should be briefly summarised here.

Response:
We made the change in the introduction.
Comment:
2. Methods, The Instrument. ‘Accessible language’ presumably means understandable to patients? Please clarify this. Please also state that items have five-point descriptive scales (as in the original questionnaire). As it stands the reader is left unclear if the scaling is end-point only or all-point descriptive scales.

Response:
We clarified these points in Methods.

Comment:
3. Methods, The Instrument. If patients can add conditions to the list, then in theory the scores can be higher than 105.

Response:
The reviewer is right. We corrected the text.

Comment:
4. The T1 survey data were not used to assess validity because the DBMA was originally developed as a postal questionnaire. However, patient-reported outcome measures are often administered in a clinical setting and hence the DBMA might also be administered in such a manner in future applications. It would also be interesting to compare the two sets of results for T1 and T2.

Response:
We agree that it would be interesting to compare the DBMA administered in a clinical setting with the mail questionnaire. However, we realized that the waiting room was not an appropriate place to complete the questionnaire for the reasons discussed in the article. We think that our data at T1 is not good for such a comparison.

Comment:
5. Introduction. “...another index of multimorbidity” can be deleted.

Response:
The text was changed.

Comment:
6. Translation. Replace “brought” with “made”.

Response:
Done

Comment:
7. P 9, Reliability. Move “correctly” to before “completed” in the first sentence.

Response:
Done

Comment:
8. P10, Discussion. Replace “relationships” with “associations”.

Response:
Done
As requested, we have provided written responses following each point made by the reviewers. We hope that you now find our revised manuscript suitable for publication in your journal and look forward to hearing from you.

Sincerely,

Marie-Eve Poitras, RN, MSc