Author’s response to reviews

Title: Evaluation of Data Quality of interRAI Assessments in Home and Community Care

Authors:

Sophie Hogeveen (s2hogeve@uwaterloo.ca)
Jonathan Chen (x2chen@uwaterloo.ca)
John Hirdes (hirdes@uwaterloo.ca)

Version: 1 Date: 10 Jun 2017

Author’s response to reviews:

Dear Dr. Jiang Bian,

I would like to resubmit the revised version of the following manuscript for publications in BMC Medical Information and Decision Making: Evaluation of Data Quality of interRAI Assessments in Home and Community Care (MIDM-D-17-00098). I appreciated the thoughtful comments and suggestions put forward by the reviewers and have attempted to address each one. Specific reviewer comments are in bold with my response underneath each one in regular font. Where changes were made, page and line numbers are provided as a reference (revised version, with all changes accepted).

One common criticism was a lack of clarity and over-use of abbreviations that made it difficult to read and understand. I have created two tables (tables 1 and 2) to help readers better understand interRAI assessment systems and the variables included in the analyses. I have also removed several unnecessary abbreviations and clarified parts of the Methods section.

Thank you in advance for reviewing the responses and for considering this manuscript for publication.

Sincerely,

Sophie Hogeveen
Reviewer reports:

Nan Liu (Reviewer 1):
Thanks for giving me the opportunity to review this article. This is a well-designed study with promising results for home and community care. I only have the following minor suggestions.

Thank you for your encouraging comments.

- I would like to suggest the authors to rename Table 1a and 1b to Table 1 and 2, and Table 2a and 2b to Table 3 and 4, respectively. These changes have been made, although the numbering has changed as new tables have been added (see pages 29-37).

- The authors may also discuss possible implementation (including potential challenges) of the assessment tool in the whole country. Further, are there any plans to apply the system in other countries with similar healthcare settings?

The Resident Assessment Instrument – Home Care (RAI-HC) has been widely implemented across Canada, with eight Canadian provinces and territories having implemented its use as the mandated assessment for home care services (pg. 4, line 9). We chose to evaluate Ontario and British Columbia RAI-HC data as the largest data holdings available were from these provinces. Beyond Canada, there are ten countries internationally (United States, Iceland, Finland, Belgium, Switzerland, Italy, France, New Zealand and Hong Kong) with large-scale implementation of the RAI-HC or interRAI Home Care planned or underway (pg. 4, line 13). I have added this point to the Background section. Currently in Canada, the interRAI Community Health Assessment (CHA) is only being used in Ontario. Internationally, several countries in Europe and the Middle East as well as New Zealand, Australia and Hong Kong have used the interRAI CHA regionally/locally or for testing.

Ying Wang (Reviewer 2):

This paper proposed an interesting topic, aiming to evaluate data quality of interRAI assessment in home and community care. However, to make it easy-reading, this paper requires more clarity about methods and more improvement on the ways to describe data setting and to report results. The statistical analysis results to support the statement of similarity should also be provided.

Thank you for your thoughtful comments. I have addressed each point below.
1. Abstract needs a lot clarity. If possible provide solid indicators/measures/evidence of data quality, such as consistency to support the descriptive results. Please clarify the various statistical techniques used in paper and explain what does 'good quality' mean and what is the 'new method' provided in the Conclusions.

The abstract has been edited. The statistical techniques used have been specified and some specific values of internal consistency and convergent validity have been provided, although in a summary form (pg. 2, lines 7-8 and lines 11-14). Unfortunately, the word limits of the abstract do not allow for a specific description of the results of the data as there are so many variables, across different settings and years. Good quality data has been explained as being trustworthy to be used for decision-making at the organizational and policy levels (pg. 2, line 21). The distinction of the ‘new method’ has also been explained (pg. 2, line 24).

2. Introduction: The introduction runs a little bit overly long to show the study aims and importance of the topic. It is good to give a general overview of the data quality studies in health care, however it stated too many detailed results in other related studies but not a focus of this paper.

The description of different data quality problems has been shortened, to focus on those issues mainly addressed by this paper. (pg. 3, line 18). I am unclear as to which detailed results of other studies were referred to but would be happy to remove additional parts of the Background section if clarified.

3. The paper is overflowing with abbreviations. It makes readers confused and actually increases the difficulties of understanding.

I appreciate that the abundance of abbreviations may be confusing for readers. Two tables have been added to help with the understanding of the various acronyms and assessment systems (tables 1 and 2). Several abbreviations have been replaced with the full names/words to facilitate understanding.

4. A highly related systematic review paper was missing. 'The Resident Assessment Instrument-Minimum Data Set 2.0 quality indicators: a systematic review, BMC Health Services Research 2010 10:166, DOI: 10.1186/1472-6963-10-166'.

Thank you for bringing this interesting article to my attention. Its findings are particularly interesting as they relate to the presence of systematic error discovered. I added a reference to
this paper where I described systematic error in the Background section (pg. 3, line 11). In an effort to avoid making the introduction even longer and providing too many details on an assessment (RAI-MDS 2.0) and studies that are not a focus of this paper, I decided not to describe this review in detail in my paper.

5. Data sources: p7, line 19. Four sources should be explained afterwards the statement: 'Data for the present study were obtained from four sources'. Or the article would reads better to move the notes about additional files providing more information about each data setting to the beginning of the Data sources.

The four sources have been listed after the statement and the note about the additional file was moved up to the beginning of the data sources (pg. 8, line 3).

6. A table showing all the variables would be very helpful to summarize the detailed information and make paper easy-reading.

Thank you for the suggestion, I have included this table (see table 2).

7. P 12, line 6, there were only five indicators explained after the authors stated that six particular sets of indicators. What is missing?

Thank you for raising this concern. The statement was unclear and has been clarified to read more clearly as six particular sets of indicators (pg. 13, line 2).

8. P22, line 18, please explain what is enough longitudinal CHA data? Was there gold standard or criteria to evaluate 'enough'?

Reassessments with the interRAI CHA in the community support services sector do not happen regularly. Therefore, it is not possible to examine rates of potential auto-population in the CHA data by linking consecutive assessments. I clarified this in the Analysis section (pg. 13, line 14).

9. P14, the authors stated that 'there were not major differences in the percentage of heart failure clients in HC clients in both provinces …'. Was this 'not major difference' statistically significant? If possible provide p value to support this statement.
The data from Ontario and British Columbia for this section of the analysis are in separate datasets and would be difficult to merge to compare the rates in a statistical way and obtain a p value. I have edited this sentence to remove any statement as to the significance of the difference in rates between the two provinces (pg. 14, line 19). The sentence is now as follows: The percentage of heart failure clients ranged from 11.5% to 14.4% in Ontario HC clients and from 14.2% to 14.9% in BC HC clients, although the percentage among CHA-assessed CSS clients was lower.

10. The result section is to report the basic findings, rather than discussion and comparison of results to other similar studies. It would be better to move the comparison to Discussion section.

Thank you for the suggestion. I removed several sentences from the Results section that compared the findings to the results of Hirdes et al. (2013) evaluation of RAI-MDS 2.0 data. I then made these statements in the Discussion section.

11. P17, line 11, what are the cut-off points?

The cut-off points were chosen according to those cited in a paper by Hirdes and colleagues (2013). I have clarified the statement describing these as follows: An alpha value of 0.70 or higher indicated acceptable reliability and an alpha value of 0.80 or higher indicated excellent reliability to make it more clear for the reader (pg.18, line 4).

12. On 22, line 4-5, the sentence to describe the sheer number of RAI-HC does not belong to the limitation section. Move it to somewhere else.

Thank you for the suggestion, I moved this sentence to the Conclusions section (pg. 23, line 2).

Hui Hu (Reviewer 3):

This is a well written manuscript describing the evaluations of data quality of the interRAI in home and community care. Yearly trends were generated for convergent validity, reliability, and indicators of auto-population. However, one concern is the proportion of the data excluded from the analyses and the lack of descriptions of how data were excluded. Alternatively, the authors can retain most of the data by generating monthly trends instead. In addition, there are many
abbreviations in this manuscript, making it hard to read. I suggest the authors to add some figures/tables to summarize these instruments, systems, and variables. Below are my detailed comments:

Thank you for your encouragement and recommendations. I have addressed each point below.

1. **Background:** many abbreviations were used. It would be better if the authors can provide a chart to illustrate the relationship of the different system and instruments (i.e. CCRS, RAI 2.0, RAI-HC, interRAI Home Care, HCRS, CCAC, OACCAC etc.) in addition to the descriptions.

I appreciate that the abundance of abbreviations may be confusing for readers. Two tables have been added to help with the understanding of the various acronyms and assessment systems (tables 1 and 2). Several abbreviations have been replaced with the full names/words to facilitate understanding.

2. **Methods-Data sources:** you mentioned for the CHA data that the assessment closest to July 1st per individual per year was included in the analyses. How about the RAI-HC data? Did you use the same criteria to select the assessment for analyses? Why not randomly pick one assessment per individual per year?

Thank you for your questions, I appreciate that these criteria were unclear. The same criteria were used to select one assessment per client per year for RAI-HC data as for CHA data (assessment closest to July 1st). I have clarified this point throughout the Data sources section (pg. 8, lines 14, 19; pg. 9, line 12). Clients generally only have one or two assessments per year, and are rarely reassessed more often so it is not necessary to randomly pick one assessment out of a multitude. When there were multiple assessments, the assessment closest to the midpoint of the year was chosen to try and ensure a distribution of assessments throughout the calendar year. The absolute median number of days between the assessments and July 1st overall was 92 (25th percentile=48 days; 75% percentile= 136 days). The absolute number of days between assessment and July 1st ranged from 0 days to 183 days. It appears that the assessments are somewhat dispersed throughout the year, with half of them happening 1.5 months to 4.5 months from July 1st.

3. **Methods-Data sources:** please move the descriptions of sensitivity analyses to the Analysis section and put the results of the sensitivity analyses to the Results section.
Thank you for the recommendation. I moved both the description and results of the sensitivity analyses to the Analysis section (pg. 12, line 4). I felt that the results should remain with the description to make it easier for the reader to understand.

4. **Methods-Variables:** again, a table summarizing the variables and specifying the type, range, and cutoffs of the variables will be helpful.

Thank you, I have done so (see table 2).

5. **Analysis:** Given the large sample size you have, it might be more informative to conduct the analyses using monthly time series trends to explore if there is any seasonal pattern in validity and reliability. And you will be able to retain most of the data since you can keep one assessment per individual per month (currently over 42% and 15% data were excluded from the analyses for the RAI-HC data from Ontario and BC, respectively).

Thank you for your suggestion. At this point in time, we are more interested in trends over the years of RAI-HC and CHA implementation in Canada than seasonal patterns in validity and reliability. The assessments included in the analyses are distributed throughout the year so seasonal patterns should not affect the findings. As described above, the absolute median number of days between the assessments and July 1st overall was 92 (25th percentile=48 days; 75% percentile= 136 days). The absolute number of days between assessment and July 1st ranged from 0 days to 183 days. Further, the tables included are already quite extensive and would become unbearably large if monthly trends were reported.