Author's response to reviews

Title: Comparing the Functional Independence Measure and the interRAI/MDS for use in the functional assessment of older adults: A review of the literature

Authors:

Christine Glenny (cglenny@uwaterloo.ca)
Paul Stolee (stolee@uwaterloo.ca)

Version: 2 Date: 24 August 2009

Author's response to reviews: see over
Cover Letter

Department of Health Studies and Gerontology

University of Waterloo, Ontario, Canada, N2L 3G1

To

The Editor,


We are very grateful to the two reviewers for their very thoughtful and detailed comments on the previous draft of our paper. Attached please find a revised version of the paper; we believe that the comments and suggestions have resulted in a much stronger manuscript.

Our responses (in italicized text) to each of the reviewers’ comments are detailed below:

Reviewer’s report

Reviewer: Ying-Chih Wang

Major Compulsory Revisions
1. Page 2, 2nd paragraph - I am not sure that “Instruments that are designed for younger, healthier, and more homogenous groups are unlikely to have the same psychometric properties with older adults” because of “their frailty, comorbidity, and heterogeneity”. Because functional status is an end product of how they perform daily tasks and is assessed in a standardized procedure by trained clinicians, the final score should reflect the elderly subject’s overall functional status. That is, if they are fragile, they would obtain a lower (i.e., worse) functional status score. Therefore, the psychometric properties should be fine. What do you mean by “Instruments …are unlikely to have the same psychometric properties with older adults”? What psychometric properties are you indicating?

We understand the reviewer’s perspective, however the point we are making relates more to the psychometric properties of an assessment when tested or used in a population of subjects, rather than to the results that are obtained when the instrument is used with a particular subject. Given that reliability, for example, is calculated as a ratio of subject variance to subject variance + rater(or test) variance + error variance, the amount of variability present in a population will have a bearing on the calculated reliability of the instrument when used in that group. In the extreme case, subjects who all score at the lowest functional level assessed by an instrument may all be accurately rated, but because there is no variation in the subject pool, the instrument will be unreliable. The level of variability would also affect the extent to which instruments would correlate with other instruments used for construct validation. Additional references were added to clarify how the reliability and validity of an instrument are dependent on characteristics of the sample being measured.

2. Page 8, start from “Both instruments were consistently found to be reliable… “till the end of results session on page 12 – please add more information about the results of reliability or validity coefficients
reported from previous literature when you described the results. For instance, when you mentioned that “internal consistency was high for the FIM total score…” we would get a better idea if you mentioned a range of internal consistency values such as “internal consistency was high (0.90-0.97) for the FIM total score”. The same comments with other sentences as well. For instance, if you indicated that “Dallmeijer and colleagues concluded that the FIM motor (0.89-0.98) has slightly higher internal consistency than the FIM cognitive (0.68-0.88), then we as an audience will get what you meant by higher internal consistency.

This is a good suggestion, and numerical results were added where possible.

3. Page 12, line 22, under Discussion – Could you describe briefly about those measurement properties of these two assessment tools that are inconsistent between the older and younger population?

As discussed in response to question 1.

4. Table 1 - I will suggest including a brief definition of each reliability and validity terminology under Table 1. That is, what is internal consistency reliability? What is criterion validity?

We have added a table (Table 2) that provides brief definitions of the reliability and validity categories discussed in this paper.

5. Table 2-5 – In general, the authors make a considerable amount of effort to collect and report the results. But I personally suggest that these summary results should be simplified. The main reason is that readers may have difficult time going through all the details and digesting so much information. The results should be more organized in a way that report similar findings systematically across previous studies. Please take a look at Barak and Duncan’s manuscript (Barak S, Duncan PW. Issues in selecting outcome measures to assess functional recovery after stroke. NeuroRx. 2006 Oct;3(4):505-24.).

Our goal for this paper was to give a more comprehensive review; therefore, we believe the level of detail in table 3-6 is necessary. We also found that it was not possible to systematically group the studies with similar findings because many of the articles had multiple methods and results. We appreciate the suggestion and attempted to reorganize the table by setting; however, we then found that it was potentially more difficult for the reader to quickly locate desired articles when they were not alphabetized. There may be no ideal way to present this information; however the current format will aid the reader who, when going through the text, wishes to refer to a particular study.

Minor Essential Revisions

1. Page 2, line 5 (under Background) - The abbreviation of PAC is inconsistent with the wording. Suggest changing post acute rehabilitation (PAC) to post acute care (PAC) in rehabilitation or removing the abbreviation.
2. Page 3, line 22 - “The (FIM) motor subscale collects information involving self care, sphincter control, transfer (?), and locomotion …”
3. Page 3, line 25 - I suggest changing “high scores on the FIM denote patients that have a high level of independence…” to “higher scores on the FIM denote patients that have a higher level of independence…”
4. Page 4, line 5 (under InterRAI/MDS) - Please capitalize the “interRAI … is an international research consortium that…”.
5. Page 5, line 2 - add the abbreviation after Cognitive Performance Scale (CPS).
6. Page 8, line 1 – Is Table 1 or Table 1-5 that summarizes the total sample of articles that met the criteria?

Minor essential revisions 1-6 were amended as suggested by the reviewer.

Reviewer: Yee Sien S Ng

I have no major compulsory revisions indicated.

Minor Essential Revisions:

The main minor essential revision which I hope the authors address is that although it is an excellent review, there is little in the way of practical suggestions as to how best to apply the results of the review. The conclusion in the abstract that “Additional psychometric research is needed on both the FIM and MDS, especially with regard to their use in different settings and with different client groups” as well as a similar comment that more research has to be done in the concluding paragraph of the manuscript, is while well noted; has no immediate practical application. A summary opinion from the authors would be well appreciated from clinicians like me. For example: Both instruments are well suited for use in the disabled elderly population, particularly in the case of FIM for rehabilitation and the MDS for nursing home residents in the present state. However further research should focus on….”

To address this we added an additional section to the conclusion:

“This review assembled and compared available evidence of the reliability and validity of two major systems for the functional assessment of older adults. Overall, we found that there is evidence for the reliability of both instruments; however, the majority of FIM studies were carried out in inpatient rehabilitation settings and most of the MDS articles were conducted with nursing home residents. Before clinicians can confidently use the instruments outside of these settings, additional psychometric research is needed on both the FIM and MDS, especially with regard to their use in different settings and in different client groups. We also found that there is considerably more literature examining the validity of the FIM than is available for the MDS instruments. This supports the continued used of the FIM as a component of the NRS. Nonetheless, it is also important to consider that this analysis only included the ADL and cognition items from the MDS which contains a more comprehensive set of items that may enhance its utility. The compatibility of the interRAI instruments across multiple health care setting should also be considered before determining which tool is the most appropriate outcome measure for this population. We suggest that, in particular, more research is needed to investigate the construct validity of the outcome measures derived from the MDS instruments. Lastly, a direct “head to head comparison” of both tools in the same population would yield valuable information, especially in terms of the assessment of their responsiveness to change. Such a study could also allow for analysis (using Rasch methods, for example) that would facilitate direct statistical comparison of results obtained using the two instruments. While such analyses could theoretically lead to the development of a hybrid instrument, it is unlikely that such an instrument would gain broad acceptance given the extensive investments already made into the two systems. It is more likely that the results would facilitate better understanding of the results of each instrument by users of the other system.”

The other minor essential revisions would be to refocus the results of the ‘data collection and analysis’ section. Points are:
In general, many of the readers will find it difficult to plough through the complex analyses unless this is their area of special interest; and the text already assumes a fair amount of knowledge. I would imagine terms like face validity and confirmatory factor analysis would be difficult to appreciate. Could I suggest:

1. Table 1 is an excellent table and need not be further elaborated in the text. For example, there is no need to repeat sentences like “x number of articles evaluated internal consistency.”

We removed many of these comments from the results section

2. Below table 1 or a separate table, provide one sentence explanations of the key terms: internal consistency, intra-rater, inter-rater, content, construct, criterion and face reliability/validity. Eg “Inter-rater reliability is the agreement which 2 or more raters have which each other; intra-rater reliability is…”

We have added a table (Table 2) that provides brief definitions of the reliability and validity categories discussed in this paper

3. Focus on the topic, which is a comparison of the FIM and MDS, not a review of the properties of the FIM itself. For example the detailed paragraph: “...Eight articles investigated the construct validity of the FIM using Rasch analysis. These had mostly consistent findings: eating and stair climbing were seen to be the easiest and most difficult FIM motor items respectively; expression and problem solving are the easiest and most difficult FIM cognitive items; bowel, bladder, eating, and stair climbing are common “misfit” items on the FIM motor; the distribution of FIM scores has a sigmoidal structure and the number of response options should be reduced...” Whichever are the easiest or most difficult items or whether the FIM has a sigmoidal structure are not relevant to the topic; as there is no comparison on which are the most easiest or most difficult items in the MDS, nor a comment on whether the MDS scores are similarly sigmoidal, skewed or simply difficult to chart.

We feel that the primary aim of this article was to summarize the current literature for both tools; therefore, we feel that it is relevant to include all available data for both each tool, including where comparable data were lacking for the other tool – this helps to identify areas where additional research may be warranted on a particular instrument.

4. There are quite a number of analyses in the discussion section which should be in the analysis section. For example:
   a. “Conversely, Stineman and colleagues [51] investigated this relationship in a sample of community residents and concluded that internal consistency was excellent and no items should be removed for any of the 20 UDSMR impairment types.”
   b. “Daving and colleagues [40] used clinicians to investigate the reliability of the FIM in community residents. They found that the reliability ranged from poor to excellent where the least reliable assessments were completed at different times by different raters.”

We have eliminated several references which we agree were misplaced; however, we feel that some summary information is required in the discussion due to the large volume and complexity of data being reviewed.

Discretionary Revisions:
1. A comment into how the raters in both the FIM and MDS are accredited. If accreditation is difficult and cumbersome, it limits the general use of these instruments across countries as compared to say the Barthel Index which has no strict accrediting bodies.

2. Head to head comparisons of the FIM/MDS into average time/labor/costs taken to complete assessments.

_These are both addressed as limitations of this review in the discussion section – we have added the point that while our review has not addressed accreditation or training requirements, time requirements, software costs, and other administrative expenses, these would be relevant considerations for organizations considering adoption of one of these instruments._

3. In relation to 2, are there more rehabilitation inpatients (ie FIM users) or more nursing home inpatients (ie MDS users) in Canada? It makes more sense to me that all else being equal, I would use the instrument that will require the least time for the larger patient type numbers.

_This is an interesting point; however, it may not be consistent with the focus of this review. Also, more information on both tools is needed before we can determine which instrument is the most appropriate in each setting. This also can be seen as related to the limitation we discussed related to the previous point._

4. I would suggest a ‘birds-eye’ view opinion from the authors. The over-riding concern from a practicing clinician would be to ask if there is/or how pressing is the need for current geriatric practitioners of the FIM to learn the MDS and vice versa; as the authors have pointed out that many of the items overlap anyway for ADLs. For example if a geriatrician would to spend most of the time in an inpatient rehabilitation facility and only a little time in a nursing home; I would imagine there would only be a marginal benefit learning the detailed RAI in addition to the FIM at a lot of additional angst training for the RAI!

_This review is aimed as a preliminary step to guide future research, more data are needed before conclusions regarding which instruments is the most appropriate for each care setting can be made. We believe that the paragraph we added in relation to the main revision suggested at the beginning of this review also applies to this point._

5. The authors could also opine on whether future research should be aimed at refining the psychometric properties of both instruments separately in their current state, or should the focus be to develop a hybrid FIM-MDS system to allow generalization to more patients.

_We have made our suggestions for further psychometric research more explicit, and have also made some comments about the possibility of a hybrid system, in our response to the main suggested revision at the beginning of this review._