

Author's response to reviews

Title: A Rasch analysis of the Manchester Foot Pain and Disability Index

Authors:

Sara Muller (s.muller@cphc.keele.ac.uk)

Edward Roddy (e.rodny@cphc.keele.ac.uk)

Version: 2 **Date:** 14 September 2009

Author's response to reviews: see over

Arthritis Research Campaign National Primary Care Centre
Primary Care Sciences
Keele University
Keele
Staffordshire
ST5 5BG

14 September 2009

Dear Profs Menz and Potter,

We are pleased to resubmit the following article, now entitled "A Rasch Analysis of the Manchester Foot Pain and Disability Index" for submission to the Journal of Foot and Ankle Research as a research article.

We thank the reviewers for their comments and have addressed each of the points raised below. We have also included higher quality TIFF figures in this submission.

Yours truly,
Sara Muller

Mrs Sara Muller
Arthritis Research Campaign National Primary Care Centre
Primary Care Sciences
Keele University
Keele
Staffordshire
UK
ST5 5BG

Email: s.muller@cphc.keele.ac.uk

Tel: +44 (0)1782 734853

Fax: +44 (0)1782 733911

Reviewer's report

Title: The Manchester Foot Pain and Disability Index: A Rasch analysis

Version: 1 Date: 17 August 2009

Reviewer: Anne-Maree Keenan

Reviewer's report:

Overview of study:

This study assesses the internal construct validity of the MFPDI, using the Rasch model, which is part of an ongoing programme of work exploring the utility and properties of the MFPDI. The study is a thoughtful and well written piece of work, which managed to convey some of the very dry Rasch principles in a relevant and appropriate manner.

Specific Comments:

Background: I think that the authors should explain the rationale of the Cook recommendation of reduction of the one item being based on it not fitting a two-domain approach to the questionnaire. As it is not one in which the original MFPDI was based (ie four domains) this is relevant background, particularly in the context of item selection or deletion.

Cook et al make this recommendation based on an exploratory factor analysis that resulted in a two-factor solution. One item (My feet are worse in the mornings) failed to load onto either factor and was removed to create the Modified MFPDI that they developed. We have added something to this effect to the Background on page 3.

It is not surprising that the appearance subscale did not fit several of the Rasch parameters, due to the small number of items. This should be included in the discussion and the abstract.

We have included an expanded discussion of the problems surrounding the appearance subscale in both the discussion section of the main paper (page 10) and abstract and hope this satisfactorily addresses the reviewer's point. (See below also – Specific changes (i)).

General Criteria

1. Is the question posed by the authors new and well defined?

Yes

2. Are the methods appropriate and well described, and are sufficient details provided to replicate the work?

Yes

3. Are the data sound and well controlled?

Yes.

4. Does the manuscript adhere to the relevant standards for reporting and data

deposition?

Yes

5. Are the discussion and conclusions well balanced and adequately supported by the data?

There are areas of discussion that need to be developed a little further.

In response to your comment above, we have added further to the Discussion on the appearance subscale.

In response to point made by Reviewer 2, we have added to the Discussion of DIF and included a paragraph about the application of the Rasch score in other populations, although we have had to conclude that there is no consensus in the literature here and we do not wish to draw a firm conclusion as to the best course of action. We think this is up to the individual wishing to use the score to decide.

6. Do the title and abstract accurately convey what has been found?

The title is a little clumsy and I suggest it should be changed to A Rasch Analysis of the Manchester Foot and Pain Disability Index.

We thank the reviewer for this suggestion and have changed the title accordingly (page 1).

7. Is the writing acceptable?

Yes. The writing is of an excellent standard.

Specific Changes

(i) Abstract Page 2, Line changed from "...suggest that items from the appearance subscale fit the Rasch model, though this may be due to the small number of participants without extreme.... " to "...suggest that items from the appearance subscale fit the Rasch model, though this may be due to the number of items in the subscale (only two) and the small number of participants without extreme.... "

We thank the reviewer for this suggestion and have changed the manuscript accordingly (page 2).

(ii) Patients and methods. I would include a sentence at the end of the Rasch model paragraph (pg 4, paragraph 2) a sentence outlining that the following (ie the model, unidimensionality etc) are characteristics explored within the Rasch model and how each are evaluated.

We thank the reviewer for this suggestion and have added an appropriate sentence (page 4).

(iii) The issue of uni-dimensionality and the t-test is a difficult one to explain. Can I suggest the following amendment:

....It is essential that any scale is measuring only a single construct [26]. To ensure that the FPDI scales were unidimensional, a principal components analysis of the residuals was performed. The aim of this is to identify patterns of the residuals once the "Rasch factor" has been extracted. This is important in order to identify any subsets of items that may be loading together, and therefore may represent a different construct. The absence of any meaningful pattern in the residuals will be deemed to support the assumption of local independence of the items. In order to explore this, the two most different groups of items (i.e. those items whose fit residuals load negatively and those that load positively onto the first component). These two sets of items produce the most different

estimates of person location. Using these two sets of person locations.....

We thank the reviewer for this suggestion and have changed the manuscript accordingly (page 4-5).

(iv) Discussion Page 10, para 4: delete the “only” just before moderate – 645 is generally considered a good sample size.

We thank the reviewer for this suggestion, but have not changed this as suggested, as we were not able to use all 745 people who responded to the Health Survey Questionnaire in the Rasch analysis of the MFPDI. The 745 people did not necessarily have foot pain, and so did not all complete the MFPDI. At the end of the ‘Study sample’ Section under Results in our original and revised manuscripts, we have described that we had only 131 people who completed the MFPDI function subscale without an extreme score, 133 such people for the pain subscale and 36 such people for the appearance subscale.

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:

I declare that I have no competing interests

Reviewer's report

Title: The Manchester Foot Pain and Disability Index: A Rasch analysis

Version: 1 Date: 19 August 2009

Reviewer: Gabrielle van der Velde

Reviewer's report:

Reviewer's Assessment Report:

Thank you for the opportunity to review this manuscript. The authors are commended for wishing to promote good measurement practice in the research field of foot disability.

Major Compulsory Revisions:

The main limitation with this study is related to its conceptualization: 1) what are the justification, rationale, and evidence to support the validity of conducting an analysis of the three FDPI subscales, and 2) how will these sub-scale scores be used by users such that fundamental measurement is fulfilled? What is the practical application of their findings?

1. Background on FPDI requires more detail and should be summarized in a clearer fashion. For example:

a) The authors should explicitly describe the underlying construct(s) the FPDI was designed to measure, and provide background on its development and the conceptual framework that was used to develop the instrument, and so forth. For example, when the FPDI was originally developed, what was rationale for having 4 constructs for the measurement of foot disability? Why was 'personal appearance' conceptualized to be related to disability?

Whilst we understand the reviewer's desire for this information, it is difficult for us to comment on this in detail because we did not develop the FPDI. In fact, part of the rationale for our previous paper (Roddy et al 2009 Rheumatol) was that we did not think that appearance should be included in a definition of disability, which we considered to be about physical function. We have however, added a sentence into the manuscript to explain that the items in the FPDI were developed in conjunction with patients with foot pain and that original 4 constructs were arrived at through exploratory factor analysis (page 3).

b) The authors should describe the scoring of the instrument, the range of possible scores, and the interpretation of scores, as originally proposed by Garrow, and how these were changed over time by other users of the FPDI, and whether these changes were based on some statistical / measurement approach or just ad hoc.

We have added some more information to the Background section on page 3.

c) The authors need to more clearly describe the evolution of the utilization and scoring of the FPDI over the years, starting from Garrow, then Cook, etc. For example, how were 3 constructs ‘confirmed’ statistically and why was one ‘construct’ dropped? What was the rationale / basis for summing scores as did Waxman et al (RCT)?

We feel that we have answered this point in our responses to points 1a and 1b above. The additions to the manuscript are on page 3.

d) Are the 4 ‘constructs’ are in fact ‘subscales’ or ‘factors’ that are thought to or conceptualized to map onto the construct ‘foot disability’?

We are not quite clear here as to what the reviewer would like us to change in the manuscript. Please could the reviewer elaborate further?

Note: some of this information requested above could be included in the methods under a section title, for example, ‘Manchester Foot Pain and Disability Index Description’

2. Background: Authors should present evidence to support statement that Waxman et al. score is an ordinal score. A requirement for ordinal scoring is confirmation / evidence of unidimensionality of the FPDI (i.e., measures only one construct) through some form of statistical analysis (e.g., factor analysis, Rasch analysis, Mokken analysis). The authors state that the FPDI was originally designed to have 4 constructs; this statement suggests that the FPDI is not unidimensional, and therefore the score by Waxman et al. was not ordinal.

We thank the reviewer for this comment on a point that we had overlooked. We have also inserted details of two clinimetric studies, which along with Waxman’s trial, have summated scores to produce a total FPDI score. We have changed the manuscript to say the sum score was ‘at best’ ordinal (page 4). We have not discussed the issue of dimensionality here as it is covered elsewhere in the manuscript.

3. Background: The authors state that ‘the only way to derive such a score from ordinal item responses’ is through the use of the Rasch model. Actually, the intent of a Rasch analysis is first to determine whether a scale in question represents unidimensional, interval-level measurement, and if not, one can attempt to adapt the scale to determine whether adaptations to the scale can lead to unidimensional, interval-level scaling through analytic iterations that consist of adaptations (e.g., response option collapsed, items removed, etc.) followed by repeated Rasch analysis.

We agree with the reviewer with regard to the intent of a Rasch analysis. However, in this paper our objective was to investigate the legitimacy of deriving an interval-level score from the ordinal responses to the FPDI items. In order to meet this objective, we checked that the data met Rasch model specifications such as unidimensionality. We subsequently concluded that we were able to obtain this interval level measurement. If we have misunderstood the reviewer’s request, we would be happy to reconsider this point.

4. Objective: A clearer statement of the study objective would be helpful (presumably the last paragraph [2 sentences] of the Background section). The statement of objective is contradictory to the sentences that precede it. The preceding sentences discuss an RCT that used an overall score, and ‘way to derive such a [single] score. So the reviewer is unclear whether the objective is it to determine whether the 3 subscales are unidimensional, interval-level scales (3 scores?) or the FPDI as a scale (single score)?

We acknowledge that our objective may not have been clear and have changed ‘such a score’ to ‘interval-level scores’ (page 4). We have also specifically stated the objective of the study in the final sentence of the Background.

5. Study Sample: The reviewer suggests that the description of the study sample provided in 'Results' be synthesized with the description of the study sample in 'Methods' since there is some redundancy and results of the survey are not those of this Rasch analysis. The description of the sample should be provided in the Methods section only.

Whilst we thank the reviewer for this suggestion, we have chosen to leave the study sample description in the Results section. We have done this because, although not a result of the Rasch analysis, they are previously unpublished findings that we cannot reference and so we think are justified 'results' of this study. We would be grateful if the Editorial team could provide guidance regarding this formatting issue.

6. Methods:

a) As noted above, there should be a section that provides a description of the FPDI, its conceptual framework, development, validation, psychometric properties, etc.

We feel that we have addressed this point in the changes that we have made to the Background section.

b) Rasch analysis is specifically concerned with assessing the fit of data to the Rasch model, not with fitting the Rasch model to data. Yet the latter is suggested throughout the manuscript, for example, by the Results heading: 'Fit of the Rasch Model'.

We thank the reviewer for pointing out these typographical errors and have corrected them.

c) The explanation and description of the principles of Rasch analysis and the methods are rather insufficient. While it is true that the Rasch model has been described elsewhere, the current summary description of the model provided in the text is insufficient for the reader to understand the basic concept behind this model.

At this stage, we have not added any additional description to the paper, as Reviewer 1 has stated that they are appropriate, well described and sufficient. We would be grateful if the Editorial team could guide us as to how we should approach this difference of opinion between the reviewers.

d) The Methods are not adequately reported. Here are examples. The authors should explain why the partial credit model is used. The explanations in the Methods section are unbalanced, where in some cases, concepts or methods are fully explained (e.g., method for testing unidimensionality) but virtually no explanation for other concepts (e.g., threshold plots and what they represent). Why did the authors expect the response categories to be hierarchically ordered? Why must a scale measure only one construct and be unidimensional? What was the correlation cut-off value used to establish response dependency? What considerations were made to determine whether the sample size was appropriate for the various analyses (e.g., fit to model, DIF, etc.)

We used the partial credit model to allow for possible differences between items in the logit distances between thresholds, which we think might conceptually occur. We have chosen not to include this in the manuscript, as we think that it is a technical detail that would make the paper unnecessarily complicated.

With respect to the expected threshold ordering, we have added some more detail on what we expect this ordering to be. We think that it is self-explanatory why we expect this to be the empirical ordering of the response options.

We have referenced that scale must be unidimensional. The paper that we have referenced explains why this is the case, and for the current paper, we feel that this is sufficient.

As we have stated in the manuscript, we used Andrich's suggestion that the residual correlation between two items should be noticeably larger than other correlations, rather than using a set cut-off value. We have referenced our choice of methodology here.

We have discussed the implications of the sample size in this study in the Discussion (page 10-11).

7. Results: The results for testing for unidimensionality of the pain sub-scale do not suggest that it is unidimensional as 5% is not contained within its confident interval. See: Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Rheum* 2007;57:1358–62.

Whilst we appreciate that Tennant and Conaghan state that the confidence interval for the unidimensionality test should overlap 5%. We understand that the reason for this is to demonstrate that fewer than 5% of persons produce different person location estimates from the two subsets of items. As the percentage of persons with significantly different person locations in our sample is 0.8%, we do not think that the confidence interval needs to overlap 5%. We have used the text quoted below from Tennant and Conaghan to come to this conclusion.

“This test takes the patterning of items in the residuals, examining the correlation between items and the first residual factor, and uses these patterns to define 2 subsets of items (i.e., the positively and negatively correlated items). These 2 sets of items are then used to make separate person estimates, and, using an independent t-test for the difference in these estimates for each person, the percentage of such tests outside the range -1.96 to +1.96 should not exceed 5%. A confidence interval for a binomial test of proportions is calculated for the observed number of significant tests, and this value should overlap the 5% expected value for the scale to be unidimensional.”

8. Results: The overall item fit is inadequate for the subscale appearance.

We thank the reviewer for pointing this out. We have changed the text on page 8 to reflect the poorer fit of the appearance subscale items than the function and pain items.

9. Results: Authors should describe graphical results of Item Characteristic Curve for each item.

We are concerned that to include graphical figures for all 17 ICCs would be excessive and not add greatly to the value of the manuscript. We have however included the ICC for the item relating to having constant pain, which was the only item showing significant misfit on the F-test and borderline misfit on the chi-square test in Table 1. This is now Figure 1 and is referred to on page 8.

10. Results: There is DIF noted, which is a breach of unidimensionality – this challenges the authors conclusion that that a summary score can be calculated from the function subscale items, that the analysis confirms the unidimensionality of all three scales, and that an interval score is produced by the subscales.

We state that DIF can be a breach of unidimensionality in the Methods section on page 6, and further discuss this particular issue around Item 6, on page 9 in the Discussion section. We have now added to the first sentence of this discussion point in order to clarify further than this could be a breach of unidimensionality. However, we do not consider, given the other tests of unidimensionality that we carried out, that the function subscale breaches unidimensionality and have not changed our conclusions

11. Discussion: Many important statements in this section are not referenced, and ought to be, for example: statement that there was sufficient statistical power.

We have cited Linacre 2004 with respect to the power of the sample (page 10). We have added other references where appropriate.

12. Discussion: Authors should discuss how these findings are to be applied.

We have added a paragraph to the Discussion about the application of the Rasch score in other

populations, although we have had to conclude that there is no consensus in the literature here and we do not wish to draw a firm conclusion as to the best course of action. We think this is up to the individual wishing to use the score to decide.

Minor Essential Revisions:

1. The manuscript would benefit from general editing, for example, paragraphs should contain more than one sentence (e.g., Background, Methods), the authors should write in one tense (present or past), there ought to be consistent formatting (e.g., 'Overall fit to the model' in the Methods section: is this a sub-title?), and so forth.

We thank the reviewer for pointing out these typographical errors are have corrected them.

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable