Reviewer's report

Title: A Within-Subjects Trial to Test the Equivalence of Online and Paper Outcome Measures: The Roland Morris Disability Questionnaire.

Version: 1 Date: 4 February 2010

Reviewer: Henrik Lauridsen

Reviewer's report:

Title
A Within-Subjects Trial to Test the Equivalence of Online and Paper Outcome Measures: The Roland Morris Disability Questionnaire.

The article reports on the equivalence of the RMDQ using online and paper formats in a cohort of chiropractic patients and patient support groups. The authors report on equivalence of total scores, missing data rates and participants perceptions on completing online and paper formats.

The study suffers from a serious methodological flaw in the choice of ‘within-subject equivalence’ design. This design is influenced by bias from the ‘memory effect’ and the authors are of the opinion that this is only a minor problem using arguments which are primarily speculative. Furthermore, there are several smaller methodological problems which I have outlined in the ‘Major Compulsory Revisions’ section.

Note:
1:3.5 = page 1, paragraph 3. Line 5 in paragraph 3
1: Design.2 = page 1, Design section. Line 2 in design section

Major Compulsory Revisions

4: Design
I think there are pros and cons for choosing each design. Surely a randomised controlled trial controls for confounding, and the argument used by the authors that ‘one cannot be confident that groups are identical’ is unwarranted. When the sample size is large enough this design will control for biases. The within-subject equivalence design has the weakness that we do not know what happens in between the two scoring times. Do the pain or the way the participants understand the questions change (response shift)? And how much of the equivalence is due to memory effect? I recommend leaving the discussion of why this design was chosen to a section describing ‘Weaknesses’.

4: Design.2
It is unclear what the authors mean by ‘Order of completion was determined by a list of
randomised participant numbers, balanced in groups of 10’. Please clarify. Secondly, this information probably also belongs to the ‘procedure’ section.

5: Participants and Procedures
The authors state that participants scoring # 1 on the RMDQ is equivalent with having LBP. Even though the RMDQ is an excellent tool in this patient group it is problematic to include patients with scores falling within instrument measurement error. It is generally accepted that RMDQ measurement error is around 2-5 points. I suggest omitting patients scoring # 5 as we cannot be sure if it is measurement error or true LBP.

5: Participants and Procedures.1
What is meant by ‘patient support groups’? Is it family and relatives to the patients? If it is, what is the rationale behind carrying out the equivalence test been on these participants? For the equivalence test to be valid it should be carried out on patient populations similar to those it will be applied to. It is not clear from the paper how many participants this group comprises, and I have no suggestions to a solution.

5: Outcome Measures.1
The original RMDQ 24 item does not have ‘yes’ and ‘no’ options, only ‘yes’ options. Thus, the authors did not use the original RMDQ version, but a breed between the RMDQ 23 item version by Patrick et al. and the original.

6: Data Analyses
This section of the paper is long and I suggest shortening it a bit.

6: Data Analyses.3
Since a breed between the original and 23-item version was used, the authors do not state how missing items were handled. Was proportional recalculation used or did the authors just ignore missing items? If missing items were ignored this introduce calculation error which will bias the data. I recommend that the authors describe what they did.

8: Results
I would suggest that you report the mean (SD) time between filling in the two questionnaires. This enables the reader to assess the possible memory effect which can bias the results.

8: Participants.5
The authors state that ‘some of the background data were not collected for 52 participants”; however, this is not clear from Table 1. Table 1 gives the impression that background data was collected for all participants. The reader needs to see what was not collected (i.e. include n for data where participants are missing). In this way he/she can make up his/her mind if it is a problem. Secondly, it would be nice if Table 1 was simplified e.g. give only mean and range for age.
The authors discuss the possibility of their results being an artefact of memory and give three reasons for why this is unlikely to fully explain their results:

Firstly, many participants completed the two versions of the questionnaire in distinct locations at different times (e.g. one at clinic and one at home).

I do not see why locations should have an impact on memory. It is usually time between measurement points that has the most pronounced effect - not location.

Secondly, the researchers observed many of the participants completing the questionnaires and reported little evidence of this response strategy.

I do not see how a researcher can observe a memory effect in participants filling in questionnaires? Perhaps the authors can explain this?

Thirdly, no participants reported doing this in the open-ended questions.

Well no, why should they? They were asked about the advantages/disadvantages of each version of the RMDQ – not if they remembered the questions.

In summary, the arguments given for why memory effect is not a problem are flawed. I suggest the authors reflect more objectively on why memory effect is a possible bias and what the consequences are if it is present. Consequently, the conclusion ought to be more cautious.

Figure 1

I believe this figure is irrelevant. Describing how the electronic questionnaire was designed in the methods section should be enough for the reader to get an idea of what was done.

Minor Essential Revisions

2:keywords: I would probably include the word ‘Roland Morris Disability Questionnaire’ and not ‘outcome measures’. The study is specific to the RMDQ and cannot be generalised to other questionnaires.

3:last paragraph. The section on epidemiology of LBP is irrelevant compared to the purpose of the article. This could be left out. The relevant part is the description of the RMDQ.

10: Discussion

The authors write: ‘…it would be possible to programme future online versions of the RMDQ to prevent missing data.’ What exactly do they mean by that? Is it the design of the questionnaire or is it how the participant proceeds? If a participant is denied to continue if a particular question is not answered this would without doubt reduce the response rate. Please clarify.
**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests.

Yours sincerely

Henrik H. Lauridsen