Author's response to reviews

Title: The size of the treatment effect: do patients and proxies agree?

Authors:

Femke AH van der Linden (fah.vdlinden@vumc.nl)
Jolijn J Kragt (jj.kragt@vumc.nl)
Jeremy C Hobart (jeremy.hobart@pms.ac.uk)
Martin Klein (m.klein@vumc.nl)
Alan J Thompson (a.thompson@ion.ucl.ac.uk)
Henk M van der Ploeg (vanderploeg-stapert@planet.nl)
Chris H Polman (ch.polman@vumc.nl)
Bernard MJ Uitdehaag (bmj.uitdehaag@vumc.nl)

Version: 3 Date: 22 October 2008

Author's response to reviews: see over
Dear Editor,

We are grateful for the opportunity to submit a revised manuscript and appreciate the valuable comments of the reviewers, which we used to improve the manuscript. We were pleased that two of the three reviewers were positive towards accepting the manuscript for publication. The comments of reviewer 1, as well as the other two reviewers, will be addressed in the following pages of this letter.

We hope that the manuscript in its current version can be accepted for publication in the BMC Neurology.

With kind regards,

Yours sincerely,

Ms. F.A.H van der Linden
Reviewer 1: Alessandra Solari

Reviewer's report:

1. The study is of limited size (55 patient-proxy couples) and the population is poorly delineated. It is not at all clear whether these are preliminary results of an ongoing study and, if so, it is not specified the target number of patients to be enrolled, or how this number was determined. No information is given regarding the recruitment period.

The results presented in the manuscript are not part of a preliminary study. The recruitment period was two years: from October 2004 until October 2006. We agree that this was not clearly stated and this information was added to the result section.

2. Attrition was high (24%), with analyses performed on only 42 patient-proxy couples. This high proportion of missing subjects seriously undermines the internal validity of the results.

We agree with the reviewer that the percentage of missing data was high. This was mainly caused by the fact that the MSIS-29 was not returned at follow-up. The high rate of attrition is acknowledged in the discussion section.

3. The authors compared measures of internal responsiveness (“effect size”) and external responsiveness (receiver operator characteristic – ROC – analyses) in people with MS and their proxies. This comparison is useful but only complimentary to the stated objective of the study: not to assess responsiveness of the MSIS-29, but rather the agreement between MS patients and proxies respondents on changes in the 29-item MS impact scale (MSIS-29). It is thus vital to assess inter-rater agreement on physical and psychological score changes, by estimating for example the intraclass correlation coefficient.

We specifically chose not to use an ICC to assess patient-proxy agreement on change scores in this study based on experience of another study1. In that study, proxy respondents appeared to report a larger change than the patients did themselves, but these differences were not significant. On the other hand, ICCs between the change scores on the MSIS-29 were poor, indicating a low level of patient-proxy agreement on the change scores. However, since an ICC is based on the variance of the sample, a lack of variance in change scores could also have caused the low ICCs, rather than lack of patient-proxy agreement. True variability might have been low and by using an ICC we might have only assessed measurement error. Therefore we chose a different approach to calculate t-tests between change scores for the analyses in this study.

1Van der Linden et al. Longitudinal proxy measurements in multiple sclerosis: patient-proxy agreement on the impact of MS on daily life over a period of two years. BMC Neurol 2008, 8:2
Minor points:

1. The abstract is uninformative: no figure is given regarding number or participant characteristics; numerical data are totally absent, even from the results section.
The abstract was extended with the number of participants and other numerical data.

2. In the results section, the descriptive data are insufficient: for the EDSS only median values are given, without any indication of dispersion (range or interquartile range).
The range of median EDSS values were added to the results section.

3. A table reporting the number of patients and proxies scoring minimum and maximum; and the number of missing items at the two MSIS-29 scales is necessary.
At baseline, none of the patients and proxies scored the minimum of zero or the maximum of 100 on the physical or psychological scale.
At follow-up, 1 patient and proxy couple scored 0 on the physical scale. There were no maximum scores on for the physical scale. No minimum or maximum scores were seen for the psychological scale.
The number of missing items for the two scales was as follows:

<table>
<thead>
<tr>
<th></th>
<th>Physical scale</th>
<th>Psychological scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Proxy</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

The number of missing items is very small and therefore we believe that it would not be of additional value to add this to the results section. We hope the reviewers agree, but if the reviewers believe otherwise we will include it in the results section.

4. The fact that only those with more than 50% missing items on both MSIS-29 scales were excluded from analyses (page 8) is a questionable choice, and should be justified with documentation.
The choice of excluding only those patients with 50% or more missing items is based on the HTA report of JC Hobart et al. This was justified with the appropriate reference (17). One patient and one proxy respondent had more than 50% missing items on both scales of the MSIS-29 because they forgot to complete the second page of the questionnaire.

5. Finally, ROC curves should be provided, and the area under the curves reported together with their confidence intervals (or statistical significance). This is important in consideration of the limited numbers and (as stated by the authors) variability of the findings.
ROC curves are provided in the revised manuscript. Initially we chose not to include the ROC curve to minimize the number of figures. AUC values and confidence interval were also included.
Reviewer 2: Michael Hutchinson

Reviewer's report:

This is a well-written report of a study of the concurrent use of a patient report measure, the MSIS-29, by both a patient with MS and their proxy respondent. A group of 42 patients who were treated with steroid therapy were asked to describe also whether they had improved of were unchanged. The authors demonstrate quite remarkable mean agreement between patients and partners in the change/nochange in the MSIS-29 recorded over a quite short time scale of 6-8 weeks. The ROC curves produce remarkably good sensitivity, specificity, NPV and PPV. The value of 7-8 change in the MSIS-29 physical as being the point at which change in the MSIS-29 is clinically significant is in keeping with a previous study. Whether such concurrence in the patient & partner assessments would hold up in the longer term (1 year plus) should be the subject of a further study.

ONLY CAVEAT. The abstract should contain actual results and numbers of patients studied

The abstract was extended with the number of participants and other numerical data.

Reviewer 3: A. Simon Pickard

Reviewer's report:

This is a well-written paper utilizing analytic techniques common to the literature on patient-proxy agreement, and one of the first to be performed in MS.

Minor Essential Revisions:

1. Under methods/procedures. Did all patients and proxies complete the questionnaire at initial visit by themselves, or was there interviewer assistance?
   All the patients and proxies completed the questionnaire by themselves at the initial visit. This was preceded by a verbal instruction of the researcher.

2. Under results, first paragraph. How many patients/proxies declined to participate? [n, %] 
   Unfortunately, data on how many patients declined to participate were not documented.

3. A general observation in the proxy literature is that proxy raters tend to rate patients as having poorer HRQL than patients themselves (Sneeuw et al, 2002). How do your results (e.g. Table 3) compare to studies of proxies in other conditions (e.g. stroke, cancer)? It seems curious that proxy respondents in the "not improved" group reported a 4-point decline in MSIS-29 scores (ES=0.30), in the opposite direction of proxy respondents for the improved group. 
   This was indeed a curious finding, but this does fit into the general observation that proxy raters tend to rate the patient as having al poorer HRQL even when the patients themselves indicate that they did not change.

4. Table 3: Although not statistically significant, effect sizes of difference between changes scores appear to be ~0.30. Note in limitations that you may be underpowered to detect a true difference in scores given the small sample sizes in the subgroups (not improved, improved). 
   We appreciate this valuable addition to the limitation section.