Author's response to reviews

Title: Assessment of fall-related self-efficacy and activity avoidance in people with Parkinson's disease

Authors:

Maria H Nilsson (Maria_H.Nilsson@med.lu.se)
Anna-Maria Holmbäck (Anna_Maria.Holmback@med.lu.se)
Peter Hagell (Peter.Hagell@med.lu.se)

Version: 2 Date: 29 June 2010

Author's response to reviews: see over
Dear Editor,
Please find enclosed the revised version of our manuscript with the title “Assessment of fall-related self-efficacy and activity avoidance in people with Parkinson's disease” (MS: 1133101740372345).

We thank the Reviewers for their comments and suggestions, which we have dealt with accordingly. Below follows our responses and an outline of the modifications made in the revised manuscript (all changes and additions are highlighted in yellow). The corresponding references are to be found within the manuscript.

In addition to the changes outlined above we have made a few minor revisions of language. These are also highlighted in yellow in the new manuscript version.

We hope that the revisions and responses will be sufficient for our manuscript to be accepted for publication in BMC Geriatrics.

All authors have approved the submitted manuscript. We have no conflicts of interest (also stated at the end of the manuscript).

Yours sincerely,

Maria H Nilsson, Anna-Maria Holmbäck and Peter Hagell

Editorial requests:
- The study was approved by the Advisory Committee for Research Ethics in Health Education (Lund). This is inserted on page 8.
- A heading has been inserted before the competing interests (page 16).

Reviewer 1:

1) I would like to see more detailed information on fear of falling in PD in the introduction (personally I feel paragraph 3 is unnecessary).

Authors’ reply:
We have extended the paragraph about FOF in people with PD (page 4) with the following sentences:

“Fear of falling (FOF) and decreased balance confidence is more common in people with PD than in healthy controls (Ref). In a study investigating 119 people with PD, 59% of the participants reported having FOF (Ref). FOF is even more common and pronounced among fallers, and it can cause ADL restrictions and social isolation (Ref). After adjusting for prior falls and PD-specific impairments, FOF has been shown to be a
significant predictor of future falls (Ref). It has therefore been recommended to include FOF when assessing balance performance in persons with PD (Ref).

We find paragraph 3 to be of importance since it describes and defines the investigated construct, and it therefore remains unaltered.

2) In the procedure, could you please state whether the participants were in the on-phase or off-phase. Were all participants assessed in roughly the same phase of their on-off cycle?

Authors’ reply:
This is clarified on page 7 (Procedure), and on page 10 (Results):

Page 7: “When completing the questionnaires, the participants also rated their present mobility as either “good (i.e. on)”, “good but hyperkinetic” or “bad (i.e. off)”.

Page 7: “The participants of the outpatient sample were also assessed clinically. Assessments were scheduled at a time point when they regularly felt at their best.”

Page 10: “At time 1, 64 out of 79 (81%) participants were in the on condition when completing the questionnaires, i.e. mobility was good or good but hyperkinetic. Fourteen rated their mobility as bad. One had missing data.”

3) The statistical analyses could be more concise:
- internal reliability: Cronbach alpha for whole scale + Pearson’s correlations between items

Authors’ reply:
Inter-item correlations are not reported.

- Test–retest reliability: intra-class coefficient between scores obtained at time 1 and time 2

Authors’ reply:
This has been revised as suggested (page 8):
“Test-retest reliability was analyzed by the intraclass correlation (ICC) coefficient between scores obtained at time 1 and 2.”

- Validity: independent t-tests (use parametric if you can, by normalizing your data) examining between-group differences in total scores according to age, etc. You have now also included correlations between the FOF scales and PF, TUGT, UPDRS, etc. It would be better if you would examine between-group differences in total scores according to those measures (by dichotomizing them).
Authors’ reply:

While this certainly is an option it also has disadvantages as it is not the traditional manner in which construct validity is examined (see, eg, Campbell & Fiske 1959), particularly when comparator variables are measured by tools lacking widely acceptable and evidence based cut off scores. Furthermore, comparing subgroups (defined by, e.g., the median) will be even more sample dependent than using correlations since the latter takes the full empirical outcome space into account but the former does not. For these reasons, we have not changed this.

We have not used parametric statistics since the data are not linear but ordinal. This cannot be rectified by normalizing scores.

4) I do not understand how you analysed floor and ceiling effects. Could you explain this more clearly. Also, please include the distribution (skewness).

Authors’ reply: We have revised the description (on page 9):

“Floor- and ceiling effects are thus calculated for those having total scores: the number of participants that score at maximum (or minimum)/ the total number of participants (multiplied by 100) (Ref).”

We do not understand the reviewer’s request of also including the distribution in relation to these aspects.

5) Some general points in the discussion:

Please note that a Cronbach’s alpha should be above 0.70 without exceeding 0.90.

Authors’ reply: There are differences in rules of thumb and the opinion expressed by this reviewer is not shared by everyone, as it is often recommended that the coefficient needs to exceed 0.9 in order for a scale to be acceptable for individual patient assessment (ie, for use in clinical practice). We have added the following on page 12 (discussion):

“Cronbach’s alpha was ≥ 0.95, and it has been recommended to be at least 0.90 in a clinical application (Ref). High alpha values may also indicate a redundancy of items (Ref), suggesting that there may be room for item reduction of the scales. However, further studies in larger samples addressing this issue more specifically are needed to allow any firm conclusions.”

6) On page 13, you present new results. If you would like to expand the scope of your paper to evaluating the level of FOF in people with PD, the paper will need to be rewritten accordingly.

Authors’ reply: We assume that the reviewer refers to the following section on page 14:

“Items of SAFFE which attained the highest scores (“more avoidance”) were items 8 (go out when it is slippery), 10 (go to a place with crowds), 13 (walk half a mile) and 15 (travel by public transportation).”
Although this is not specified as part of the aim, we disagree with the reviewer that these findings are new results as these data are presented in Table 3. We believe that highlighting this information in the discussion provides useful information for clinicians, in addition to psychometric information.

Reviewer 2:
Minor essential revision

1) Abstract
The abstract is very long and not sure whether this complies with Journal’s format. It reports on too much detail, such as gender differences in FOF. The addition in the conclusion, that further studies are needed to interpret the change scores, seems self-evident.

Authors’ reply: According to the instruction, the abstract should not exceed 350 words. In our first version, the abstract consisted of 313 words. We do, however, agree with the reviewer that the abstract could be shortened and that some details may be omitted. We have therefore revised the abstract accordingly and it now consists of 267 words.

2) Background
First paragraph. The study of Latt et al 2009 is missing, which determined fall risk factors in PD.

Authors’ reply: This reference has now been added to the first paragraph on page 4. Latt et al. identified abnormal posture, freezing of gait, frontal impairment, impaired balance and reduced knee extensor strength as independent risk factors for future falls (Ref).

3) Participants
Why were STN-patients included in this sample.

Authors’ reply: STN is an increasingly used therapy for people with advanced PD and, as such, it is an integral part of the profile of the PD population. Excluding surgical patients from this study would consequently leave us questioning whether the present results could be generalized.

4) Methods
Both scales include items that may not have been relevant to the daily activity profile of the patients, such as cleaning and cooking. What happened if patients failed to fill in these items.
Authors' reply: If the patients failed to fill in any of the included items, this is reported as missing data (see Tables 2 and 3). Prepare simple meals (item no 3 in FES(S) and item no 11 in SAFFE) had no missing values. Clean your house (item no 2, SAFFE) had one missing value, whereas item 10 of FES(S) (i.e. “Housecleaning”) had no missing values. The present results thus speak against that the included items would be perceived as irrelevant, but further qualitative work may be needed in order to elucidate these issues.

5) Discussion
(Former) Page 12, I would not recommend to use the Rasch as a “modern” psychometric Analysis

Authors’ reply: While we agree (given its 50 years history), Rasch analysis (together with IRT) is typically referred to as “modern test theory” (MTT), in contrast to “classical test theory” (CTT). The distinction lies in the reliance on linear associations (correlations) that is central to CTT but irrelevant in MTT. We have therefore not changed this in the revised manuscript.

Major compulsory revisions

6) The background does not explain or justify which aspects of validity were addressed in the study.

Authors’ reply: We have now added brief comments regarding this (as well as regarding reliability) to the manuscript. A new paragraph has been inserted (Background) on page 5.

“This is unfortunate since traditional indices of such properties (e.g. scaling assumptions, reliability and validity) are sample dependent. A reliable score contains little measurement error and is reproducible. This can be assessed in various ways, of which Cronbach's alpha and test-retest reliability are the most common (Ref). Construct validity relates to the extent an instrument produces scores that are representative of the variable it is intended to represent. It includes convergent validity and divergent validity, which can be assessed by examining the pattern of correlations with other variables. It also includes whether scores distinguish between groups that that are expected to differ in relation to the investigated construct (Ref).”

7) Participants and methods
It seems crucial to include patients that were non-demented as the study consisted of a postal survey. How and by whom was it decided that patients were ‘non-demented’.
Authors’ reply: This section is now revised (page 6):
“Data were collected from two samples including 79 people with clinically diagnosed idiopathic PD (Ref) (Table 1). All participants were recruited from a Swedish university hospital, where the patients are regularly evaluated (clinical tests and questionnaires) as either in- or outpatients. Selection of participants was done by the treating neurologist or by a specialist nurse working within the same Movement Disorders team. Dementia was an exclusion criterion for both samples.

One sample consisted of 50 people with varying degrees of motor complications and/or gait disturbances, receiving out-patient care at two movement disorder neurologists’ clinics during one year. Eight patients declined participation and five patients were unable to attend the study visit. The remaining 37 participants responded to the included questionnaires and they were also assessed clinically.

An additional sample consisted of 42 people, who had received and four who were on the waiting list for Deep Brain Stimulation (DBS) in the subthalamic nucleus. They were invited to participate in a postal survey. A total of 93 patients were eligible, but 47 were excluded for the following reasons: age >75 years, dementia, or previous participation in the first sample. Forty-six patients were invited to participate in the survey and 42 consented (Table 1). “

All of the included participants are thus regularly evaluated, and each participant is well known and previously investigated by the person performing the selection procedure.

8) It is stated that neurologists invited patients to participate, based on which inclusion criteria? A selected cohort has its drawbacks.

Authors’ reply: See above (number 7, Reviewer 2).

9) A major drawback of this study is that it is based on a postal questionnaire methodology, which makes it difficult to control for patients’ cognitive difficulties, true understanding of the scales and control for medication intake. Being in on or off may make a crucial difference in perception of FOF. However, reliability statistics were clearly not influenced by this problem, which is suprising. Please comment on these issues in the discussion.

Authors’ reply:

We do agree that being in on or off may influence their ratings. However, our personal experiences of using FES(S) as an interview administered version is that people with PD tend to ask whether they are to rate in relation to their “off-state” or “on-state”. When completing FES(S) and SAFFE, the participants were therefore in the present study instructed to make a general estimation when rating. This is now clarified on page 6:
“When completing FES(S) and SAFFE, they were instructed to make a general estimation when rating.”

See also our response to Reviewer 1 (number 2).

This is also addressed in the Discussion (page 15):

“It is also conceivable that responses may be influenced by whether respondents with PD are “on” or “off” while completing questionnaires, which could not be controlled for here since data collection was conducted by postal surveys. However, the high test-retest coefficients argue against this being a major source of bias.”

10) One of the problems with the FES is that when people have to fill this in by themselves they easily interpret their ‘confidence’ to carry out the 13 activities as:’’their ‘ability’ to carry them out. In my view this asks for an interviewer-based administration. I miss any interpretation about self-assessment via postal questionnaires in PD or any awareness in the discussion that this may have been a drawback. It questions the validity of these instruments.

Authors’ reply:
To our knowledge there is no published study that has investigated whether a self-administered version of FES(S) renders different results compared with using an interview administered version when investigating people with PD. This would be an interesting study to perform in the future. This is now acknowledged in the Discussion (page 15):

“Finally, it should be borne in mind that the data presented here was collected through self report and it is unknown whether responses are influenced by mode of administration. The equivalence between self completion and interviewer based administration of the FES(S) and SAFFE needs to be assessed in additional studies.”

11) (Former) Page 8 of the analyses section discusses which aspects of construct validity were under scrutiny. This and some of the hypotheses on subgroup differences may be better placed in the introduction.

Authors’ reply:
We do not agree but acknowledge that this is a matter of tradition and style, and will be happy to change this if the Editor insists.

12) The calculation of the SEM as a measure of error is limited. I would recommend to calculate responsiveness of the tests by using the Smallest Detectable Difference (SDD), the SDD is calculated on the basis of the standard error of measurement, (*1.96*#2) assuming that the measurement errors are constant across the range of possible scores. In order to allow comparison of responsiveness between the 2 tests, the Reliable Change Index (RCI) can be determined for each measurement by calculating the
SDD as the percentage of the maximal feasible score.

**Authors’ reply:**
This is not uncontroversial and it is more complex than what is implied by the Reviewer. We have not made any changes for the following reasons: (1) the SDD is directly based on the SEM ($SDD = 1.96 \times SEM \times \sqrt{2}$) and is therefore easily computed from available data by anyone interested in doing so. (2) There are inherent problems related to the interpretation of SDD since SEM is not symmetrically distributed around the observed score (but relates to the latent true score). (3) A number of studies suggest that the SEM is a reasonable estimate of the minimal important difference. This is briefly acknowledged in the Discussion (page 12-13):

“In this study, we used the SEM which expresses the error of the hypothetical (unknown) “true score” based on the observed raw scores (Ref). Although not unchallenged, a number of studies have suggested that the SEM is a reasonable estimate of the minimal important difference (Ref).”

**13) Discussion**
There is no further interpretation of the Cronbach alpha results. What do the authors recommend, is summing the scores allowed and valid? And is further work needed?

**Authors’ reply:**
We have extended the section about reliability (on page 12, Discussion):

“Our observations provide support for the reliability of SAFFE and FES(S) scores when investigating FOF in people with PD, as alpha and ICC values of both exceeded the recommendations of 0.80 (Ref). Cronbach’s alpha was $\geq 0.95$, and it has been recommended to be at least 0.90 in a clinical application (Ref). High alpha values may also indicate a redundancy of items (Ref), suggesting that there may be room for item reduction of the scales. However, further studies in larger samples addressing this issue more specifically are needed to allow any firm conclusions.”

The first paragraph of the discussion has been rewritten: “This study provides initial support for the scoring assumptions (i.e. summing scores), validity and reliability of FES(S) and SAFFE among people with PD.”

Scoring assumptions thus include summing the scores, which has been clarified in the revised version.

**14) There is no attempt made to compare the scales. Is it better to measure activity avoidance than feelings of insecurity? (see comment earlier on the difficulties to keep patients focused on their feeling of insecurity).**

**Authors’ reply:**
On page 13 (Discussion) the following is stated:
“The correlation coefficient between FES(S) and SAFFE was -0.74. In accordance, about 55% of the variance in scale scores can be explained by the other variable. This suggests that although fall-related self-efficacy and activity avoidance are related, they are not interchangeable constructs, and scale selection should be based on the objectives at hand.”