Author's response to reviews

Title: Cross-cultural adaptation and reliability and validity of the Dutch Patient-Rated Tennis Elbow Evaluation (PRTEE-D).

Authors:

Mathijs van Ark (m.van.ark@umcg.nl)
Johannes Zwerver (j.zwerver@umcg.nl)
Ronald L Diercks (r.l.diercks@umcg.nl)
Inge van den Akker-Scheek (i.scheek@umcg.nl)

Version: 2 Date: 13 September 2013

Author's response to reviews: see over
Author’s response to reviews

Title: Cross-cultural adaptation and reliability and validity of the Dutch Patient-Rated Tennis Elbow Evaluation (PRTEE-D).

Authors:
Mathijs van Ark (m.van.ark@umcg.nl)
Johannes Zwerver (j.zwerver@umcg.nl)
Ron L. Diercks (r.l.diercks@umcg.nl)
Inge van den Akker-Scheek (i.scheek@umcg.nl)

Version: 2 Date: 14 September

Author’s response to reviews: see over

The reviewers provided us with useful feedback and suggestions. We think that these suggestions contributed to an improved version of this manuscript.
Reviewer's report

Title: Cross-cultural adaptation and reliability and validity of the Dutch Patient-Rated Tennis Elbow Evaluation (PRTEE-D).

Version: 1  Date: 21 May 2013

Reviewer: Daniel McWilliams

Reviewer's report:

Major Compulsory Revisions
I think that the inclusion of a substantial group of normal people must have biased the assessment of validity and reliability. Most of the normal people scored a very low or non-existent LE score, whereas the LE patients had much higher scores. The Altman-Bland plot in of repeated PRTEE-D measurements shows highly consistent normal scores, followed by more variable LE scores. Additionally, many of the LE scores fall outside of the limits of agreement. I would like to see how the LE and normal groups performed separately. Perhaps different symbols on the Altman-Bland plot could be used. I would also like the authors to state how reliable the PRTEE-D is for LE patients only.

The reviewer rightly states that the inclusion of normal people influenced the assessment of reliability. It was a good suggestion to provide the Bland-Altman plot for LE patients separately as well. Additional analyses for test-retest reliability (ICC's and Crohnbachs alphas) are now also provided. By providing Bland-Altman plots for the complete group as well as for LE patients separately, this limitation of the study is further tackled. The data of LE patients only show that all but one of the measurements fall within the Limits Of Agreement. The outcome of the analyses for LE patients only did not differ much from the whole group. As to be expected the ICC scores were somewhat lower (total = .88, function = .89, pain = .78) for the LE patients only, but the ICC values were still excellent. Five questions dropped to fair to good correlations, but were still all higher than 0.59. Crohnbachs alpha values decreased, but these values are still excellent. Statements about the reliability of the PRTEE-D for LE patients only are now included in the manuscript. The validity is not biased by the inclusion of normal people in this study, because only data of the LE patients were used for the assessment of validity.

Minor Essential Revisions

In the Introduction, please state clearly whether the PRTEE questionnaire is designed to screen for LE or assess the severity of LE. The manuscript implies that PRTEE is designed to assess severity. Please confirm this.

This is correct, this is now more clearly stated in the manuscript.

Please briefly give more details of the ethical approval and informed consent process.
More details of the ethical approval are provided in the method section of the manuscript. The Medical Ethical committee of the University Medical Center Groningen reviewed the study protocol and concluded that the study was not subject to the Medical Research Involving Human Subjects Act. No formal ethical approval was therefore needed. All participants received the PRTEE-D questionnaire with an accompanying letter, informing about the study and its goals and explaining that return of the questionnaire will be taken as consent to participate.

During the one week reliability assessment, do we know if the patients’ LE severity changed? For example, did some people feel that they were having a ‘bad day’? The reliability data are based upon the assumption that the condition is consistent across the 1-week time-span. Are there any data to support this?

To our knowledge, no 1 week follow up data exists in LE patients. Besides the PRTEE, DASH and VAS pain, no other data was recorded in this study to assess a possible severity change. LE patients showed a 3.7 ± 8.9 point mean decrease in PRTEE score. This is in accordance with an approximate 4/100 decrease in VAS-pain during one week (calculated from a one month follow-up) (Smidt et al, 2006). Despite these possible minor changes, we chose a one week period in this study to prevent ‘copying’ from a subject’s memory (Deyo et al 1991).

Was this study a community-based or hospital-based study? How were the LE patients recruited? Are they a representative samples?

Physiotherapists, general practitioners and sports physicians in and in the area around the University Medical Center Groningen were contacted. Patients diagnosed with LE were asked to participate in the study. Since patients contacting different (para)medics were included in the study, this seems to be a representative sample.

Make sure that Methods and Results are written in the past tense.

This is adjusted in the manuscript.

Discretionary Revisions

Propose a strategy about how to score people who avoid performing tasks with their affected arm. Should they be given a maximum score for that item, or is it more appropriate to leave the value missing?

A strategy is proposed in the discussion section on this subject. We think the best solution for this, is to use the average score of the subscale for this question, like stated in the PRTEE manual for missing questions (MacDermid, 2010). Another option, use the maximum score, seems less suitable because patients are most likely able to do the activity without unbearable pain but choose the less painful option (using their other arm).
Critically compare the methodology between this study and those that established PRTEE questionnaires in other languages.

The differences in methodology between this and other studies investigating the reliability and validity of the PRTEE is more elaborately described in the current version of the manuscript. Next to the already described differences in follow-up time and the use of healthy participants for the test-retest reliability part, other items of the methodology are discussed as well now.

Discuss why question 4 on the PRTEE-D pain subscale showed much less reliability than all of the other questions?

It is indeed notable that question 4 showed less reliability than the other questions. We don't have a good explanation why this is the case. This was not the case in other studies which investigated ICC per question. It would only be speculative to provide an explanation for this.

Level of interest: An article of limited interest

Quality of written English: Needs some language corrections before being published

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

Declaration of competing interests: I have no competing interests within the scope of this study


Reviewer's report

Title: Cross-cultural adaptation and reliability and validity of the Dutch Patient-Rated Tennis Elbow Evaluation (PRTEE-D).

Version: 1 Date: 2 August 2013

Reviewer: Angelo Cacchio

Reviewer's report:
The authors adapted the Patient-Rated Tennis Elbow Evaluation (PRTEE) Questionnaire into Dutch and tested the measurement properties of the adapted questionnaire.

There is a clear need to adapt HRQOL scales to other cultures, so the focus of the manuscript is important.

The process of cross-cultural adaptation used here appears to be rigorous and consistent with previously published guidelines.

I commend the authors on a nicely written paper on a highly relevant topic.

Major Compulsory Revisions:

1. To highlight the two components of reliability, I'd suggest to reformulate the paragraph on page 7 as follows: "Reliability is a generic term used to indicate both the homogeneity (internal consistency) of a scale and the reproducibility (test–retest reliability) of scores (Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. Statistic and strategies for evaluation. Contr Clin Trials 1991;12:142S-158S.)."

   This useful addition is now added to the reliability section.

2. I suggest to indicate the model (2, in this case) and form (presumably 1) of the ICC (e.g., ICC(2,1))

   The model and form of the ICC are now indicated in the manuscript.

3. In addition to the ICC, please report the SEM and MDC

   SEM and MDC values are now also provided in the results section.

4. I suggest to define as the result of internal consistency was interpreted: (e.g., "Internal consistency is considered excellent when Cronbach’s alpha exceeds 0.80, adequate when it falls between 0.70 and 0.79, and inadequate when it is lower than 0.70 (Andresen EM. Criteria for assessing the tools of disability outcomes research. Arch Phys Med Rehabil. 2000;81 Suppl 2:S15-S20.)."

   Good suggestion, criteria for the Cronbachs Alpha values are now included in the method section.
5. Among the psychometric properties of the adapted questionnaire, the authors assess the reliability and validity, the establishment of which are necessary for the questionnaire to be widely adopted. Although the quality of measurement questionnaires has usually been evaluated by considering the reliability and validity of such questionnaires; it has, however, been suggested that responsiveness should be another criterion in the choice of a measurement questionnaire (Guyatt GH, Kirshner B, Jaeschke R. Measuring health status: what are the necessary measurement properties? J Clin Epidemiol. 1992;45:1341-1345.). A lack of responsiveness measure in the validation of a translated and adapted questionnaire is a limitation, and I would suggest that this limitation is highlighted in the Discussion paragraph.

The responsiveness of the PRTEE-D is indeed still to be investigated. This is a good point and is now added to the discussion section as a limitation.

6. Also the assessment of factor validity (factor analysis) was important in this type of study and also this lack should be addressed as a limitation of the study in the Discussion paragraph.

A factor analysis is now provided in the manuscript.

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** Yes, but I do not feel adequately qualified to assess the statistics.