Author's response to reviews

Title: Testing the proficiency in distinguishing locations with elevated plantar pressure within professional groups of foot therapists.

Authors:

Nick A. Guldemond (ngu@orthop.unimaas.nl)
Pieter Leffers (p.leffers@epid.unimaas.nl)
Fred H.M. Nieman (fni@oazg.azm.nl)
Antal P. Sanders (asan@rev.azm.nl)
Nicolaas C. Schaper (nsc@sint.azm.nl)
Geert H.I.M. Walenkamp (gwa@sort.azm.nl)

Version: 4 Date: 4 November 2006

Author's response to reviews: see over
Dear Ms. Moylan,

We would like to thank the editorial board and the referees for the helpful comments on our manuscript "Testing the proficiency in distinguishing locations with elevated plantar pressure within professional groups of foot therapists."

We have carefully considered the comments of the reviewers and we have changed the manuscript. Please find the list of the comments and changes below.

The changes made in the manuscript are highlighted in yellow. We hope that the revised manuscript is now appropriate for publication in 'Musculoskeletal Disorders'.

We thank you for your interest in our article.

Yours sincerely,

Nick Guldemond, clinical researcher
Also on behalf of the fellow authors

Dept. Orthopaedic Surgery
University Hospital Maastricht
P.O. Box 5800
6202 AZ Maastricht
The Netherlands
Phone: +31 (0)43 3875031
Fax: +31 (0)43 3874893
Mobile: +31 (0)6 42393696

e-mail: ngu@orthop.unimaas.nl

Encl.
- Revised manuscript
- List of the changes in the manuscript.
Our reactions to the issues the reviewers have raised, together with the changes made in the manuscript are listed below.

Comments Keith Rome

The authors have made a substantial effort in responding to all my comments. I am happy with the comments although there are some minor typo errors - change devises to devices.

We corrected the text.

Comments Margaret Wan Nar Wong

Reference list

We corrected the reference list.

Comments Anna Hart

General

I am not familiar with the statistical package used.

The chapter ‘Generalizability theory’ of the latest edition (3rd) of Streiner and Norman’s ‘Health Measurement Scales’, contains a reference to the statistical software GENOVA: (Brennan, R.L. 2001. Generalizability theory. Springer-Verlag, New York). After consultation with Crick and Brennan, we will be happy to provide this software.

It is difficult to tease out what the data mean in this study (except that there does not appear to be good agreement BETWEEN professional groups). Rather than aiming to get ICC values, the authors should also devote some effort to describing the data, and the nature of the disagreements. To my mind, it is difficult to discuss agreement between groups without also describing the level of agreement within those groups.

We elucidated the concept ‘method agreement’ as multivariate measure, which taking into account variance between therapists and variance between ratings and golden standard by rewriting the pertaining paragraph in the results section. In addition, we supplemented results of a more conventional analysis of within group or interrater agreement.

1. I found the paper curious in that the title refers to ‘within groups’ agreement, while the text only seems to refer to comparisons between professional groups. I would recommend that either the title be changed, and a rationale given for not checking agreement within groups, or that within groups analyses are also done (It is possible for readers to estimate this from the Figures, but this is difficult).

We concur with the reviewer’s remark that "...the title refers to 'within groups' agreement, while the text only seems to refer to comparisons between professional groups.". This impression may exist, because during the previous editing of this manuscript we have decided to omit ICC results to avoid too much complexity. These ICCs containing the method agreement between the proportions of elevated pressure as measured by the 'gold standard' and the ones found by subjective professional judgments, specified for the 3 types of therapists and the 6 types of foot regions.

These ICC’s were averaged over both feet, averaged over the ten therapists per group and averaged over the 3 patients. See table 4. Reported in the text were only the ‘marginal’ ICC’s of this table (averaged over the six foot regions) and the overall method agreement (next to regions averaged over all three groups). We understand that this may gave the impression
that Table 6 (the estimated mutual agreement between the 3 groups) is all that was analyzed within the ANOVA model we used.

The 'marginal' ICC's reported in the text concern the relevant variance components of the patients*method*therapist-interaction for each of the 3 types of therapists. Next to this, their higher interactions patients*method*therapists*regions, patients*method*therapists*side and finally, the patients*method*therapists*regions*side components are also relevant in calculating the method agreement ICC. The summation of these 4 sources of potential error are contrasted to the summation of 12 other variance components, like the one for patients, the one for regions within patients, the one for side within patients, the one for patients, etc. So, the method agreement ICC for one group of therapists is the ratio of the 12 sources of patient variance components not-containing the patients*method*therapists-component divided by the sum of all 16 patient variance components. We include these formulas in the appendix of the manuscript.

Fact is, that within-therapists variations have been included in the relevant variance components used in the ICC calculation. To avoid the impression mentioned above we have at first decided to change the title of the manuscript into: "Testing the proficiency to distinguish locations with elevated plantar pressure within and between professional groups of foot therapists", and secondly, to provide the reader in the text with a guideline (2nd paragraph in the results section) how to interpret the 'within groups' ICC's. This way, the title will correspond more clearly with the results and more emphasis is being given to the results on the agreement of therapists with the 'gold standard'. For this, table 3 is also helpful for ICC interpretation.

Apart from this, we have observed that a few errors in the calculation of the 'marginal' and overall ICC's have occurred. Some of the patient-related variance components had not been included within the formulas. This is now corrected in the new text, and the overall 'marginal' and overall ICC's appear to lie slightly higher than in the previously reported ones. With apologies for the possible inconvenience.

2. The researchers have defined a 'gold standard'; however, this appears not to be free from controversy, and I wonder how golden it really is. This warrants more discussion in the paper. I suspect that this does not affect the results in Table 4, but this needs to be made explicit. It certainly affects the interpretation of Table 3. The authors should discuss this.

The 'gold standard' used in the manuscript is formed through measurements using a pressure sensitive platform. The technical validity of these devices and reliability of methods for obtaining plantar pressures are evaluated in numerous studies which appear to be rather good. One can discuss the cut-off criterion (higher than 700 kPa for peak/elevated pressure) used in this study or the assumption of the relation between plantar pressure and foot pathologies. We addressed these issues in the fourth paragraph of the discussion section.

3. I find Table 3 unhelpful. Proportions may be similar, but this would not constitute agreement. For example, consider 3 people X, Y and Z rating 3 objects A, B and C as high or not high. X selects A only as high, Y selects B only, and Z selects C only. All 3 people have selected 1/3 of the objects, but there is no agreement between them. Moreover, in the data presented, the ratings are averaged over feet – how do we interpret this? These data need to be removed or discussed properly.

The results of Table 3 indeed show only the shift in marginal proportions between the results of the 'Gold standard' elevated pressure at one hand and the scores of judgment of elevated pressure at the other. This shift can also be a interesting parameter: in our results one can see from Table 3 that therapists generally tend to overestimate the elevated pressure in patients. This result has been added to the text. Next to this, internal relationships between the proportions tell us something about the actual method agreement between both ways of scoring. The remark about the fact that ratings are averaged over both feet merely
represents that we have analyzed the ICC's within a multivariate ANOVA model. Removal of the results of Table 3 does not seem to make the text more clearer, so we have tried to give more explanation what is presented as results.

4. Table 4 may be correct if the correct formula has been used. The authors should state explicitly which formula for ICC, from Streiner and Norman, was used. (It may even be helpful to include a table of the estimated variance components, although it might be unwieldy). I think I understand what is meant by 'a randomly chosen therapist' in the paragraph 'Statistical analysis', but it is not very clear. Readers need a clear definition of what sort of agreement is being measured. I think the authors are saying that they are estimating agreement between raters in general, not the specific people that were in their study.

We include the formulas in the appendix of the manuscript. We have added a line in the text, that interested readers may ask for further information on the calculation and estimation of the ICC's presented. The paragraphs of Streiner and Norman concerning ICC calculation and estimation are very hard to summarize in the manuscript, because in their text a methodology of generating formulas of multivariate ICC's is developed. Classical measurement theory is here extended to Generalizability Theory using mixed-model multivariate ANOVA.

The interpretation on the type of agreement made by the reviewer is quite right. We use the generalized version of method agreement, in which raters in the study are assumed to be representative for a larger group of raters (sometimes named as interrater agreement, Portney & Watkins, 1993, page 509-516, ref. #40). We have added a line in the text to make this more clear to the readers.