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: 3 Date: 1 August 2006

Author's response to reviews: see over
To: Elizabeth C Moylan PhD  
Senior Assistant Editor  
BioMed Central  
32-42 Cleveland Street  
London W1T 4LB  
UK

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

Abstract:
1. Re-word ‘in the way they usually do’ to “in everyday clinical practice”.

We changed the text as follows (abstract and methodology section): ‘Therapists were instructed to examine the patients according to the methods used in their everyday clinical practice.’

2. The results section should include p-values.

We supplemented the following text to the results section: ‘The differences in portions of elevated plantar pressures between the GS and the ratings per discipline show significant discrepancies (all p-values < .05).’

Background:
3. Re word ‘felt-therapy’ to “padding and strapping therapy.”
4. Re-word ‘shoeware’ to “footwear”.

We changed the text as follows: ‘Off-loading strategies could include padding and strapping therapy, foot orthoses, therapeutic footwear or total contact casts.’

5. Reference(s) required after various techniques such as footprints…….

We added the following references:

6. Can the authors please explain the ‘amount’ of plantar pressure? The evidence about 700KPa as the threshold needs to be fully justified.

We changed the text as follows: ‘Advanced pressure sensitive devises i.e. platforms and in-shoe devises, provide information about the location as well as the magnitude of plantar peak pressure’

Seven hundred kPa as cut-off criterion for categorizing peak pressure into elevated or non-elevated pressure was based on values which are frequently cited in the literature pertaining diabetes, rheumatoid arthritis and metatarsalgia. Studies in healthy subjects showed that peak pressures were significantly lower than 700 kPa. We are aware that 700 kPa as a risk-threshold is based on weak evidence. We deal with this issue in the discussion section and show the results of lower cut-off points.

We changed the text as follows and added the following references: ‘In the literature, peak pressure values similar or higher than 700 kPa are frequently considered as abnormal or elevated. Also, studies in healthy subjects showed that peak pressures were significantly lower than 700 kPa.’

We adopted the suggestion of the reviewer.

Bare foot peak pressures were estimated per foot, by calculating the mean over the readings of 5 consecutive measurements. After instruction and practice of the plantar pressure measurement procedure, data of five steps per foot were collected. The procedure for masking was according to the preceding BMC paper, but now applied to bare foot plantar pressure data.

We changed the text as follows: "Bare foot peak pressures were estimated per foot, by calculating the mean over the readings of 5 consecutive measurements. This was done for six discrete regions: big toe (BT) and metatarsal one (mt-1) to five (mt-5) defined through Novel 'create any mask®' software and verified through anterior-posterior radiographs."


10. I cannot comment on the statistical tests. I assume the data was normally distributed to use a mixed-model ANOVA. Please explain the principles of Streiner and Norman.

We would kindly suggest chapter 8 ‘Generalizability Theory’ of the excellent book ‘Health Measurement Scales: a practical guide to their development and use’ by Streiner D.L. and Norman G.R. (1995). A summary of the described principles is no alternative for their concise and clear text. Although we think it would not be suitable to integrate the following text in the present paper, we shall attempt to elucidate the principles described by Streiner and Norman:

Streiner and Norman’s calculation of the ICC (intraclass correlation coefficient, which can be used as a measure of method agreement: ICC = R), involves the definition of individual patient variation plus
the definition of specific patient variation originating from differential methods. The individual patient variation can be estimated by the variance component for random factor patient i.e. $\text{Varcomp} (P)$, the specific patient variation originating from differential methods can be estimated by the variance component for the interaction of the random factor patient and the fixed factor method i.e. $\text{Varcomp} (Pm)$.

The ratio of $\frac{\text{Varcomp} (P)}{\text{Varcomp} (P) + \text{Varcomp} (Pm)}$ results in the ICC for method agreement. However, in multivariate Generalizability Theory, the calculation can be more complex because other factors besides $P$ and $m$; may be involved in the ANOVA design: in the present study ‘therapist’ within discipline, ‘side’ within patient and ‘region’ within foot.

Streiner and Norman suggest a strategy to incorporate variance components involving interaction between patient and aforementioned factors in both numerator and denominator of the ratio. Define $X$ as the set of other factors. All variance components involving factor $P$ are used in the calculation of the concerned ICC or so-called rho coefficient:

$$R = \frac{\text{Varcomp} (P) + \text{Varcomp} (Px)}{\text{Varcomp} (P) + \text{Varcomp} (Px) + \text{Varcomp} (Pm) + \text{Varcomp} (Pmx)}$$

Where

$P$ = patient or subject

$x$ = other relevant factors, such as discipline, therapist, side (left/right), region (within foot), etc.

$m$ = method i.e. Gold standard vs. clinical rating

11. It is interesting to note that the authors decided to accept 0.80 as being acceptable. Previous authors have used a less robust system, where 0.6 is acceptable. For example, Portney and Watkins [2] suggest that an ICC > 0.75 is deemed good; between 0.50-0.75 as moderate and <0.50 represents poor reliability. Bruton [3] suggests that an intra-class correlation of at least 0.6 is to be considered useful.

We agree there are no absolute standard values for acceptable reliability using the ICC. The whole context of measurement and type of discipline determines how much reliability is needed. Portney and Watkins indeed state that ‘.75 is indicative of good reliability’, but they continue with the statement that ‘for many clinical measurements reliability, reliability should exceed .90 to ensure reasonable validity’ (p. 565). As we point out in the second paragraph of the discussion, we consider that identification of locations with elevated pressures is an important aspect in foot care decision-making.

Results:

12. I am unclear of Table 3. Please can explain in more detail the concept of observed proportions of elevated pressure? For example the GS for the big toe is 0.50. What is the SI unit?

The figures in table 3 represent dimensionless proportions, which allow to compare the ratio of regions with elevated pressure according to the GS and ratio of regions indicated by therapists. We supplemented the legend of table 3 with the following text: ‘The proportion of elevated plantar pressure per region according to the GS is determined by the number of regions with observed elevated pressure to all regions for that location: for example, three of the six BT regions (three patients both feet) had elevated pressure i.e. $3:6 = 0.50$. For each region, the ratings per discipline were averaged over both feet of the patients.’

13. Please delete the word ‘sometimes’ and ‘most of the time’ with appropriate scientific language.

We changed the text as follows: ‘On average, one can see an underestimation in the big toe region and an overestimation in the other regions when compared to the GS. The standard deviations indicated large variation around the point estimate.’

14. The estimated method of agreement (ICC) should include confidence intervals.

Unfortunately, we cannot comply with this request. When only ‘patient’ and ‘method’ are used as factors in Generalizability tests, one can fall back on standard errors and confidence ranges, like for instance Fleiss (1986) mentions. This involves Satterthwaite’s 1946 approximation of the standard error. In multivariate designs, like the six factor design used here, it is only possible to calculate the estimated standard error of the variance components

15. Table 4 reports ICC high correlations – 0.70, 0.64 and 0.78 but you state in the results section that all measures were below the critical level. Please can you explain? Furthermore, you state ICCs of 0.46, 0.51 and 0.60 – how were the figures derived?

For the argumentation for use the critical level, please refer to comment 11. The figures were derived through ratios of variance components estimated with mixed model repeated ANOVA models.

Discussion:

16. 2nd line has a question mark – please explain or delete.

We removed the question mark.

17. I am still unclear on how the data extraction has lead to the conclusion that the big toe is underestimated? Please clarify.

The pertaining figures in table 3 show lower rates i.e.: 0.43 ± 0.50 (podiatrists), 0.45 ± 0.50 (pedorthists), 0.18 ± 0.39 (orthotists), compared to the rate based on the GS measurements: 0.50. The overall (disciplines) rate 0.36 ± 0.48 is also lower than the GS ratio.

18. There is further discussion of the current results in diabetics. The current study only evaluated two healthy older subjects and a further subject with psoriatic arthritis. I am concerned there is too much speculation from the authors about diabetic feet compared to the current study.

We supplemented the following text to the discussion: ‘In the present study three patients with metatarsalgia were evaluated. Studies in patients with other foot pathologies are warranted to put our findings in the perspective.’

19. Page 6 – first line should be either in the introduction or methodology. Again the study was based on patients with diabetic neuropathic foot ulceration. Does this critical value change considering the type of patients used in current study?

In accordance with the context of the preceding text, this sentence refers to the difference regarding plantar pressure between dynamic and static measurements. We do not think that this phenomenon will be significantly different for other patient groups. As Cavanagh stated, "...the standing foot should always be seen in the context of the pressures and loads experienced by the foot in dynamic activities which the patient may be capable of performing. It is likely that any abnormalities detected in the pressure distribution during standing will be amplified by other more stressful activities of daily life..." (Cavanagh, Pet al. 1987. "Pressure distribution under symptom-free feet during barefoot standing." Foot Ankle 7:5, 262-76). We also would like to refer to the comment on the issue under point 6 and the fourth paragraph of the discussion.

20. The first paragraph needs further clarification – the statement of 5 instead of 3 feet and false-negative needs to be explained.

We adopted the suggestion of the reviewer and we supplemented the following text to the first paragraph of the discussion: ‘For example, the proportion of elevated plantar pressure per location according to the GS is determined by the number of regions with observed elevated pressure to all regions for that location: for BT, three of the six regions (three patients both feet) had elevated pressure i.e. 3:6 = 0.50. The proportion for BT based on ratings by the three disciplines was 0.36, indicating an underestimation.’
We supplemented the following text to the second paragraph of the discussion: ‘Nota bene: false negative means that a region was indicated as one without excessively high pressure, while the actual peak pressure was above the criterion level i.e.: 700 kPa and vice versa for false positive.’

We supplemented the following text to the fifth paragraph of the discussion: ‘If we lower the cut-off point to 600 kPa, than the GS proportion of elevated pressure in the big toe region would increase. This is illustrated by a shift of the cut-off point from 700 kPa to 600 kPa in figures 2 to 5. As a result now five BT regions are categorized as elevated, instead of three regions. This leads to a larger underestimation or false-negative rate: 5:6 = 0.83 versus 0.36.’

21. The authors suggest that other pressure variables maybe important. I suggest to the authors they fully justify the current data of only peak pressure.

We used peak plantar pressure because it is the most studied measure which is associated with plantar tissue stress and foot pathology. Also peak pressure and pressure time integral for a specific region, are strongly correlated. Moreover, the location of peak pressure and the location with the highest pressure time integral is identical. We supplemented the following references:


Conclusion:
22. The authors state no differences between professions - is that a significant difference? If so, please state.

We stated that …‘There appears to be no difference in the proficiency to distinguish regions with elevated plantar pressure between podiatrists, pedorthists and orthotists.’… We consider these small differences irrelevant and therefore should not to be subjected to statistical testing, since they are all far below the criterion level. We have changed the text, so now it reads: …‘There appears to be no relevant difference in the proficiency to distinguish regions with elevated plantar pressure between podiatrists, pedorthists and orthotists.’…
23. This paper seems to be written with data collected during the same study on 'Comparison of foot orthosis made by podiatrists, pedorthists and orthotists regarding plantar pressure reduction in Netherland', which has been published in BMC in 2005. The introduction and methodology parts closely simulates the previous publication.

We consider this comment as a general remark.

24. There is major doubt on the validity of comparing clinical foot examination with pedobarographic measurements: Physical examination is most useful in detecting structural abnormalities and bony prominences associated with focal high pressure areas. The dynamic plantar pressure measurement measured by the pedobarograph detects contact force/area. As they detect different things, the comparison may not be a valid comparison. The absolute plantar pressure could be elevated to beyond 700 kPa without foot structural abnormality and bony prominence.

The research question was: can therapist identify regions with elevated pressure, as categorized through quantitative plantar pressure measurements, through clinical examination? We are not primarily interested in which and how information from clinical examination therapists use for their judgement, but are they able to distinguish low from high plantar pressures. We hope that the relevance of this question came clear from the arguments and the literature cited in the paper and commentary in this letter. We can support your statement with our experiences in numerous early stage patients with diabetes which have virtually healthy feet, but with plantar peak pressures higher than 700 kPa. It is therefore, that it must be clarified that either clinical examination is adequate to identify regions with elevated plantar pressures or that one better rely on quantitative plantar pressure measurements. Nota bene: almost all therapists used foot print devices complementary to their physical examination.

25. The professionals 'were asked to identify locations with excessively high plantar pressure': Were the professionals told that they need to detect locations with pressure increase over 700 KPa? Or that the criteria was areas at risk of ulcerations? If not, the threshold for detecting 'locations with excessively high plantar pressure' by the professionals may not correspond to that set by the investigators for the pressure platform. The subsequent comparison would not be a fair comparison, and the results and conclusion drawn would be misleading.

We also recognize that there is subjective aspect in the interpretation of plantar pressure. We asked the therapists to identify locations with excessively high plantar pressure, instead of just elevated pressures. Statements in the literature indicate 700 kPa as criterion for various foot pathologies, which seem to us as an reasonable standard against clinical judgement can be measured. In paragraph five of the discussion we considered lower cut-off points, which did not result in a more favourable outcome for the therapists. Regarding the condition of the patients who participated in this study, we consider cut-off points lower than 500 kPa inappropriate. However, if lower pressures than 500 kPa must be considered as excessively high plantar pressure, than there is the problem of regions with higher pressures than 500 kPa which were not indicated as such. Moreover, even within one foot, regions with lower pressures were frequently rated as elevated and other regions with higher pressures were not.

26. Figure 1: Not necessary. Figure 2,3,4: Gives little information except to illustrate the highly variable results, which has been mentioned in text. It would be much more useful and informative to the readers to replace with a table/chart illustrate the ‘clinically high pressure area’ detected by the three groups vs the GS high pressure areas.

We kindly would refer to table 2, for information about clinically rated high pressure areas versus the Gold Standard values. Figure 2, 3 and 4 provide a comprehensive description of the dataset from which information about the proportions for each ‘low’ and ‘high’ pressure can be derived. Their also essential to illustrate the effect of shifting the cut-off point. We would like to leave the decision whether figure 1 is redundant, to the editor. We would prefer to keep it.

27. Metatarsalgia is actually very non-specific diagnosis. It would be better if a specific diagnosis can be stated, especially ones) that can account for the pressure increase. Otherwise the three patients could be having very heterogeneous conditions.
We agree that metatarsalgia is a non-specific diagnosis and for that reason we provided supplementary information in table 1. In theory, heterogeneous conditions, and possibly diverse plantar pressure patterns, are in advantage of therapists: ICC’s tend to decrease in studies with a homogenous population and increases with a heterogeneous population. It is easier to differentiate a certain characteristic in a heterogeneous population.

28. The median professional experience differs to a large extent among the three groups, which may affect the results obtained. Please include the range of experience in methodology.

We adopted the suggestion of the reviewer.

29. Specific training on detection of high plantar pressure areas by the professionals would be relevant information to be include.

We agree with the reviewer and supplemented the following text: ‘….after validation and standardisation, the best method should be trained and implemented as a standard guideline.’

30. Discussion line 2: ‘the proportions ? Based on’. Something missing?

We removed the question mark.

31. Table 3: There is a discrepancy between the table legend and the contents. It is uncertain whether the rations stated are with respect to the GS or not.

We supplemented the following text: ‘The proportion of elevated plantar pressure per region according to the GS is determined by the number of regions with observed elevated pressure to all regions for that location: for example, three of the six BT regions (three patients both feet) had elevated pressure i.e. 3:6 = 0.50. For each region, the ratings per discipline were averaged over both feet of the patients.’

32. In Results: ‘sometimes one can see an underestimation.....and most of the time an overestimation....’. The actual false positive/ false negative rates, and positive/ negative predictive values would be useful.

Positive and negative predictive values are indeed useful measures. For their calculation a two by two table is needed. Regardless the correctness of positive and negative predictive values for this design, calculation will result in 180 2x2 tables (6 plantar regions and 30 therapists). In addition, we are facing a nested design, where region (BT, MT’s) is nested in foot, foot (left/right) within patient and therapist nested within discipline (3 groups). Therefore, in the context of this study, the proportion of elevated plantar pressure and the ICC’s are the most appropriate summary measures.

33. It is interesting to note that the peak bear foot plantar pressure in this study as stated in table 2 differs dramatically from baseline in shoe peak pressure presented in ‘Comparison of foot orthosis made by podiatrists, pedorthists and orthotists regarding plantar pressure reduction in Netherland’ table 2. It is understood that barefoot plantar pressure measured would be different from those taken in shoes. However, the pressure distribution pattern is not expected to change excessively, unless there is a insole in situ. The authors may need to think about what accounts for the dramatic change. Though this is not strictly a review question on this paper, it may bear significant implications on the validity of the study, and is worth some detailed examination.

To our perception, the difference between peak pressure measured under the bare foot and in a shoe presented in our papers, is not different to what is found in the literature. As a frame of reference for bare foot plantar pressure in various populations, we would like to refer to the literature stated under point 6. If you compare various inshoe pressure studies with aforementioned references, for example in the paper of ‘Tsung, B. Y., et al. (2004). “Effectiveness of insoles on plantar pressure redistribution.” J Rehabil Res Dev 41(6): 767-74.’ where the reviewer was co-author, than one would conclude that all ‘only shoe’ mean peak pressures values are below 400 kPa. Although the ‘shoe only’ peak pressures in our study tend to be higher, the values are similar. Since our data is not very different from other studies, we do not understand the issue addressed and the classification ‘dramatically’.
Comments Karl Landorf

General:
34. Having read the new manuscript and the previous publication I make the following comments: Yes, it is essentially from the same study as before: same practitioners/assessors and patients (and authors). However, the investigators have endeavoured to answer a different question to the previous study that was covered in the first (previously published) paper. The question they have attempted to answer in this manuscript would not have sat well in the first paper. Therefore, on the surface it would appear that it warrants a new paper. Interestingly, the study in question (i.e. the new manuscript) only included 10 orthotists (rather than the 11 from the first study), even though they had the same number (10) of podiatrists and pedorthists in both studies – I am not certain as to why one orthotist was left out of this secondary study.

Because the procedure of analysis (GENOVA) requires balanced data, we left out one orthotist through random sampling.

35. Yes, the methods are similar; although the investigators used an EMED platform-based plantar pressure system as the primary measurement system rather than a Pedar in-shoe plantar pressure system as they did in the first study. Therefore, to some extent it is new data that answered a different research question; however it was housed in the same overall project that the previous paper (i.e. the first study) was generated from. I guess the question to be answered then is: is this necessarily problematic?

We consider this comment as a general remark.

36. There is one problem that I am concerned about with all of this data (i.e. data from the one overall study) and that is the investigators have conducted a large number of hypothesis tests with what appears to be no pre-specified plan of analysis. Therefore, there is the very real likelihood of a Type 1 error. I don’t have the skills to work out how many hypothesis tests the investigators would have performed but it would appear that there were many. Given that the data included in this manuscript is from the same overall project as the first study (using the same assessors and participants), this may be a problem. This is a tricky issue; where do you draw the line on number of hypothesis tests? As this is new data, is it OK to do this? Or, because the data has been extracted from the same practitioners/assessors and participants, does this increase the likelihood of Type 1 error. At the very least, I think the authors need to address this issue in their limitations section (this probably should have been addressed in the first paper too). An adjustment of the level of significance may also be required.

We can assure the reviewer that in this experimental design all hypotheses were ‘a priori’ formulated and are therefore not data driven. Also, we were not interested in testing hypotheses on the (fixed part of the) experimental design, but in the variability of the random part of the design, i.e. in only those variance components pertaining to method agreement (gold standard vs. clinical rating). Results are estimated variance components and ratios of summated estimated variance components. Moreover, data presented in both papers were independently collected in different datasets.

The multiple testing issue and the need for adaptation of the p-value is controversial. We would like to refer to chapter 12 in ‘Modern epidemiology, Rothman & Greenland, Lippincott Williams & Wilkins; 2nd edition (1998)’ for a relevant discussion of this topic. We think that adding remarks about this topic do the discussion paragraph would not make the article more informative and certainly less readable.

37. One minor issue, although, not much can be done about it, the references in the initial paper are out of sequence. The easiest reference to detect this is reference number 29 in the text is actually reference number 30 in the reference list. I tried to work out where the sequence went astray and I think it is from about reference 22 or slightly prior to that one. I assume the authors added an extra reference at some stage, which has messed up the sequence.

We are very sorry about this shortcoming in the published paper. If possible, we will be more than happy to be of assistance in putting this right.