Author's response to reviews

Title: Strength of agreement between diagnoses reached by clinical examination and available reference standards: A prospective validity study of 216 patients with lumbopelvic pain and/or symptoms referred into the lower extremity

Authors:

Mark Laslett (mark.laslett@xtra.co.nz)
Barry McDonald (B.McDonald@massey.ac.nz)
Hans Tropp (Hans.Tropp@lio.se)
Charles N Aprill (caprill@magnoliaclinic.com)
Birgitta Oberg (birgitta.oberg@ihs.liu.se)

Version: 2 Date: 10 March 2005

Author's response to reviews: see over
Response to reviewer’s comments

Reviewer: Jonathan Hill

The STARD[1,2] recommendations are well known to the authors and are used in other papers generated from this research. The results of this analysis of data cannot be presented in the manner recommended for the following reasons:

1. The recommended manner of presentation requires a single index test be compared to a single reference (gold) standard. This paper presents diagnoses derived from multiple tests compared to multiple reference standards.

2. The current paper essentially evaluated agreement between clinical examiners and expert diagnoses based on reference standard methods. Kappa is appropriate when sufficient numbers appear in each cell, but is inappropriate here because most diagnoses have low prevalence in the sample. PCC was used to estimate chance agreements and actual agreement was compared to this. We could find no other method of making this comparison that was relatively simple and understandable by readers. We would certainly be open to suggestions or alternatives.

3. Sensitivity, specificity and the summary likelihood ratio statistics are appropriate when evaluating single tests against single reference standards, but as pointed out above, this is not possible in the comparison applied in this paper.

Diagnostic accuracy data is available for many of the tests used in the clinical examination. Some results are already published[3-11] and others (from the current data) are in separate papers currently submitted for publication or pending submission. The data in the papers based on the current material present diagnostic accuracy data in the STARD recommended format with 2X2 tables, flow diagrams etc.

It is agreed that more information on the index tests for SIJ and ZJ pain is desirable. However, to do so would increase the length of the paper dramatically, and (we believe) to an unacceptable degree. The references used do have full descriptions of the index tests and we feel this is adequate for the current paper.

The STARD recommended flow diagram has not been provided for reasons stated above, but a flow diagram of patient recruitment has been added to give the reader a better idea of the project process. A sentence has been added to the Methods section referring to this new diagram.

A sentence in paragraph 1 of Methods has been altered to indicate duration between index and reference standard tests. “These clinical examinations required between 30 and 60 minutes and were carried out immediately before the reference standard diagnostic tests.”

Appendix 1 is added (and the other 2 Appendices re-numbered) to provide data on sensitivity and specificity of specific index tests used. This data is derived from the literature available at the time the study commenced. Reference to later publications is inappropriate since the study design was set in stone in early 2001. 95% CIs of these tests have not been included in the new Appendix since most of the papers did not provide them. It is possible we could generate them from the data available from a few papers if the reviewer believed it essential. However, it is our belief that such detail is unwarranted in this paper which does not calculate these statistics on the current data.
Reviewer: Elaine Thomas

1. It is agreed that reference in the text to Table 4 is confusing. The 16 categories resulted from applying the methods detailed in Appendix 2. The text has been moved to its correct place prior to the section that refers to the even simpler list of 6 patho-anatomic categories.

2. Reordering of the text as stated above should resolve the reviewer’s concern 1(b)

3. PCC was not calculated for the physiotherapy diagnoses because it is inappropriate. PCC is the chance that a physiotherapist would guess the reference standard diagnosis, hence it is the PCC of the reference standard that matters. Calculation of PCC for the physiotherapy diagnoses would be calculated to estimate the chance that the reference standard would agree with the physiotherapy diagnosis. While this is an interesting perspective, it is not the one we set out to evaluate.

4. Some descriptive variables have normalish distributions and others are highly skewed. To keep the table as simple as possible yet provide the appropriate statistics, an additional column for standard deviation has been added for the continuous variables. Results for the dichotomous variables (sex, time off work etc) remain unchanged since SE and STD have little meaning for them.

5. Point 1 (c) PCC of 0.319 and exact agreement of 0.532 refers to the 16 categories. The reordering of text as described in point 1. above should clarify this.

6. There were 9 cases with two patho-anatomic diagnoses. These cases were therefore represented in two cells apiece. Therefore 112 cases had single patho-anatomic diagnoses. A sentence has been added to provide this data.

7. The reviewer correctly identified confusing presentation of data regarding the simplified list of patho-anatomic diagnoses. Multiple representations of single cases in Tables 5 and 6 are hard to describe easily. To make the presentation of the data more readable, only cases that do not cause multiple entries are included. The number of cases is now 103. The text and tables have been altered to reflect these changes. We hope this is more intelligible. Cases where either or both physiotherapist and reference standard clinician failed to make a patho-anatomic diagnosis are excluded from these tables, which accounts for the reviewers concern 2(c).

8. Minor essential revisions: Yes, the 13 patients seen by SBY (second therapist) are included. It was hoped that she would see many more patients (thus increasing the sample size significantly), but this did not eventuate. These patients are included because the original design of the study did set out to make a comparison between the two therapists and to improve the generalizability of the results. It was a disappointment, but we felt it necessary to include all patients to conserve the consecutive nature of the sample.

9. Notes at the end of the Appendices have been modified.

10. Standard error (SE) for proportions were calculated: \( \sqrt{\frac{(P *(1-P))/N}{}} \). The 95% CIs for PCC were calculated using the formula \( P \pm (1.96*SE)[12] \)

11. Table 2 “Note 2” removed and replaced by ‘*’ as suggested

12. Use of ‘Z’ for ZJ pain has been altered in Appendix 1

13. All cases of ‘exact’ agreement are included within ‘clinical’ agreement. A sentence has been added to the data analysis section of methods to make this clear.

14. PCC and percent agreement has been presented as percentages as suggested

15. All spellings of ‘disk’ converted to ‘disc’

References


