Author's response to reviews

Title: Strength of agreement between diagnoses reached by clinical examination and available reference standards: A prospective 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: 3 Date: 5 April 2005

Author's response to reviews: see over
Response to Reviewers April 5 2005

General Remarks

1. I thank the reviewers for their very helpful comments, criticisms and suggestions. This material has been difficult to present in a succinct and informative manner and the reviews have helped enormously. My only concern is the suggestion (point 3) by Jonathon Hill that we are being less than open in presenting the data. I am sure that this was not his intention however.

2. I have completely revised the presentation of data, making it much simpler with fewer tables along the lines suggested. This also makes one appendix unnecessary. The agreement results are similar however.

3. There are now just three tables. Table 1 remains as the demographic and profile descriptions of patients. Table 2 is now a complete cross-tabulation of diagnoses for 216 patients by the reference standard / expert opinion clinician (27 categories) and physiotherapist clinician (19 categories). This is essentially the raw data. On the basis of this table, condensations to the diagnostic categories were carried out. Because both clinicians provided multiple diagnoses for many patients, there 368 diagnostic items i.e. a single patient with a multiple diagnosis will provide a diagnosis count in more than one category. Table 3 provides this information and it can be checked against Table 2. Table 3 also has dashed lines transecting the table, separating the six main patho-anatomic categories from the non-patho-anatomic categories (Illness behaviour, Indeterminate, Other and Instability). This data within this truncated patho-anatomic section is the basis for the analysis of agreement on the patho-anatomic diagnostic categories.

4. On the surface, agreement does appear to be weak even though better than chance. However, this does not concern us greatly. This is a pragmatic report of the overall performance of a low tech clinical examination and diagnoses achieved, compared to highly sophisticated and predominantly invasive procedures with the most complex group of patients I have ever encountered in 30 years of clinical practice. The high proportion of cases deemed to be diagnostically indeterminate or displaying confounding illness behaviours (in the opinion of the clinicians) attests to the complexity. The achievement of about 50% agreement on patho-anatomic categories was hoped for prior to commencement of these studies. As a consequence of this work, certain significant modifications to the clinical examination will be made and subsequent studies may demonstrate an improvement on what we achieved. ‘Exact agreement’ is a very demanding requirement when multiple pathologies are present. ‘Clinical agreement’ as described in the paper is really all that can be expected of a low tech clinical examination of patients with low back pain – a symptom commonly described as ‘non-specific’ in the low back pain literature.

Reviewer: Jonathon Hill

Point 1. It is correct that overall agreement contains ‘pockets’ of agreement associated with specific tests. However, to describe the tests used in any useful way is a monograph on its own. In the Methods section, the basic methods used to reach patho-anatomic diagnoses by clinical means is outlined with appropriate references. One paper comparing the centralization phenomenon with discography results is accepted for publication elsewhere. A second paper on other clinical variables with discography is under review elsewhere. One paper comparing a specific clinical method with facet joint blocks has been published in this journal and another comparing other specific clinical tests with facet joint
blocks is also under review. One paper on clinical diagnosis of sacroiliac joint blocks has been published and a second paper is accepted for publication. These papers are referred to in this current paper and we have relied on this to simplify the presentation of general agreement data.

With regard to expanding the description of agreement for the whole sample including those with multiple diagnoses: Table 2 provides the raw data and the reader is invited to evaluate any comparison of interest. I hope this is sufficient.

**Point 2.** Tables 2 and 3 of the previous document have been removed as outline in General remarks / point 3 above. The reviewer suggested a 13X13 table. Again Table 2 in the current revision provides the full data. The reference standard clinician provided more diagnostic categories than the physiotherapy clinician, which makes a symmetric table impossible. Part of the problems with the first two manuscripts was an attempt to condense the information into symmetric tables (for possible kappa statistic analysis). I believe that this was a mistake and the current presentation is a more accurate presentation.

**Point 3.** The current presentation of the raw data should reassure the reviewer.

**Point 4.** The table structure has been simplified as described in General Remarks (point 3 above).

**Reviewer: Elaine Thomas.**

**Point 1.** It is agreed that the non-patho-anatomic diagnoses were problematical. The new structure of the data presentation (Table 2) provides the raw data. Table 3 allows isolation of the patho-anatomic data in a way that can be checked against the raw data. The presentation of prevalence of all categories is still accessible through these tables. It is agreed that the agreement analysis that includes categories without satisfactory reference standard is a problem. But it is interesting! The data from Table 2 does invite an agreement analysis which is provided for by assessing chance agreements through PCC and achieved agreement ('exact' and 'clinical'). Perhaps one might consider this a pragmatic 'real world' comparison of methods, which is appropriate. The subset of patho-anatomic categories accessible from Table 3 is the source of the agreement analysis and kappa statistic calculations.

**Point 2.** The revised table structure should resolve the issues raised.

**Point 3.** Again, the new table structure permits examination of the raw data and the subsets derived.

**Point 4.** It is agreed that PCC is seldom used in this context. Personally I was unaware of it myself until Dr McDonald suggested its use as a ‘rule-of-thumb’ method of estimating chance classifications. Kappa is the most appropriate, but the large number of categories in the raw data, asymmetric nature of the table (Table 2 now) and the fact that most cells had zero or low numbers makes this impossible on the raw data. This is why we have used the kappa only on the small subset of patho-anatomic categories. If the reviewer can suggest any other method beside PCC that can be used to evaluate chance agreement on the raw data (Table 2) or even the full table (Table 3) we would be most appreciative. I have sought advice in numerous places, but PCC was the only simple method that emerged. The reviewer’s comment about the “agreement over and above chance” is helpful. I have modified the Discussion to include this concept. Essentially agreement over and above chance is between 19 and 24%.
Point 5. The specific details of the range of agreement are now slightly different and the current presentation simpler, but I understand the point. If we were to summarize the range of agreement using confidence limits, that should also include an upper limit also, don’t you think? That makes the figure so wide (26-64%) so as to be almost meaningless. I have stuck with the estimates and pointed out that the upper estimate of chance agreements and the lower estimate of achieved agreement do not overlap and the data from which these calculations are available.

Point 6. Changed table structure addresses these issues.

Point 7. Achieved agreement for the six patho-anatomic is higher, but the agreement over and above chance is only 5% more than for the raw data (Table 2) and the whole dataset of condensed categories (total of Table 3). This is discussed briefly in the Discussion.

Minor essential revisions
1. Table 1. My apologies. The table has been slightly restructured, giving the IQR and removing the Mean and SD for non-normal data variables. In addition to duration of symptoms and duration of time off work, the Roland questionnaire data was heavily skewed. The other variables are reasonably normal so Mean, SD are retained for these. STD has been replaced by SD. The SE of the Mean has been removed.
2. The flow chart has been amended. The 294 was a typo and the misspelling of “examining” corrected. The figures should now match those in the Tables.
3. It is agreed that Kappa adequately serves to estimate agreement adjusted for chance agreement. However PCC is more intuitive and clinically meaningful, so it is retained.
4. The details are now somewhat different, but (I hope!) consistent now.

Concluding remark
I have added a paragraph to the end of the Discussion regarding the weak agreement referred to by the reviewers which is essentially the same as the point 4 of my General remarks.