Reviewer's report

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

Version: 2 Date: 24 March 2005

Reviewer: Elaine Thomas

Reviewer's report:

General
I thank the authors for the information added to this revised version of the manuscript which has aided in my understanding of their study. In this revised version there are several important issues that need to be addressed covered below:

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

1) As the main objective of this study was to look at the reliability of pathoanatomical classification, I feel that the data referring to non-pathoanatomical diagnoses (given no reference standards was applied for these diagnoses) are confusing the agreement analysis. I understand that it is important to report the “prevalence” of such data but I feel that it cannot be involved in the agreement analysis. The emphasis for the classification should be on the pathoanatomical diagnoses.

2) The data given in Tables 2 & 3 are confusing. Only the totals for the first row and first column have a meaning, ie. total number of disc diagnoses and number of disc only diagnoses, the others do not. Given that there are few participants whose diagnoses change dramatically after applying the simplification rules given in Appendix 3, and the reduction of diagnoses if removing the non-pathoanatomical participants as suggested in Point 1, I feel that these individual tables should be removed as the main concern of this study is the comparison of two methods of diagnosis.

3) One of my main concerns is the presentation of the multiple-diagnosis data. As data for the current 16 class are shown separately (Table 4) the reader cannot see where the agreement/disagreement lies. If, as suggested in Point 1 above, the authors were to reduce the number of diagnostic categories by removing those without a pathoanatomical diagnoses, this reduced classification could be presented in a standard cross-tabulation allowing the readers to get a better idea of the good and poor areas.

4) The PCC is not frequently used in the field of agreement and I feel that the authors need to explain it’s role better. I now understand that the PCC is the proportion of correctly classified people expected by chance, due to the marginal totals according to the reference standards. Hence it is in fact a point of reference against which to compare the level of observed agreement and has no real meaning as a stand-alone statistic. For example, the exact agreement calculated for the comparison of the data show in Table 4 is given as 31.9% and the PCC for the same data is 18.3%. Does this mean that the “exact agreement” above and beyond chance is 31.9%-18.3%?

5) In the Discussion the authors summarise the range of agreement as 32%-68%. I assume that the maximum given (68%) is the upper confidence limit of the kappa figure for the data given in Table 6. Therefore, should the minimum figure for this range not be the lower confidence limit associated with a chance-corrected statistic for “exact agreement”?

6) The data in Table 5 is a replicant of that given in Table 6 and as Table 6 uses it more informatively my suggestion is to remove Table 5.

7) The data given in Table 6 is presented for only those subjects with a single pathoanatomical diagnoses. The levels of agreement in this, perhaps, simpler group of participants is substantially higher than that seen for the whole group. The authors should discuss this as I feel their main
The statement of the usefulness of this classification criteria (page 13, paragraph 1, final sentence) rests solely on the data presented for this reduced, simpler group. The levels of agreement for the more complex situation where participants are allowed to have more than one diagnosis is poor, though greater than chance, for both “exact agreement” and “clinical agreement”.

---

**Minor Essential Revisions**

(such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

1) In the first review I made some suggestions for the appropriate presentation of the data given in Table 1. These recommendations were not carried out. I have therefore below re-iterated these recommendations.

   a. There is a need to give appropriate summary data dependent on the distribution of the observed data - mean & SD for Normal, median & IQR for non-Normal – not give both for all. If presented well, this will not make the table confusing. It is clear that the variables recording duration and weeks from work are not Normal and hence the mean & SD are not appropriate

   b. STD not standard short-form for standard deviation – should be SD and a reference to this abbreviation should be given as a foot note to the table as for VAS etc

   c. The standard error of the mean should be removed from the table as this is not appropriate when presenting descriptive, and not inferential, statistics.

2) The flowchart is helpful to the reader but there are some small errors in the text of the flowchart and an inconsistency when comparing data in the chart to that in the text of the manuscript:

   a. The chart reports that 296 participants were invited to take part whereas in the text this figure is given as 294.

   b. I think error has effected the figure given for the number of patients included as 296-78 exclusions should be 218 and not 216 given, but 216 is the correct total.

   c. There is a typo in the middle diamond (Exmaining..)

3) Given the authors present the kappa statistic for the data in Table 6, I feel that the PCC is not needed also as the kappa is adjusted for chance agreement.

4) On Page 12 you give the proportion of agreement in Table 6 as 70.9 but calculating it from the data in Table 6 leads to 71.8.

---

**Discretionary Revisions**

(which the author can choose to ignore)

---

**What next?:** Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Acceptable

**Statistical review:** No

**Declaration of competing interests:**

I declare that I have no competing interests