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: Jonathan Hill

Reviewer's report:

General
I am content with the response provided by the authors that the STARD recommendations do not need to be religiously adhered to for this study as clinical judgement regarding diagnoses are being compared rather than single tests being applied in isolation. However, much more still needs to be done to make this paper clearer for readers.

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

1. The authors response to my comments give some clarity on their objectives, but I would welcome further clarification. Their aim may be to compare agreement between a physiotherapist’s overall assessment of specific diagnostic classifications, when using a range of clinical tests against a diagnostic reference standard. However, the emphasis of the paper currently emphasises agreement for individual tests within a selected sample of patients as is highlighted by the data presented in table 6.

Their findings appear to show that:
- overall chance agreement on diagnostic classification was calculated to be around 18% in this population
- this was improved to 31% agreement with the physiotherapist’s clinical examination for “exact agreement” and 53% for “clinical agreement”
- when individuals with a multiple diagnosis or indeterminate diagnosis were excluded then agreement improved to 71%, with chance agreement calculated as 37% for this population.
- of these individuals that were selected out, having only a single diagnosis, this study found that discogenic pain and hip pain were most accurately diagnosed.

If this paper aims to present overall agreement for a range of tests including individuals with multiple diagnoses as argued in their response to my first review - then the information on the agreement for the whole sample studied including individuals with multiple diagnoses must be expanded. Most clinicians do not have the luxury of excluding difficult cases. These individuals with more than one diagnosis are clearly harder to examine, but the data provided does not enlighten the reader as to why this is the case. It is likely that the reduction in agreement from 71% to 53% in the whole of their sample is due to certain pockets of agreement around specific tests but unfortunately this information is lacking.

2. I suggest a solution would be to remove tables 2 and 3 as they are unhelpful (because they are confusing and the ‘total’ columns don’t mean anything) and this information could be better presented by replacing them with a 13x13 table similar to table 6 (use the diagnostic categories from table 4 with ‘indeterminate’, ‘illness behaviour’ and ‘other pathologies’ removed [n=147]). This would provide a similar level of information to table 6, for individuals with multiple diagnoses and would
inform readers how the different tests contributed to provide 53% overall agreement. In addition, when information is provided on which tests were more accurate (e.g. hip and discogenic pain) this needs to be quantified.

3. Once an honest presentation of the physiotherapist’s agreement has been presented for both the whole sample and selected sample then the discussion can be re-written to be more open and realistic about the study limitations and the usefulness of the clinical assessment will be better understood by readers.

4. Table 5 is not useful and can be removed.

5. Appendix 1 and 2 provide useful information and the flow diagram that has been added, significantly improves the paper.

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

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'