Reviewer's report

Title: Protocol for Determining the Diagnostic Validity of Physical Examination Maneuvers for Shoulder Pathology

Version: 2 Date: 27 June 2012

Reviewer: Robert Borden Hopkins

Reviewer's report:

Major Compulsory Revisions
The guideline that was submitted for the study was a CONSORT guideline which is for randomized control trials. The appropriate guideline statement for studies of diagnostic accuracy is the STARD statement.

Minor Essential Revisions
The study is well thought out and will provide important answers to the clinical questions. There are enough details to follow the clinical study design, but there are some deficiencies in other preliminary areas. For example, the description of the systematic review is weak. The description of the systematic review to identify tests is worthy in length of a separate manuscript, if the authors have not already thought of this. In particular, the details that are missing can be viewed in the Amstar checklist: which databases (Pubmed, Medline, EMBASE, etc.) were search, with what keywords or MeSH headings, over what time period, by whom, etc. Similarly the modified Delphi panel is worthy in length of a separate publication, if the authors have not already thought of this. A guideline for reporting Delphi panel results has been suggested by Boulkedid (Boulkedid R, Abdoul H, Loustau M, Sibony O, Alberti C (2011) Using and Reporting the Delphi Method for Selecting Healthcare Quality Indicators: A Systematic Review. PLoS ONE 6(6): e20476. doi:10.1371/journal.pone.0020476). For both the systematic review and modified Delphi panel, the outputs of these activities would be viewed as limitations unless they have been subject to review by their academic peers. This could be added to the discussion section.

I had a bit of trouble in my first read through following the manuscript, or it could just be me. The inclusions of a flow chart in the study design would be most helpful, although the abstract is easier to follow than the text.

The Discussion tends to repeat material in the introduction and is not necessary (Sackett classification, etc.).

The section title RESULTS should be avoided, this implies that the study is done and there is no need to submit the study protocol. The paragraph that describes the results of the systematic review could be moved to page 9 before the title “Clinical Examination Testing”. The second paragraph should be deleted.

There are a few minor statistical issues. The sample size calculation is based on a margin of error of 0.10. A citation of where this value comes from is needed.
i.e., similar to other studies (cite one). The method for sample is okay, although there are no adjustments for subgroups, etc. This may be required in the discussion section as a limitation.

I suppose the biggest confusion in the study design is the two-step process, where only some patients that are enrolled move onto physical examination. The sensitivity/specificity is calculated for those patients who have had no physical examination and those with physical examination. With this approach, an addition to the discussion that some patients who should have moved onto physical examination but did not, is important. (Without this, the complete 2x2 table cannot be estimated). This will be highlighted in the MRA results of the non-surgical candidates. Similar to this, the diagnostic accuracy of MRA should be provided, in addition to being cited.

The choice of measures to report is sufficient (sensitivity, specificity, Likelihood ratios, predictive values, pre-test and post-test probabilities). Fancier methods such as area under the curve for a receiver operator curve could be done, but often this adds little interpretive or clinical value.

Discretionary revisions
The overall format of the protocol would read better if it sequentially followed the STARD statement.

After you have identified the diagnostic accuracy of the components of the history or physical examination, there are alternatives to express the results to aid clinical guidance. You have suggested the best 1, best 2, best 3, etc. This needs to be broken down as the best 1, best 2, best 3 for sensitivity (to rule in disease) and then for specificity (to rule out disease). Some tests will be highly sensitive and others highly specific. From this, you could also list the best combination that has highest area under the curve (a joint measure of both sensitivity and specificity), which now may have value. Alternatively if you use SAS software in a regression approach, you can estimate the best model which looks at all combination of variables and picks the best combination of tests to make inference.

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

None.