Reviewer's report

Title: Development and evaluation of an instrument for the critical appraisal of randomized controlled trials of natural products.

Version: 1 Date: 29 January 2009

Reviewer: Tanya Horsley

Reviewer's report:

Overall, the authors have completed a well done, interesting report outlining the development and evaluation of a critical appraisal instrument for evaluating RCTs of conventional medicines for use with single entity natural products. It is well known that critical appraisal tools offer a means for estimating potential biases and factors that may affect study findings; with CAM trials a burgeoning area of research the development of a tool specific to CAM is not only necessary, but timely. The authors use a traditional Delphi approach using 11 of a potential 14 experts for establishing and finalizing a list of items for use on the final evaluation tool. The paper is generally well written, although lengthy, its strength lies in the level of methodological detail for how the tool was developed and evaluated.

As a general comment, critical appraisal tools have often been criticized as being cumbersome to use when lengthy and thus are often not used by authors within many reports (particularly systematic reviews that include many studies). This tool is rather lengthy (28+ items) and may impede uptake of the tool. A suggestion would be to include this as a limitation of the tool to its potential use for authors.

The authors have indicated that a tool has been developed for responding to questions within the tool – this should be included within the publication as well if available. I would urge the authors to also include information (if not already included) on interpreting findings and how to incorporate the information into reports. This is a common limitation for previous critical appraisal tools.

- Major Compulsory Revisions

1. The strength of the Delphi process lies within the ability to choose the appropriate ‘experts’ to be involved within the process. It would be prudent for the authors to include a brief outline or descriptor for how they decided on experts and what the criteria was for determining candidates.

2. The search for critical appraisal instruments for conventional medicines was taken from a previously published systematic review; however, the review search date was from 2006 (uncertain of specific month) which jeopardizes the currency of the information. It is suggested that authors update the search and do a quick scan to ensure the information is up to date (even if only in the major databases)

3. I am deeply concerned with the methodology of using ‘frequency’ of items to determine relevance for inclusion. Did the authors not calculate kappa’s for
agreement between experts? They report in a way that is somewhat unclear -- for example ‘there was general agreement’, or ‘no significant effect’ but do not provide the data (p values) or methods for assessing significance. These need to be reported clearly.

- Minor Essential Revisions

4. I’m uncertain as to what a ‘revised CONSORT’ statement is. If the authors revised it they should provide a description of what was changed.

5. Comments and discussion between anonymous expert participants is a strength of the Delphi process, as such, the author may want to report how many of the participants in fact provided comments and descriptions of why they included certain items along with responses to the questions. This would help strengthen the transparency of the process.

6. Information was emailed to participants; the authors should include a statement as to how they maintained anonymity of participants in the process (e.g. BCC participants?). As it stands now, I’m uncertain as to how this was achieved.

7. The authors include instruments that are ‘modified’ from their original versions within the inclusion of the report. Although the authors are unlikely to change their approach (and have indicated this as a limitation); they should have been excluded entirely. They are not validated in their new form and skew the data on frequency (which they relied on for determining criteria).

8. Although they randomized participants to groups, the authors should provide a description of the experience/expertise in critical appraisal (match comparison?) of participants for the field test of the instrument. Were there differences between groups at baseline with such a small sample? Did they include a broad range of expertise? It would be interesting to see the findings if a range of expertise (novice vs. expert) existed. There is a brief comment on this in the discussion section; but should appear clearly and concisely within the methodology section.

9. The criteria used for determine that an item was ‘essential’ or not is concerning. These criteria are neither validated and nor ‘standard’ as part of the Delphi process. Are these arbitrary criteria created ad hoc by the authors for including items? If so, the authors should state this and provide a discussion for its limitations. The question that follows is – why go through this entire process of using experts etc. to resort to using previous validated tools? Why not simply update a validated tool and just add the section for single NP products?

- Discretionary Revisions

10. A general comment – this report could be shortened by editing extraneous information within the text. Although stylistic, it would be less arduous to read through.

**Level of interest:** An article of importance in its field
Quality of written English: Acceptable

Statistical review: Yes, but I do not feel adequately qualified to assess the statistics.

Declaration of competing interests:
I declare that I have no competing interests