

Research article

Open Access

## Development and evaluation of an instrument for the critical appraisal of randomized controlled trials of natural products

Tannis Jurgens\*<sup>†1</sup>, Anne Marie Whelan<sup>†1,2</sup>, Melissa MacDonald<sup>1</sup> and Lindsay Lord<sup>1</sup>

Address: <sup>1</sup>College of Pharmacy, Dalhousie University, 5968 College Street, Halifax, Nova Scotia, B3H 3J5, Canada and <sup>2</sup>Pharmacy Consultant, Department of Family Medicine, Dalhousie University, 5968 College Street, Halifax, Nova Scotia, B3H 3J5, Canada

Email: Tannis Jurgens\* - tannis.jurgens@dal.ca; Anne Marie Whelan - anne.marie.whelan@dal.ca; Melissa MacDonald - mmmacdon@dal.ca; Lindsay Lord - lordl@dal.ca

\* Corresponding author †Equal contributors

Published: 23 April 2009

Received: 11 December 2008

*BMC Complementary and Alternative Medicine* 2009, **9**:11 doi:10.1186/1472-6882-9-11

Accepted: 23 April 2009

This article is available from: <http://www.biomedcentral.com/1472-6882/9/11>

© 2009 Jurgens et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** The efficacy of natural products (NPs) is being evaluated using randomized controlled trials (RCTs) with increasing frequency, yet a search of the literature did not identify a widely accepted critical appraisal instrument developed specifically for use with NPs. The purpose of this project was to develop and evaluate a critical appraisal instrument that is sufficiently rigorous to be used in evaluating RCTs of conventional medicines, and also has a section specific for use with single entity NPs, including herbs and natural sourced chemicals.

**Methods:** Three phases of the project included: 1) using experts and a Delphi process to reach consensus on a list of items essential in describing the identity of an NP; 2) compiling a list of non-NP items important for evaluating the quality of an RCT using systematic review methodology to identify published instruments and then compiling item categories that were part of a validated instrument and/or had empirical evidence to support their inclusion and 3) conducting a field test to compare the new instrument to a published instrument for usefulness in evaluating the quality of 3 RCTs of a NP and in applying results to practice.

**Results:** Two Delphi rounds resulted in a list of 15 items essential in describing NPs. Seventeen item categories fitting inclusion criteria were identified from published instruments for conventional medicines. The new assessment instrument was assembled based on content of the two lists and the addition of a Reviewer's Conclusion section. The field test of the new instrument showed good criterion validity. Participants found it useful in translating evidence from RCTs to practice.

**Conclusion:** A new instrument for the critical appraisal of RCTs of NPs was developed and tested. The instrument is distinct from other available assessment instruments for RCTs of NPs in its systematic development and validation. The instrument is ready to be used by pharmacy students, health care practitioners and academics and will continue to be refined as required.

## Background

The use of Complementary and Alternative Medicine (CAM) has steadily increased in North America in the last 2 decades. Coinciding with the increase in use, the scientific evaluation of the efficacy and safety of CAM practices and products is also on the rise. In fact, a recent editorial reported that in 2008 more than 7500 CAM trials were indexed in MEDLINE [1].

The evaluation of the efficacy of CAM practices and products using the same methodology used for conventional practices and products is controversial. It can be argued that a randomized controlled trial (RCT), considered to be gold standard methodology in evaluating conventional medicine, may not adequately approximate how the CAM is used in practice and therefore may not be a fair way to assess efficacy. Of all the practices and products that make up CAM, natural products (NPs), including herbal medicines and natural sourced chemicals, most closely resemble conventional medications in terms of method of administration and the ability to quantify the administered dose. Like conventional medications, NPs are, therefore, well suited to evaluation using RCTs.

Health care professionals are accustomed to using evidence based information regarding the efficacy of conventional medicines as one part of their clinical decision making and therefore prefer to use evidence based information when helping patients make decisions about the use of CAM such as NPs. RCTs, conducted to determine efficacy of NPs, are being performed and published with increasing frequency. However, the critical appraisal of these trials, necessary to establish evidence based information, is more complex than with conventional medications due to the potential for variation in chemical content of NPs. RCTs completed using the best methodology do not reliably assess efficacy if the content of the NP being evaluated is not known or reported. The importance of the content of the NP and its potential for variation in content and thus activity is often not recognized or is underestimated. [1] This leads to difficulty in applying the results of RCTs of NPs to a patient's question about the utility of a particular NP.

The need to be able to assess the quality of RCTs of NPs prompted us to search the literature to identify an instrument to use in critically appraising RCTs of NPs. Our systematic review identified 16 published instruments for evaluating the quality of RCTs of NPs, however, none of the instruments stated that they had been validated and there did not appear to be a widely accepted gold standard instrument [2]. The CONSORT Statement for Herbal Interventions [3] came closest to fulfilling requirements for a quality instrument to critically appraise RCTs of NPs, however its content is specific for use with herbal medicines. There is therefore a need to develop an instrument

that will have a broader use in critically assessing the quality of RCTs of NPs that include herbs as well as single, natural sourced chemicals. Initially we proposed identifying a critical appraisal instrument that was designed for evaluating RCTs of conventional medicines and then adding a section specific for NPs. We subsequently determined that although there are many instruments available, there did not appear to be agreement as to which instrument was the "gold standard".

The purpose of this project was to develop and evaluate a critical appraisal instrument that is sufficiently rigorous to be used in evaluating RCTs of conventional medicines, and also has a section specific for use with single entity NPs, including herbs and natural sourced chemicals.

## Methods

The development and evaluation of a critical appraisal instrument for evaluating RCTs of NPs was divided into 3 distinct phases: establishment of a list of items essential in describing the identity of an NP, establishment of a list of non-NP items important for evaluating the quality of an RCT and, finally, a pilot study to compare the new instrument to a published instrument for usefulness in evaluating the quality of RCTs of NPs and in helping to answer a clinical question.

### **Phase 1. Establishment of a list of items essential in describing the identity of an NP**

#### *Selection of items*

The initial list of items for this study was compiled from items contained in published critical appraisal instruments designed for RCTs of NPs as well as from items suggested by the research team.

#### *Participants*

A list of 14 potential participants from disciplines including: natural products chemistry/pharmacognosy (n = 6), botany (n = 1), ethnobotany (n = 2), pharmacology of natural products (n = 1) and critical appraisal (n = 4) was compiled. Each participant was assigned a unique code (known only to the research assistant) and was sent an individual recruitment letter via e-mail. In anticipation of willingness to participate, potential participants were also sent the initial list of NP items along with instructions for completion

#### *Procedure*

The Delphi process [4] was used to achieve consensus among a group of experts as to which items describing the identity of an NP were essential to consider when critically appraising an RCT of an NP. The consensus building process was conducted in 2 rounds using email. All email contact with participants was conducted through individual emails (not one email containing all 14 addresses) to maintain anonymity. The research assistant was responsi-

ble for receiving and collating all results, using participants' codes. All collated results viewed by participants had names and codes removed to maintain anonymity. Consensus was considered to have been reached when 80% of participants were in agreement with an item being designated as essential to include in the instrument.

#### *Delphi Consensus-Round 1*

participants who agreed to participate were instructed to review the list of NP items on the Round 1 study form and indicate, using "yes, no, or uncertain", whether they felt that each item was essential to include in an assessment instrument used to evaluate RCTs of NPs. Participants were encouraged to submit comments regarding the wording of items and suggestions for additional items they felt should be included. Round 1 study forms were to be completed and returned by e-mail within two weeks of receipt. A reminder e-mail was sent to participants who did not respond to the initial email request for participation or had agreed to participate and had not returned the completed Round 1 form.

#### *Delphi Consensus-Round 2*

Round 2 study forms were created by compiling all responses, comments and suggestions for addition, with all identifiers removed. Follow-up letters and Round 2 study forms were sent to each participant via individual e-mail. Participants were instructed to comment on the revised list of items as well as on any comments or suggestions recorded from Round 1. A reminder e-mail was sent to each participant who did not respond within 14 days.

#### *Analysis*

Round 1: Responses, comments and suggestions from all completed Round 1 study forms were recorded and reviewed, using participants' codes to maintain confidentiality. A revised list of items was compiled based on participants' feedback and investigator expertise and was used for Round 2. Round 2: Responses and comments on all completed Round 2 forms were recorded and analyzed and a final list of items considered to be essential by the study participants and investigators was assembled

### **Phase 2. Establishment of list of important non-NP items**

#### *Identification of critical appraisal instruments for RCTs of conventional medications and NPs*

A systematic review of the literature was conducted to identify critical appraisal instruments of RCTs of conventional medications and NPs, with the purpose of identifying items used in published instruments. Details from the portion of the systematic review pertaining to instruments designed for NPs were published previously [2]. Databases (PubMed, Embase, Web of Science, CINAHL, IPA and Cochrane Library) were searched from inception to June 2006 to identify potentially relevant articles that

described the critical appraisal of RCTs of conventional medications. Search terms were: critical appraisal, scale, checklist, instrument, randomized controlled trial, systematic review, meta analysis, tool, form, evidence based medicine, quality control, quality assessment and standards. Searches were limited to articles written in English. Evidence-based medicine websites, textbooks and investigators files were searched for additional instruments. Bibliographies of relevant references were reviewed to identify additional articles.

Full articles of potentially relevant citations were retrieved and evaluated for relevance by two investigators, with disagreements being resolved by a third investigator. Articles were deemed relevant if they contained an instrument used for the assessment of RCTs of conventional medicine, described how to report an RCT of conventional medicine or discussed the importance of quality assessment and provided suggested items essential to the assessment. Articles were excluded if they contained an instrument that had been previously published, contained an instrument designed specifically for the assessment of a meta-analysis, a systematic review or for a trial design other than RCT or contained an instrument designed for the assessment of RCTs of therapies other than medication.

#### *Compilation of items from critical appraisal instruments for conventional medications*

a list of all items contained in assessment instruments for conventional medications and guidance documents was compiled, using a process similar to that used for the NP instruments [2]. A table, using the Revised CONSORT statement [5] as a framework, was used to record all items. Items contained in assessment instruments, but not found in the Revised CONSORT statement, were recorded in a separate section labeled as miscellaneous items. Items from instruments that were reported to have been validated and/or there was empirical evidence provided for the importance of including the items were noted.

#### *Analysis*

The list of all items from conventional instruments was examined for items essential for inclusion in the new critical appraisal instrument. To be designated as essential to include in the new critical appraisal instrument, an item had to meet at least 1 of the following 2 inclusion criteria: it had to have been contained in a published instrument that was documented as having been validated or must have had empirical evidence to support its inclusion in a published instrument. Originally we planned to have a third criterion, which was that the item had to have been included in the more than 50% of critical appraisal instruments. This criterion was removed from consideration for reasons that are provided in the Discussion section.

### **Phase 3. Field test of new assessment instrument**

#### *Purpose*

to evaluate the new assessment instrument by 1) comparing the performance of the new assessment instrument to that of a published assessment instrument in evaluating the quality of RCTs of NPs and helping to answer a clinical question; 2) measuring the validity of each criterion; 3) determining the potential impact of the order of use of the instruments.

#### *Study design*

the randomized trial was approved by the Dalhousie University Health Sciences and Humanities Human Research Ethics Board, with participants and primary researchers (TJ and AMW) being blinded as to which group the participants were assigned. To be eligible to take part in the study, participants had to fit one of the following inclusion criteria: be a practicing pharmacist, be a student currently enrolled in a pharmacy program or be an academic involved in the teaching of critical appraisal skills or clinical skills to pharmacy students. People were excluded from participating if they had been involved in any of the earlier phases of development of the new instrument. A sample size of 14 was determined to be sufficient, based on a similar study [6]. Participants were given 3 weeks to complete the study and return all materials. They received a reminder after 3 weeks if the completed material was not received by the research assistant and provided additional time, if required to complete the assessment.

#### *Materials*

Three published RCTs evaluating the efficacy of soy isoflavones in treating hot flashes were selected, based on having a range of quality of trials for the study. All identifiers, including authors, affiliations and journal citation were removed from the 3 articles. A clinical case describing a woman experiencing hot flashes as a consequence of menopause, was prepared. Participants were asked to apply the results of the RCT to the case by answering the following clinical question: "In a 53-year-old postmenopausal woman experiencing 5 – 10 hot flashes per day would soy isoflavone extract decrease the frequency and severity of the hot flashes?".

The draft new assessment instrument developed in the first 2 phases of this project, along with a User's Guide was provided for the appraisal of each RCT. A published assessment instrument used for reviewing clinical trials of herbal products was selected as a comparator instrument [7]. Permission was obtained from the publisher to use the assessment instrument. All identifiers were removed from both instruments to reduce the risk of bias.

#### *Procedure*

Participants were randomly assigned to 1 of 2 groups using a technique to conceal allocation that involved selecting 1 of 2 colors of paper out of a hat. Study packages

each contained 2 envelopes, one red and one blue. Each envelope contained the clinical case and question, an assessment instrument, a User's Guide, 3 RCTs to be critically assessed and an evaluation form for feedback on the usefulness of the assessment instrument. All participants were instructed to open and complete the red envelope first. Half of the participants received red envelopes containing the new assessment instrument while the red envelopes received by the second group of participants contained the comparator instrument. Participants were told to open and complete the contents of the blue envelope only after completing the contents of the red envelope. After contents of both envelopes were completed, participants were instructed to fill out an evaluation form that asked which instrument they found easiest to use, which one was most helpful identifying strengths and weaknesses of the study and in answering the clinical question and which one was most helpful in selecting an NP comparable to one used in the RCT.

#### *Analysis*

all responses submitted for each item contained in the new critical appraisal instrument for each of the 3 articles were recorded by the research assistant (MM), using participants' codes rather than names to preserve anonymity. To test for criterion validity, Fisher's exact test was used to compare responses for each item submitted by each participant to the investigators "gold standard" answers that were reached by consensus prior to beginning the trial. Data obtained from participants' use of the comparator instrument was analyzed to determine if using the comparator instrument resulted in the same assignment of overall quality of each article as was found with the new appraisal instrument. Fisher's exact test was used to determine if the order in which the instruments were used affected responses to the new assessment instrument. Responses to questions on evaluation sheets were compiled. All comments on the evaluation sheets, including suggestions for improvement, were recorded and examined.

### **Results**

#### **Phase 1. Establishment of a list of items essential in describing the identity of an NP**

##### *Selection of items*

Sixteen items describing an NP or its placebo were selected for inclusion in Round 1 study form (Table 1, column 1). Thirteen of the 16 items were found in 3 published instruments for NPs [3,7,8]. Item #16 (success of blinding) was included in one instrument and was added to the initial list [8]. Item #7 (lot number of product) and item # 15 (description of placebo) were not found in published instruments and were added by investigators.

##### *Participants*

11 of 14 potential participants completed and returned the Round 1 study form. Follow-up letters and Round 2

**Table 1: Results of Delphi Consensus-Round 1**

Item	Compiled Responses			Participant Comments
	Yes	No	Un certain	
1. The genus of the test NP was stated.	9	-	2	<ul style="list-style-type: none"> <li>• Items 1 &amp; 2 assume an organism is used, but there are many chemical NPs, so we need to include an unambiguous name if it is a natural chemical e.g. arginine is ambiguous, L-arginine is correct.</li> </ul>
2. The species of the test NP was stated.	9	-	1	<ul style="list-style-type: none"> <li>• Need to be more specific by listing cultivars/varieties</li> </ul>
3. The plant part used to prepare the test NP was stated.	10	-	1	<ul style="list-style-type: none"> <li>• Also when and where harvested</li> </ul>
4. How the test NP was processed/extracted was described.	9	1	-	<ul style="list-style-type: none"> <li>• Chemistry and manufacturing information is required by the NP Regulations</li> <li>• Proprietary data but for government registration should be included but should not be reported in the publication</li> </ul>
5. If test NP was a commercial product, brand name was stated.	8	1	1	<ul style="list-style-type: none"> <li>• Interesting, but not essential.</li> <li>• Brand is important because of differences in manufacturing.</li> </ul>
6. If test NP was a commercial product, the name of the manufacturer was stated.	9	1	-	
7. If test NP was a commercial product, the lot number was stated.	8	2	-	<ul style="list-style-type: none"> <li>• Make sure they have enough supply for the complete trial and didn't use different lots throughout study</li> </ul>
8. The name of important chemical(s) of the test NP was stated.	11	-	-	<ul style="list-style-type: none"> <li>• The term "important chemicals" is vague</li> <li>• Full characterization or fingerprinting of the testing material</li> <li>• This could be a problem with some products as active agent(s) may not be known</li> <li>• Should have a percentage of the chemicals present listed</li> <li>• If standardized for the chemical content and the manufacturer wants to be public</li> </ul>
9. The amount of important chemical(s) of the test NP was stated.	11	-	-	<ul style="list-style-type: none"> <li>• The term "important chemicals" is vague</li> <li>• If the makers are not active compounds, there is no need to quantify each compound but it is always good to have some quantitative information</li> </ul>
10. The test NP was analyzed for chemical content.	11	-	-	<ul style="list-style-type: none"> <li>• For registration with to government yes but should be agreed upon by investigator and company before publication.</li> </ul>
11. The dosage form of the test NP was stated.	10	-	-	
12. The dose of the test NP was stated.	10	-	-	
13. The frequency of administration of the test NP was stated.	10	-	-	
14. The route of administration of the test NP was stated.	10	-	-	
15. The placebo or comparison treatment and the test NP were matched in terms of taste, smell, appearance and dosing regimens.	10	-	-	
16. The success of blinding was evaluated.	10	-	-	

study forms were e-mailed individually to 11 participants. Eight of 11 participants completed and returned the Round 2 study form. Specific reasons for not completing the study were not provided by dropouts.

*Delphi Consensus-Round 1*

contents of completed study forms were recorded (Table 1). In general, there was a significant amount of agreement with the list of items. In total, participants unanimously agreed on the inclusion of 9 of 16 items. The remaining 7 items each had 1 or 2 participants who either were uncertain or did not agree that the item was essential to include in the list. Participants offered suggestions for clarification of wording of some items. Several partici-

pants suggested the inclusion of additional items that they felt were missing from the list, such as the stability of the NP, whether the NP was prepared according to monograph specifications, if a certificate of analysis was provided, where and when the plant was harvested and details of the growing conditions.

*Delphi Consensus-Round 2*

the Round 2 form was created based on review of Round 1 results. There was at least 80% agreement on all items included on the Round 2 form. Thus these 15 items were considered by experts and investigators as essential. Twelve of the 15 items that pertained specifically to the description of the NP were grouped together (a-h) (Figure 1). Two

other items were grouped together because they were specific to the matching of the NP and placebo/comparator.

### **Phase 2. Establishment of important non-NP items**

#### *Identification of critical appraisal instruments for RCTs of conventional medications*

The initial search of databases and hand searching of investigators files, websites, textbooks, and reference lists for critical appraisal instruments for RCTs of conventional medications and NPs identified 4442 citations (Figure 2). A review of the titles of the citations resulted in the exclusion of 4032 citations. Review of the remaining 410 abstracts identified 229 potentially relevant articles, 200 of which were for instruments specifically for RCTs of conventional medication. The remaining 29 potentially relevant articles pertained to instruments for RCTs of NPs and were systematically reviewed separately [2]. Upon further examination, 99 of the 200 articles met the inclusion criteria for instruments for RCTs of conventional medication [4,6,9-105]. The 99 articles consisted of 87 instruments for RCTs of conventional medications and 12 guidance documents written to guide reporting of an RCT and/or discussed the importance of quality assessment of RCTs of conventional medicine.

#### *Compilation of items from critical appraisal instruments for conventional medications*

4 of the 99 instruments stated that they had been validated [6,28,63,104]. Therefore, all 17 general item categories contained in the validated instruments were designated for inclusion in the new instrument (Table 2). Four of the 17 items categories: sequence generation (for randomization), blinding, participant flow and baseline data were further supported for inclusion by empirical evidence. [37,40,106-110]

The new instrument for the critical appraisal of RCTs of NPs was assembled by combining the 17 general item categories identified as being important to assess with conventional medications (Table 2) with the essential items specific for the content of NPs determined in the Delphi consensus process (Figure 1). In some cases, item statements were further divided to provide clarity. For example, the general item category of "blinding" was divided into two items in the new assessment instrument: item #7. (The study was blinded) and item #24 (The success of blinding was evaluated). A section titled "reviewer's conclusion" was added to the end of the instrument to allow the reviewer to reflect on their findings after the evaluation of an article using the new instrument is complete. This included 2 questions to help the user apply the results to their practice. A User's Guide was developed to facilitate the correct and consistent use of the instrument. It included a section describing the purpose of the instrument and definition of each item

### **Phase 3. Field test of new assessment instrument**

Fourteen participants were enrolled in the RCT, pharmacists (n = 6), academics (n = 3) and pharmacy students (n = 5). Thirteen of 14 participants completed the study; 6 in the group that used the new assessment instrument first and 7 in the group that used the comparator instrument before using the new assessment instrument. The one dropout was unable to complete the study within the specified time frame due to unforeseen circumstances.

Responses, for each of the 3 articles, to each item in the new assessment instrument were recorded and compared between participants, regardless of the order in which the instrument was used. There was general agreement among participants as to the answers for most of the items in the new assessment instrument, for all 3 articles. Items producing the most discrepancy in answers were those dealing with the description of the NP.

Statistical analysis of participants' responses for each item in the new assessment instrument to the gold standard responses of the investigators for each of the 3 papers revealed that there was no significant difference in answers, indicating strong criterion validity. Statistical analysis showed that the order in which the 2 instruments were used had no significant effect on responses to any of the items in the new assessment instrument.

The critical appraisal of the overall quality of each paper produced similar results, regardless of which instrument was used. That is, the paper that was judged to be of best quality of the three papers analyzed using the comparator instrument was also found by participants to be of the best quality using the new instrument

Participants' responses to the "Reviewer's Conclusion" section of the new assessment instrument were compared to the gold standard responses of the investigators. Eight of 13 participants (50% of those who used the new assessment instrument first and 71% of those who used the new assessment instrument second) felt that there was enough detail provided in Paper A (not in Paper B or C) to allow them to select a similar product, therefore agreeing with the gold standard. The group using the new assessment instrument second reported feeling much more comfortable in finding a product to recommend.

Participants who used the published assessment instrument first were fairly evenly split as to whether, after reviewing the papers, they would recommend the treatment to their patients. This was in contrast to those from the group who used the published assessment instrument after having used the new assessment instrument. They were more likely to say that they would not recommend the treatment.

The description of the NP included:

- (a) If the NP is (or was made from) a plant, microorganism or animal:
  - I. The genus and species
  - II. The part of the plant, microorganism or animal that was used
  - II. How the NP was prepared (extracted or processed)
- (b) If the NP was composed of one or more chemical constituent(s), either derived from a plant, microorganism or animal or produced synthetically:
  - I. The complete name of the chemical constituent(s)
- (c) If the NP was a commercially prepared product (regardless if it fits into category (a) or (b):
  - I. The brand name
  - II. The manufacturer
  - III. The lot or batch number(s)
- (d) The name(s) of the active or marker chemical(s)
- (e) The amount (weight or %) of the active or marker chemical(s)
- (f) A qualitative and quantitative analysis of active or marker chemical(s)
- (g) The dosage form
- (h) The daily dose

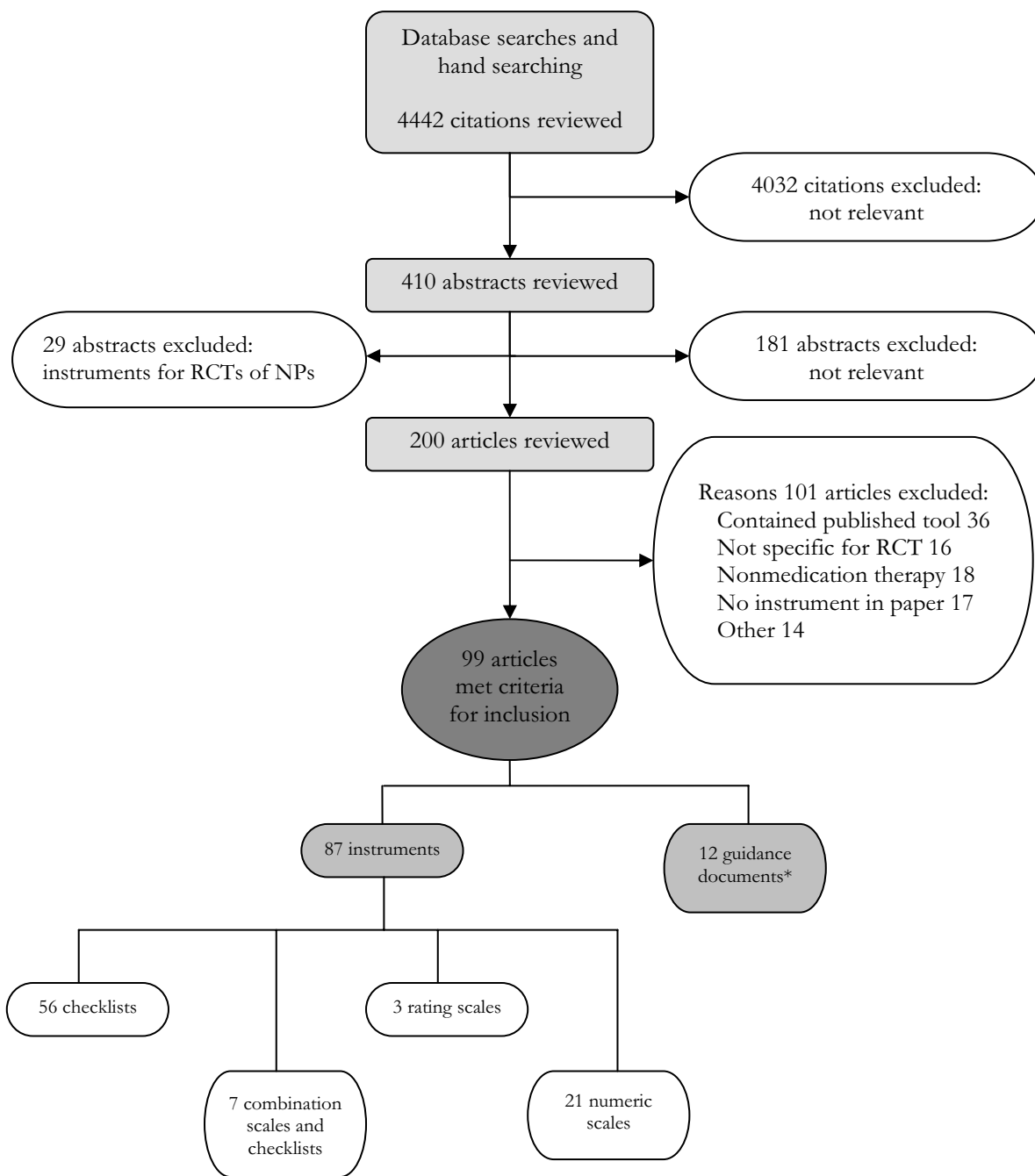
The NP and placebo/comparison treatment were matched in terms of:

- (a) Taste, smell and / or appearance
- (b) Dosing regimen

The success of blinding was evaluated

**Figure 1**  
**Final list of items essential in describing the identity of an NP.**

---



\* Guidance Documents: Articles detailing how to report an RCT and/or discussing importance of quality assessment and providing suggested items of importance.

**Figure 2**  
**Results of search for critical appraisal instruments for RCTs of conventional medications.**



**Table 2: Support for inclusion of items in critical appraisal instruments for RCTs of conventional medicine.**

General Item Categories	Part of validated tool	Empirical evidence
Participants	yes	
Interventions	Yes	
Objectives	Yes	
Outcomes (Methods)	Yes	
Sample size	Yes	
Sequence generation (Randomization)	Yes	Yes
Allocation concealment (Randomization)	Yes	
Blinding	Yes	Yes
Statistical flow	Yes	
Participant flow	Yes	Yes
Baseline data	Yes	Yes
Numbers analyzed	Yes	
Outcomes and estimation (Results)	Yes	
Ancillary analysis	Yes	
Adverse effects	Yes	
Interpretation	Yes	
Generalizability	Yes	

The new assessment instrument was favored by participants in 3 out of 4 questions that participants completed as part of the Evaluation Form. Eleven of 13 found the new instrument to be most helpful identifying strengths and weaknesses of the studies, 8 of 13 felt that the new instrument was best at helping to answer the clinical question and 11 of 12 found the new instrument the most helpful in selecting an NP comparable to one used in the RCT. Participants were fairly evenly split as to which instrument they found easier to use. Overall, 12 of 13 participants preferred the new assessment instrument, stating that they liked the detailed questions about the identity of the NP. Feedback from the participant who preferred the comparator assessment instrument reported liking the numerical score that was generated by using the comparator instrument as an indication of quality. The new assessment instrument was favored over the published assessment instrument by the group who used the new instrument second, however, the participants who used the new assessment instrument first were less clear as to which instrument they preferred.

The new critical appraisal instrument was refined, based on participants' feedback and is included as Additional file 1. The final version of the User's Guide is presented in Additional file 2.

**Discussion**

**Development of instrument**

The Delphi process is a well recognized technique used to reach consensus among a group of experts [111]. Using content experts in the Delphi process to establish a list of essential NP items to include in the critical appraisal

instrument ensured content validity of the items [111-113].

Results from Round 1 showed that participants, in general, agreed with the inclusion of the majority of the items in the new critical appraisal instrument. The depth and breadth of expertise of participants was evident in suggestions of additional items and changes in wording to improve clarity. The most challenging issue raised by participants in Round 1 was to identify the correct terminology to use for the chemicals contained in the NP. Items in the initial list asked for reports of RCTs to name "important" chemicals, their amounts and state whether the NP was analyzed for chemical content. Participants agreed that these items were essential but reported that the term "important chemicals" was vague or unclear. Assigning an unambiguous term to these chemicals is complicated by the fact that in some NPs, chemicals many have been determined to be "pharmacologically active", while in other NPs, in the active chemicals are unknown and so product content is standardized to marker chemicals.

Round 2 was used to achieve consensus on rewording of items and additional items as per suggestions from participants. The list of items was also reorganized for ease of use. Results of Round 2, coupled with investigators input, resulted in the list of essential items presented in Figure 1, which will become part of the new instrument for critical appraisal of RCTs of NPs. While the list contained many of the items found in other instruments, the Delphi consensus process, using the expertise of 11 participants, provided a forum for ensuring all items were essential to clearly define the NP being evaluated, were in language

that was clear and unambiguous to users with a variety of backgrounds and were presented in a user friendly format.

Identification of non-NP items for the instrument was achieved by examining the content of 99 published instruments and guidance documents for the critical appraisal of conventional medications. Initially it was proposed that the items that appeared most frequently in the instruments would be likely to be the ones of most importance. While good in theory, there were two limitations with relying exclusively on frequency as an indicator of importance. Firstly, although there were 99 instruments and guidance documents to examine, the actual number of unique instruments was much smaller. This was a result of the fact that some instruments were not truly new instruments but were actually adaptations of earlier instruments. We felt that it was important to include these adapted instruments in our analysis as, the fact that researchers felt their initial content was worthwhile to retain and modify suggested their value. It was also important to identify what adaptations had been made as there was the possibility of identifying new items. The limitation of including these adapted instruments in our analysis was that there was the potential that some items could have been carried along into the revised instruments, increasing the frequency score of the item overall, but without the importance of the item being adequately assessed in the revision of the instrument. The second limitation of relying on frequency as the only indicator of an item's importance is that as understanding of critical appraisal develops, new items determined to be essential have been and will be added to new instruments. Because the items are new, their overall frequency of inclusion in instruments will be low and, by our criteria, would not have been included in our new instrument. It was for these two reasons that frequency of inclusion in an instrument was removed from the list of criteria for including non-NP items in our new instrument.

While frequency of inclusion of items in instruments is an indicator of their perceived importance, to overcome some of the inherent limitations described above, it was decided to include items that had empirical evidence to support their inclusion or were contained in instruments that said they had been validated. Only 4 of the 99 instruments were described by their authors as having been validated [6,28,63,104]. Interestingly, these 4 instruments were also designed using empirical methods of development including revisions based on inter-rater agreement or comments on content validity [6,63], Delphi consensus [28], and a modified Nominal Group Technique [104].

#### **Evaluation of instrument**

The new assessment instrument, developed using a Delphi process and defined criteria to ensure that items had strong support for inclusion, was tested by anticipated

users of a critical appraisal instrument. This included pharmacists, pharmacy educators and pharmacy students, each with a range of experience in critical appraisal and knowledge of issues important to NPs. An important part of the evaluation was to have participants be able to compare the new instrument to a published instrument that was designed specifically for use with NPs and that all identifiers from articles and instruments be removed to reduce the risk of bias.

Items in the new assessment instrument producing the most discrepancy in answers dealt with the identity of the NP. The range in answers was likely a result of lack of clarity in the items and so each of the NP items was examined and modified if required, to reduce ambiguity.

Randomizing the order in which participants used the 2 assessment instruments allowed us to probe whether the order affected their responses to items in the new assessment instrument. The fact that the order of use had no statistically significant affect on answers to items on the new assessment instrument allowed us to pool answers from both groups when assessing criterion validity by comparing participants' answers to the gold standard, thereby giving a larger sample size.

#### **Strengths and Limitations**

The strength of the new assessment instrument described in this paper is the method used for its development. Identification and assessment of items for inclusion in the new assessment was based on clearly defined criteria. The use of experts in the Delphi process to guide the development of the list of items essential to include for NPs was clearly a strength, as was the evaluation of the instrument to validate criterion. This is an advantage over all published instruments for NPs that we identified.

A potential limitation of the instrument is that the initial list of both NP and non-NP items was dependent on the content of assessment instruments identified in the literature. The literature search was completed in June 2006 and it is possible, despite the fact that the investigators routinely monitor the newly published literature in this area, that newer instruments were not identified. It is also possible, due to the complexity of the nature of the database searching on this topic that not all relevant instruments were identified. Search results revealed that there was inconsistency in the key words used to index articles containing instruments and guidance documents. Both the time lag since the initial search and the difficulty in searching the databases could have resulted in some instruments not being identified thereby affecting the list of items. Additionally, because the search of the literature was limited to articles published in English, this may have also resulted in an incomplete identification of articles.

The number of items contained in the new instrument may be viewed as both a strength and a limitation. The number of items attests to its attempt to be complete in its assessment of the RCTs, however, some potential users may find the time needed to address each item prohibitive.

## Conclusion

A new instrument for the critical appraisal of RCTs of NPs was developed and evaluated. The NP section of the instrument was developed through consensus of a group of experts in NP and critical appraisal, ensuring content validity. The non-NP section of the instrument was constructed using items that were contained in instruments that had been validated or had empirical evidence to support their inclusion. The new assessment instrument was tested by a group of potential users who determined it to be helpful in evaluating RCTs to aid in answering a clinical question, as compared to a published comparator instrument. The instrument is distinct from other available assessment instruments for RCTs of NPs in its systematic development and validation. It has the additional advantage that its use is not restricted to herbs, rather it can be used with for all "single entity" NPs, such as glucosamine for example. The instrument is ready to be used by pharmacy students, health care practitioners and academics and will continue to be refined as required.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

TJ and AMW designed all studies, interpreted data and drafted manuscript. MMM conducted Delphi consensus rounds and the instrument evaluation trial. LL was responsible for database searching, relevance assessment and data extraction and some interpretation. All authors read and approved manuscript.

## Additional material

### Additional file 1

Appendix A.

*Appendix A. Final version of new assessment instrument developed for the critical appraisal of RCTs of NPs.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1472-6882-9-11-S1.doc>]

### Additional file 2

Appendix B.

*Appendix B. User's Guide for Dalhousie Assessment Instrument for Critical Appraisal of Randomized Controlled Trials (RCTs) of Natural Products (NPs).*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1472-6882-9-11-S2.doc>]

## Acknowledgements

Funding for the project and studentship support was provided by the Dalhousie Pharmacy Endowment Fund and a Dalhousie University Faculty of Health Professions Research Development Grant. Thanks are extended to Dr. David Gardner and Dr Richard Braha for helping with early project development and to Angela Hazel and Stephanie Keith for their contributions to the project. The statistical analysis was provided by Vaneeta Kaur Grover.

## References

- Chan E: **Quality of efficacy research in Complementary and Alternative Medicine.** Editorial. *JAMA* 2008, **299(22)**:2685-2686.
- Whelan AM, Jurgens TM, Lord L: **Evaluating the quality of randomized controlled trials that examine the efficacy of natural health products: a systematic review of critical appraisal instruments.** *Evid Based Complement Alternat Med* 2008.
- Gagnier JJ, Boon H, Rochon P, Moher D, Barnes J, Bombardier C, CONSORT Group: **Reporting randomized, controlled trials of herbal interventions: an elaborated CONSORT statement.** *Ann Intern Med* 2006, **144(5)**:364-367.
- Verhagen AP, de Vet HC, de Bie RA, Kessels AG, Boers M, Bouter LM, Knipschild PG: **The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus.** *J Clin Epidemiol* 1998, **51(12)**:1235-1241.
- Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, Gotzsche PC, Lang T, CONSORT GROUP (Consolidated Standards of Reporting Trials): **The revised CONSORT statement for reporting randomized trials: explanation and elaboration.** *Ann Intern Med* 2001, **134(8)**:663-694.
- Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, McQuay HJ: **Assessing the quality of reports of randomized clinical trials: is blinding necessary?** *Control Clin Trials* 1996, **17(1)**:1-12.
- Low Dog T: **Clinical Trial Reviewer's Guidance and Checklist.** In *The Handbook of Clinically Tested Herbal Remedies Volume 1*. Edited by: Barrett M. Birmingham, New York: The Haworth Press, Inc; 2004:141-148.
- Gardner DM: **Evidence-based decisions about herbal products for treating mental disorders.** *J Psychiatry Neurosci* 2002, **27(5)**:324-333.
- Turlik MA, Kushner D, Stock D: **Assessing the validity of published randomized controlled trials in podiatric medical journals.** *J Am Podiatr Med Assoc* 2003, **93(5)**:392-398.
- Vamvakas EC: **Rationale, objectives, and interpretation of randomized controlled trials.** *J Clin Apher* 1997, **12(3)**:130-139.
- Heijden GJ van der, Windt DA van der, Kleijnen J, Koes BW, Bouter LM: **Steroid injections for shoulder disorders: a systematic review of randomized clinical trials.** *Br J Gen Pract* 1996, **46(406)**:309-316.
- van Tulder MW, Koes BW, Bouter LM: **Conservative treatment of acute and chronic nonspecific low back pain. A systematic review of randomized controlled trials of the most common interventions.** *Spine* 1997, **22(18)**:2128-2156.
- Weintraub M: **How to Critically Assess Clinical Drug Trials.** *Drug Ther* 1982:131-148.
- Working Group on Recommendations for Reporting of Clinical Trials in the Biomedical Literature: **Call for comments on a proposal to improve reporting of clinical trials in the biomedical literature.** *Ann Intern Med* 1994, **121(11)**:894-895.
- Zelen M: **Guidelines for publishing papers on cancer clinical trials: responsibilities of editors and authors.** *J Clin Oncol* 1983, **1(2)**:164-169.
- Nurmohamed MT, Rosendaal FR, Buller HR, Dekker E, Hommes DW, Vandenbroucke JP, Briet E: **Low-molecular-weight heparin versus standard heparin in general and orthopaedic surgery: a meta-analysis.** *Lancet* 1992, **340(8812)**:152-156.
- Onghena P, Van Houdenhove B: **Antidepressant-induced analgesia in chronic non-malignant pain: a meta-analysis of 39 placebo-controlled studies.** *Pain* 1992, **49(2)**:205-219.
- Pater JL, Weir L: **Reporting the results of randomized trials of empiric antibiotics in febrile neutropenic patients – a critical survey.** *J Clin Oncol* 1986, **4(3)**:346-352.

19. Pua HL, Lerman J, Crawford MW, Wright JG: **An evaluation of the quality of clinical trials in anesthesia.** *Anesthesiology* 2001, **95(5)**:1068-1073.
20. Reisch JS, Tyson JE, Mize SG: **Aid to the evaluation of therapeutic studies.** *Pediatrics* 1989, **84(5)**:815-827.
21. Rochon PA, Gurwitz JH, Cheung CM, Hayes JA, Chalmers TC: **Evaluating the quality of articles published in journal supplements compared with the quality of those published in the parent journal.** *JAMA* 1994, **272(2)**:108-113.
22. Rothwell PM: **External validity of randomised controlled trials: "to whom do the results of this trial apply?"**. *Lancet* 2005, **365(9453)**:82-93.
23. Sackett DL: **How to read clinical journals: V: To distinguish useful from useless or even harmful therapy.** *Can Med Assoc J* 1981, **124(9)**:1156-1162.
24. Schulz KF, Chalmers I, Hayes JA, Altman DG: **Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials.** *JAMA* 1995, **273(5)**:408-412.
25. **Scottish Intercollegiate Guidelines Network** [<http://www.sign.ac.uk/methodology/checklists.html>]
26. Seals DR, Tanaka H: **Manuscript peer review: a helpful checklist for students and novice referees.** *Adv Physiol Educ* 2000, **23(1)**:52-58.
27. Simon R, Wittes RE: **Methodologic guidelines for reports of clinical trials.** *Cancer Treat Rep* 1985, **69(1)**:1-3.
28. Sindhu F, Carpenter L, Seers K: **Development of a tool to rate the quality assessment of randomized controlled trials using a Delphi technique.** *J Adv Nurs* 1997, **25(6)**:1262-1268.
29. Stoddard GJ, Ring WH: **How to evaluate study methodology in published clinical research.** *J Intraven Nurs* 1993, **16(2)**:110-117.
30. Strippoli GF, Craig JC, Schena FP: **The number, quality, and coverage of randomized controlled trials in nephrology.** *J Am Soc Nephrol* 2004, **15(2)**:411-419.
31. The Asilomar Working Group on Recommendations for Reporting of Clinical Trials in the Biomedical Literature: **Checklist of information for inclusion in reports of clinical trials.** *Ann Intern Med* 1996, **124(8)**:741-743.
32. The Standards of Reporting Trials Group: **A proposal for structured reporting of randomized controlled trials.** *JAMA* 1994, **272(24)**:1926-1931.
33. Harrison JE: **Clinical trials in orthodontics II: assessment of the quality of reporting of clinical trials published in three orthodontic journals between 1989 and 1998.** *J Orthod* 2003, **30(4)**:309-15. discussion 297-8.
34. **What is ... ? series** [<http://www.whatisseries.co.uk/whatis/>]
35. Huwiler-Muntener K, Juni P, Junker C, Egger M: **Quality of reporting of randomized trials as a measure of methodologic quality.** *JAMA* 2002, **287(21)**:2801-2804.
36. Imperiale TF, McCullough AJ: **Do corticosteroids reduce mortality from alcoholic hepatitis? A meta-analysis of the randomized trials.** *Ann Intern Med* 1990, **113(4)**:299-307.
37. Juni P, Altman DG, Egger M: **Systematic reviews in health care: Assessing the quality of controlled clinical trials.** *BMJ* 2001, **323(7303)**:42-46.
38. Karassa FB, Tatsioni A, Ioannidis JP: **Design, quality, and bias in randomized controlled trials of systemic lupus erythematosus.** *J Rheumatol* 2003, **30(5)**:979-984.
39. Khan KS, Daya S, Jadad A: **The importance of quality of primary studies in producing unbiased systematic reviews.** *Arch Intern Med* 1996, **156(6)**:661-666.
40. Kjaergard LL, Frederiksen SL, Gluud C: **Validity of randomized clinical trials in gastroenterology from 1964-2000.** *Gastroenterology* 2002, **122(4)**:1157-1160.
41. Kjaergard LL, Nikolova D, Gluud C: **Randomized clinical trials in hepatology: predictors of quality.** *Hepatology* 1999, **30(5)**:1134-1138.
42. Levine J: **Trial Assessment Procedure Scale.** In *Guide to Clinical Trials* Edited by: Spilker B. New York: Raven Press, Ltd; 1991:780-786.
43. Liberati A, Himel HN, Chalmers TC: **A quality assessment of randomized control trials of primary treatment of breast cancer.** *J Clin Oncol* 1986, **4(6)**:942-951.
44. Liddle J, Williamson M, Irwig L: **Method for Evaluating Research Guideline Evidence (MERGE).** Sydney: NSW Health Department; 1996.
45. Lionel ND, Herxheimer A: **Assessing reports of therapeutic trials.** *Br Med J* 1970, **3(723)**:637-640.
46. Lohr KN, Carey TS: **Assessing "best evidence": issues in grading the quality of studies for systematic reviews.** *Jt Comm J Qual Improv* 1999, **25(9)**:470-479.
47. Lopex LR: **What are the essential factors in a critique of a research study for determination of quality and applicability?** *Fla Nurse* 1987, **35(6)**:25.
48. Mahon WA, Daniel EE: **A method for the assessment of reports of drug trials.** *Can Med Assoc J* 1964, **90**:565-569.
49. Makela M, Witt K: **How to read a paper: critical appraisal of studies for application in healthcare.** *Singapore Med J* 2005, **46(3)**:108-114. quiz 115.
50. Meade MO, Richardson WS: **Selecting and appraising studies for a systematic review.** *Ann Intern Med* 1997, **127(7)**:531-537.
51. Meinert CL: **Questions and factors to consider when reading a report from a clinical trial.** In *Clinical Trials: Design, Conduct, and Analysis* Edited by: Meinert CL. New York: Oxford University Press; 1986:272-276.
52. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, Tugwell P, Klassen TP: **Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses?** *Lancet* 1998, **352(9128)**:609-613.
53. Moher D, Schulz KF, Altman DG, CONSORT GROUP (Consolidated Standards of Reporting Trials): **The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials.** *Ann Intern Med* 2001, **134(8)**:657-662.
54. Mosteller F, Gilbert JP, McPeck B: **Reporting standards and research strategies for controlled trials. Agenda for the editor.** *Control Clin Trials* 1980, **1**:37-58.
55. NHMRC (National Health and Medical Research Council): *How to review the evidence: systematic identification and review of the scientific literature* Canberra: NHMRC; 2000.
56. **Public Health Resource Unit** [[http://www.phru.nhs.uk/Doc\\_Links/rct%20appraisal%20tool.pdf](http://www.phru.nhs.uk/Doc_Links/rct%20appraisal%20tool.pdf)]
57. DerSimonian R, Charette LJ, McPeck B, Mosteller F: **Reporting on methods in clinical trials.** *N Engl J Med* 1982, **306(22)**:1332-1337.
58. Detsky AS, Naylor CD, O'Rourke K, McGeer AJ, L'Abbe KA: **Incorporating variations in the quality of individual randomized trials into meta-analysis.** *J Clin Epidemiol* 1992, **45(3)**:255-265.
59. Devereaux PJ, Manns BJ, Ghali WA, Quan H, Guyatt GH: **The reporting of methodological factors in randomized controlled trials and the association with a journal policy to promote adherence to the Consolidated Standards of Reporting Trials (CONSORT) checklist.** *Control Clin Trials* 2002, **23(4)**:380-388.
60. Dixon JS, Smith A, Evans SJ: **Reporting clinical trials.** *Br J Rheumatol* 1983, **22(3 Suppl)**:74-78.
61. Djulbegovic B, Adams JR, Lyman GH, Lacey M, Hozo I, Greenwich M, Bennett CL: **Evaluation and appraisal of randomized controlled trials in myeloma.** *Ann Oncol* 2001, **12(11)**:1611-1617.
62. Doig GS: **Interpreting and using clinical trials.** *Crit Care Clin* 1998, **14(3)**:513-524.
63. Downs SH, Black N: **The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions.** *J Epidemiol Community Health* 1998, **52(6)**:377-384.
64. Ellenberg JH: **Clinical trials.** *Neurol Clin* 1990, **8(1)**:15-30.
65. Evans M, Pollock AV: **A score system for evaluating random control clinical trials of prophylaxis of abdominal surgical wound infection.** *Br J Surg* 1985, **72(4)**:256-260.
66. Evans M, Pollock AV: **Trials on trial. A review of trials of antibiotic prophylaxis.** *Arch Surg* 1984, **119(1)**:109-113.
67. **The Focus Project.** [www.focusproject.org.uk](http://www.focusproject.org.uk). The above URL is no longer active, for further information please contact the corresponding author.
68. Fowkes FG, Fulton PM: **Critical appraisal of published research: introductory guidelines.** *BMJ* 1991, **302(6785)**:1136-1140.
69. Gardner MJ, Machin D, Campbell MJ: **Use of check lists in assessing the statistical content of medical studies.** *Br Med J (Clin Res Ed)* 1986, **292(6523)**:810-812.
70. Gifford RH, Feinstein AR: **A critique of methodology in studies of anticoagulant therapy for acute myocardial infarction.** *N Engl J Med* 1969, **280(7)**:351-357.

71. Gluud C, Nikolova D: **Quality assessment of reports on clinical trials in the Journal of Hepatology.** *J Hepatol* 1998, **29(2)**:321-327.
72. Gotsche PC: **Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal antiinflammatory drugs in rheumatoid arthritis.** *Control Clin Trials* 1989, **10(1)**:31-56.
73. Grant A: **Reporting controlled trials.** *Br J Obstet Gynaecol* 1989, **96(4)**:397-400.
74. Grimes DA, Schulz KF: **Methodology citations and the quality of randomized controlled trials in obstetrics and gynecology.** *Am J Obstet Gynecol* 1996, **174(4)**:1312-1315.
75. Gross R, Keller DR: **Appraising the evidence.** *WMJ* 1999, **98(2)**:44-48.
76. Gupta AK, Ryder JE, Bluhm R, Johnson A, Summerbell RC: **Oncychomycosis: quality of studies.** *J Cutan Med Surg* 2003, **7(4)**:312-316.
77. Guyatt GH, Sackett DL, Cook DJ: **Users' guides to the medical literature. II. How to use an article about therapy or prevention. B. What were the results and will they help me in caring for my patients? Evidence-Based Medicine Working Group.** *JAMA* 1994, **271(1)**:59-63.
78. Guyatt GH, Sackett DL, Cook DJ: **Users' guides to the medical literature. II. How to use an article about therapy or prevention. A. Are the results of the study valid? Evidence-Based Medicine Working Group.** *JAMA* 1993, **270(21)**:2598-2601.
79. Harris RP, Helfand M, Woolf SH, Lohr KN, Mulrow CD, Teutsch SM, Atkins D, Methods Work Group, Third US Preventive Services Task Force: **Current methods of the US Preventive Services Task Force: a review of the process.** *Am J Prev Med* 2001, **20(3 Suppl)**:21-35.
80. **Essential Evidence Plus** [[http://www.infopoems.com/product/methods\\_validity.cfm#](http://www.infopoems.com/product/methods_validity.cfm#)]
81. **Canadian Pharmacist's Letter** [<http://www.pharmacistsletter.com>]
82. **University of Glasgow** [[http://www.gla.ac.uk/media/media\\_64037\\_en.pdf](http://www.gla.ac.uk/media/media_64037_en.pdf)]
83. **Faculty of Medicine and Dentistry, University of Alberta** [<http://www.ebm.med.ualberta.ca/therapy.pdf>]
84. **Centre for Evidence Based Medicine** [<http://www.cebm.net/index.aspx?o=1157>]
85. Ah-See KW, Molony NC: **A qualitative assessment of randomized controlled trials in otolaryngology.** *J Laryngol Otol* 1998, **112(5)**:460-463.
86. Ajetunmbi O: **Critical appraisal of randomized clinical trials.** In *Making Sense of Critical Appraisal* Edited by: Ajetunmbi O. London: Arnold; 2002:122-145.
87. Antczak AA, Tang J, Chalmers TC: **Quality assessment of randomized control trials in dental research. I. Methods.** *J Periodontal Res* 1986, **21(4)**:305-314.
88. Avis M: **Reading research critically. II. An introduction to appraisal: assessing the evidence.** *J Clin Nurs* 1994, **3(5)**:271-277.
89. Balk EM, Bonis PA, Moskowitz H, Schmid CH, Ioannidis JP, Wang C, Lau J: **Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials.** *JAMA* 2002, **287(22)**:2973-2982.
90. Bath FJ, Owen VE, Bath PM: **Quality of full and final publications reporting acute stroke trials: a systematic review.** *Stroke* 1998, **29(10)**:2203-2210.
91. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz KF, Simel D, Stroup DF: **Improving the quality of reporting of randomized controlled trials. The CONSORT statement.** *JAMA* 1996, **276(8)**:637-639.
92. Bigby M, Gadenne AS: **Understanding and evaluating clinical trials.** *J Am Acad Dermatol* 1996, **34(4)**:555-90. quiz 591-3.
93. Bigby M, Stern RS, Bigby JA: **An evaluation of method reporting and use in clinical trials in dermatology.** *Arch Dermatol* 1985, **121(11)**:1394-1399.
94. Bland JM, Jones DR, Bennett S, Cook DG, Haines AP, MacFarlane AJ: **Is the clinical trial evidence about new drugs statistically adequate?** *Br J Clin Pharmacol* 1985, **19(2)**:155-160.
95. Brown CG, Kelen GD, Moser M, Moeschberger ML, Rund DA: **Methodology reporting in three acute care journals: replication and reliability.** *Ann Emerg Med* 1985, **14(10)**:986-991.
96. Brown CG, Moser M, Moeschberger M, Kelen GD: **Controlled clinical trials in emergency medicine.** *Am J Emerg Med* 1985, **3(1)**:67-73.
97. Byers JF, Beaudin CL: **Critical appraisal tools facilitate the work of the quality professional.** *J Healthc Qual* 2001, **23(5)**:35-8. 40-3.
98. Carneiro AV: **Critical appraisal of clinical studies: practical tables.** *Rev Port Cardiol* 2005, **24(4)**:607-614.
99. **cebmh** [[http://cebmh.warne.ox.ac.uk/cebmh/downloads/education\\_resources/overview.pdf](http://cebmh.warne.ox.ac.uk/cebmh/downloads/education_resources/overview.pdf)]
100. Chalmers I, Adams M, Dickersin K, Hetherington J, Tarnow-Mordi W, Meinert C, Tonascia S, Chalmers TC: **A cohort study of summary reports of controlled trials.** *JAMA* 1990, **263(10)**:1401-1405.
101. Chalmers TC, Smith H Jr, Blackburn B, Silverman B, Schroeder B, Reitman D, Ambroz A: **A method for assessing the quality of a randomized control trial.** *Control Clin Trials* 1981, **2(1)**:31-49.
102. Chan AW, Altman DG: **Epidemiology and reporting of randomised trials published in PubMed journals.** *Lancet* 2005, **365(9465)**:1159-1162.
103. Chan KB, Man-Son-Hing M, Molnar FJ, Laupacis A: **How well is the clinical importance of study results reported? An assessment of randomized controlled trials.** *CMAJ* 2001, **165(9)**:1197-1202.
104. Cho MK, Bero LA: **Instruments for assessing the quality of drug studies published in the medical literature.** *JAMA* 1994, **272(2)**:101-104.
105. Colditz GA, Miller JN, Mosteller F: **How study design affects outcomes in comparisons of therapy. I: Medical.** *Stat Med* 1989, **8(4)**:441-454.
106. Schulz KF: **Subverting randomization in controlled trials.** *JAMA* 1995, **274(18)**:1456-1458.
107. Noseworthy JH, Ebers GC, Vandervoort MK, Farquhar RE, Yetisir E, Roberts R: **The impact of blinding on the results of a randomized, placebo-controlled multiple sclerosis clinical trial.** *Neurology* 1994, **44**:16-20.
108. Egger M, Juni P, Bartlett C: **Value of flow diagrams in reports of randomized controlled trials.** *JAMA* 2001, **285**:1996-1999.
109. May GS, Chir B, Demets DL, Friedman LM, Furberg C, Passamani E: **The randomized clinical trial: bias in analysis.** *Circulation* 1981, **64**:669-673.
110. Sackett DL, Gent M: **Controversy in counting and attributing events in clinical trials.** *New Eng J Med* 1979, **301**:1410-1412.
111. Hasson F, Keeney S, McKenna H: **Research guidelines for the Delphi survey technique.** *J Adv Nurs* 2000, **32(4)**:1008-15.
112. Frost MH, Reeve BB, Liepa AM, Stauffer JW, Hays RD, the Mayo/FDA Patient-Reported Outcomes Consensus Meeting Group: **What is sufficient evidence for the reliability and validity of patient-reported outcome measures? Value Health** 2007, **10(Suppl 2)**:S94-S105.
113. Burns KEA, Duffett M, Kho ME, Meade MO, Adhikari NKJ, Sinuff T, Cook DJ, Abenheim L, Sinuff T, Cook DJ: **A. guide for the design and conduct of self-administered surveys of clinicians.** *CMAJ* 2008, **179(3)**:245-252.

## Pre-publication history

The pre-publication history for this paper can be accessed here:

<http://www.biomedcentral.com/1472-6882/9/11/prepub>