Improving the quality of randomized controlled trials in Chinese herbal medicine, part I: clinical trial design and methodology

Zhao-Xiang BIAN, You-Ping LI, David MOHER¹, Simon DAGENAIS¹, Liang LIU, Tai-Xiang WU, Jiang-Xia MIAO¹, Andrew K.L. KWAN², Lisa SONG³

1. School of Chinese Medicine, Hong Kong Baptist University, Hong Kong SAR, China; 2. Department of Clinical Epidemiology, Chinese Evidence-Based Medicine Centre, West China Hospital, Sichuan University, Chengdu, Sichuan Province 610041, China; 3. Chalmers Research Group, Children’s Hospital of Eastern Ontario Research Institute, Ottawa, Canada; 4. School of Chinese Medicine, Chinese University of Hong Kong, Hong Kong SAR, China

ABSTRACT  Objective: To discuss the quality of randomized controlled trials (RCTs) in Chinese herbal medicine (CHM) with respect to design and methodology, and provide suggestions for further improvement in future clinical trials. Methods: A search of the Cochrane Library was conducted to identify RCTs of CHM on line in July 2005. Quality of the RCTs was assessed using a 11-item checklist modified from the revised CONSORT statement, with 2 items specific to CHM (i.e., herb preparation form and quality control of herbs). Results: The search yielded 167 RCTs that were selected for assessment. All trials included statements about the interventions, objectives, primary outcome design, statistical methods, and herb preparation form. Although...
163 (97.6%) trials reported inclusion criteria, exclusion criteria were only reported in 26 (15.6%) trials. Fewer than 10% of trials clearly stated the random allocation sequence generation methods. and only 2.4% mentioned allocation concealment. The vast majority (86.8%) of trials were open-label, while only 13.2% used blinding. Almost half (45.5%) administered the CHM intervention as a tea or decoction. Only one trial (0.6%) reported a sample size calculation, and a single trial (0.6%) discussed quality control of the CHM intervention. Conclusion: The overall methodologic quality of RCTs in CHM was poor. It is essential to improve the design of future RCTs in this clinical area. Recommendations: (1) Investigator conducting RCTs should have formal training about clinical trial design; (2) A flow chart is recommended to ensure that all essential steps of clinical trial design are included; (3) Conducting pilot studies prior to RCTs may help improve their design; (4) Registration of clinical trials and publishing their protocols prior to enrolment may reduce publication bias and solicit peer reviews of the proposed design; (5) Collaboration between CHM investigators and traditional medicine academic research centers interested in integrative medicine may lead to quality improvement of RCTs of CHM.

KEY WORDS randomized controlled trial; Chinese herbal medicine; methodology; quality assessment


1 INTRODUCTION

Chinese herbal medicine (CHM) is becoming increasingly popular in industrialized nations as one form of" alternative" or" complementary" medicine. In order to bring it fully into the conventional medical systems of the world—and thereby utilize its considerable benefits—evidence of the safety and efficacy of herbs and herbal products are necessary. The question then is" Where can one find such evidence?" Both randomized controlled trials (RCTs) and systematic reviews are commonly thought to provide the strongest level of evidence regarding treatment efficacy of competing therapeutic interventions. The credibility of the evidence to support a treatment approach such as CHM therefore depends on the quality of RCTs. Previous reports have showed that systematic reviews in complementary medicine involving CHM lack high quality RCTs to provide clear evidence of efficacy. In fact, the quality of RCTs with intervention of CHM as a modality has been a topic of discussion for long time, though no evidence was available to make a definitive judgment on the topic.

If, as some may suspect, the quality of RCTs regarding CHM is not in fact satisfactory, a discussion must then follow regarding recommendations for improving this situation. We have therefore proposed a four-part series of articles which focus on the four basic elements of RCTs in CHM: (1) clinical trial design and methodology, (2) control group design, (3) quality control of CHM used in RCTs and (4) reporting format of RCTs in CHM. Well-designed RCTs were once widely recognized as providing the strongest evidence of the effectiveness of health care interventions. With the development of systematic review and usage of meta-analytical techniques, systematic reviews of RCTs are now thought to provide the best level of evidence about the effectiveness of an intervention. However, such claims are founded on the assumption that RCTs are of sufficient quality, especially with respect to the clinical design and methodology, and that the details are reported clearly. In fact, the methodology of RCTs determines how its results can be interpreted and the extent to which its results can be trusted. Previous studies have shown that erroneous conclusions can be drawn based on misinterpretations of a study’s design and limitations. Any inadequate methodological approaches, such as patient selection, sample size calculation, randomization procedure, outcome assessment, handling of dropouts and follow-ups, will threaten the validity of a clinical trial with potentially exaggerated treatment effects. In order to assess the quality of the clinical design and methodology in studies of CHM, the goal of this study was to review the quality of relevant RCTs in CHM, and to provide recommendations for improving them in the future.

2 MATERIALS AND METHODS

2.1 Identification of randomized trials

We identified RCTs related to CHM from the
Cochrane Library database of systematic reviews in July 2005. The search strategy used was as follows.

Step 1: “Chinese herbal medicine” was used as a search term, which yielded 25 systematic reviews.

Step 2: Search records were reviewed for their relevance to CHM. There were 11 reviews relevant to CHM that were included (Table 1), and 14 reviews that were not related to CHM and thus excluded (Table 2).

Step 3: Primary studies of RCTs of CHM from the 11 reviews were tabulated and cross-referenced to eliminate duplicates. This produced a list of 167 RCTs and 2 quasi-RCTs; the latter were excluded due to their design. This list of 167 RCTs was the basis for further quality assessment.

2.2. Assessment of methodology quality

Methodologic quality was defined as confidence that the trial’s design, conduct, analysis, and presentation minimized or avoided biases in the trial’s findings. In this paper, we adopted a nine-item checklist related to the methodology of RCT from the revised consolidated standards of reporting trials (CONSORT) statement, and added two items especially for CHM involving preparation form of herbs and quality control of CHM. Thus, the checklist was composed of eleven items (Table 3).

<table>
<thead>
<tr>
<th>Item</th>
<th>Included Cochrane Library systematic reviews about Chinese herbal medicine (year of publication)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>Liu JP, Manheimer E, Tsutani K, Gluud C. Medicinal herbs for hepatitis C virus infection. (2001)</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Item</th>
<th>Excluded Cochrane Library systematic reviews about Chinese herbal medicine (year of publication)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>Proctor ML, Murphy PA. Herbal and dietary therapies for primary and secondary dysmenorrhea. (2001)</td>
</tr>
<tr>
<td>14</td>
<td>Little CV, Parsons T. Herbal therapy for treating rheumatoid arthritis. (2000)</td>
</tr>
</tbody>
</table>

Data extraction and analysis

Two observers (Zhao-Xiang BIAN, Andrew K. L. KWAN) assessed the quality of RCT methodology according to the checklist described above. All disagreements due to inaccurate data extraction were resolved through further verification of the original articles. In all cases, consensus between the two observers was achieved prior to the analysis. The data from two reviewers were entered into an Excel file for analysis.

RESULTS

We identified 167 RCTs related to the use of CHM for 11 different conditions. Table 4 reports the basic characteristics of these RCTs, including their indication, participants (total number, sample size range and average size), number of CHM interventions tested, as well as the journal and year of publication. A total of 18,058 participants were studied in 167 RCTs, with individual trials enrolling between 20 and 374 (median 107) of them. Eighteen trials (10.8%)—not conducted in Mainland China—were published in English journals, while 149 RCTs (89.2%)—conducted in Mainland China—were published in Chinese journals not listed in the 2004 edition of Science Citation Index & Journal Citation Reports. Table 5 reports the 11-item checklist mentioned in the 167 RCTs. Overall, 36.0% (90/250) of the items on the checklist for all studies were reported in the articles. Items 2, 3, 10 in the checklists were fully met, while some (items 1, 4, 5, 6, 7, 9, 11) were partially met, and others were completely omitted (item 8). All 167 trials (100.0%) had an explanation about interventions, general objectives, primary outcome design and preparation forms of herbs used. Although 163 (97.6%) trials reported inclusion criteria, exclusion criteria were only mentioned in 26 (15.6%) trials. Only four trials (2.4%) chose the secondary outcome for evaluation of intervention. One trial reported a priori estimation of sample size. Only 8.4% (14/167) of trials clearly stated sequence generation methods. Only 4 trials (2.4%) mentioned randomization concealment. Only 22 trials (13.2%) implemented blinding in the trials, while more than 86% were open trials. For those single blind and double blind trials, not all trials gave clear explanations about the blinding method. Only one trial mentioned quality control of the CHM intervention, and this topic will be discussed further in Part III of this four-part series of articles on RCTs in CHM.
4 DISCUSSION

Our review discovered that many RCTs of CHM are of poor methodological quality, much like early RCTs from other disciplines in integrative medicine, such as homeopathy and acupuncture (9). Despite this overall weakness, most RCTs did adequately report the specific objectives of their study, as well as the interventions, primary outcomes and preparation forms of the herbs used. Other aspects of methodology, such as selection of participants, sample size calculation, randomization, allocation concealment and blinding, were mostly ignored. A discussion of the importance of these characteristics is as follows.

4.1 Selection of participants

The selection of participants serves the purpose of ensuring that all participants are equal in terms of diseases, ensuring that the findings in the RCTs
accurately represent what is for the interest of the population, and helping medical practitioners decide whether RCT results can be generalized to the patients that present to their office. Absent strict inclusion exclusion criteria, the reader may not know whether the trial results can reflect the true effect of interventions in the target population, and whether the trial results can be generalized to their own practice. It is therefore essential that every RCT should be clearly defined how subjects were selected for enrolment. The inclusion criteria aim to define the main characteristics of the target population, while exclusion criteria weed out subsets of the target population that might undermine the validity of the results. Common reasons for exclusion are as follows: a high likelihood of being lost to follow-up, an inability to provide good data, being at high risk of side effects, or characteristics that make it unethical to withhold treatment. Failure to specify exclusion criteria may threaten the internal consistency of the study. For example, one trial aimed to test an herbal mixture on type 2 diabetes used World Health Organization (WHO) criteria for inclusion, but it did not specify any exclusion criteria. Patients meeting inclusion criteria but presenting with various co-morbidities would therefore be enrolled, despite baseline differences in their health status that may affect anticipated response to treatment. On the other hand, overly restrictive exclusion criteria will jeopardize enrolment and limit generalization of study findings to other patient populations. Specifying inclusion exclusion criteria are also important for systematic reviews to decide whether groups from separate studies can be combined for meta-analysis.

4.2 Sample size calculation

Of the 167 RCTs reviewed, only 1 (0.6%) conducted a priori sample size calculation. An increasing number of health-related journals are requiring RCTs to describe this process in accordance with the CONSORT statement. This provides readers with the details used to perform the calculation, and helps to ensure that trials will be large enough so that the investigators can observe the differences they are interested in detecting. If a sample size is too small, then the research might not detect the presence of true difference between two different interventions (i.e., type II error or false negative), and the study might be a waste of resources and potentially unethical. On the other hand, a sample size must not be excessively large so as to waste resources, especially time spent in recruiting and screening potential participants, nor unnecessarily expose too many participants to an experimental intervention.

There are seven basic elements required to conduct a sample size calculation for a RCT in which efficacy is compared on a numeric outcome between control and treatment groups: (1) design the study to meet the specific aims; (2) set acceptable limits for type I (α) (i.e., possibility of falsely declaring a positive effect for the treatment when there is none, often set at 0.05) and type II (β) (i.e., possibility of falsely declaring no effect for the treatment when there is one, often set at 0.2); (3) determine the minimum clinically meaningful difference and variance between groups from prior literature or expert opinion (e.g., treatment will result in a 20% improvement compared to 0% for treatment); (4) select whether the test will be one-sided (i.e., is the treatment better than the control) or two-sided (i.e., is the treatment better or worse than the control); (5) estimate the retention rate at the final follow-up point and adjust sample size accordingly (e.g., with a 50% drop-out rate, the sample size must be doubled); (6) calculate the required number of study participants in each group; (7) revise study parameters as required.

Slight variations of α and β, and expected effect and variance in both groups will affect the final sample size calculation and therefore trial costs. For example, reducing α, increasing power, and reducing the expected difference between groups will increase the sample size. The appropriate effect size varies widely between studies, since it should represent the smallest effect that would be regarded as clinically meaningful and important. This last point must be emphasized since studies that enroll many participants may report statistically significant differences that are too small to be of interest to clinicians (e.g., a 5% reduction in pain). Reporting these differences between the treatment and control groups may give a false impression that an intervention is beneficial.

Other factors to consider for this step include disease population, treatment cost, funding for trials, etc. Therefore, estimation of the effect size of interest should reflect both clinical acumen and the potential public-health effect. Clinical investigators must participate in this process since statisticians are not likely to generate useful sample calculations without the required data. The few RCTs that reported calculating the sample size did not give the method and details used to perform this calculation.
severely limiting its usefulness. It is necessary that sample size should be calculated before the clinical trial begins. and calculation methods should be reported clearly.

4.3 Randomization

Randomization is a powerful tool to assign a target sample to treatment groups balanced on all possible known and unknown confounders, thus helping to avoid systematic bias. The quality of results from RCTs is highly dependent on the quality of randomization. The important factor in the process of randomization is the allocation sequence generation; conversely, sequence generation methods are used to judge the quality of randomization. Unfortunately, our data showed that just 8.4% (14/167) trials provided clear statements about sequence generation, while 91.6% (153/167) merely stated that participants were randomized without explanation as to how the allocation sequences were generated. For those 18 trials published in English journals, only 5 trials stated clearly about the sequence generation methods. This situation is similar to other medical specialties [16,17]. For example, in periodontology, although 91% of trials were described as randomized, adequate methods for randomization were only reported in 17% [16]. Nearly 50% of orthodontics trials published in American Journal of Orthodontics and Dentofacial Orthopedics, British Journal of Orthodontics and European Journal of Orthodontics from 1989 to 1998 did not mention the method used to generate the randomization sequence [19]. This deficiency is therefore not limited to RCTs reporting on CHM. Absence of this explanation will prevent readers from judging whether the methods were proper. The key issue for the randomization is that participants are allocated to different arms randomly, that is, by chance and not by choice (involving investigator's choice and participant's choice), either through the simple (unrestricted) randomization, restricted randomization or stratified randomization [18,19]. The sequence generation method of randomization will help determine the scientific accuracy and credibility of RCTs. Therefore, strictly following the necessary steps for randomization, as well as clearly reporting these methods, are important for RCTs.

4.4 Allocation concealment

Undoubtedly, randomization is a powerful method to produce balanced treatment groups, while allocation concealment is the critical factor to ensure the success of randomization. Random sequence generation merely means that the group assignment sequence of a participant was generated randomly, but does not ensure that the allocation sequence was consistently followed during implementation. In order to allocate the participants to balanced treatment groups, the sequence should be concealed to investigators, participants, and all other study personnel. If not, selection bias may be introduced, whereby the treatment assignment is no longer truly random and an imbalance in prognostic factors may occur between treatment groups. Previous studies have shown that inadequate or unclear allocation concealment can exaggerate clinical effects up to 40%, especially in poorly conducted trials [6,7,20], and it can also cause greater heterogeneity in results [21]. Thus proper randomization should involve both random sequence generation and complete implementation of that sequence to minimize bias.

Our study reported that although 167 trials claimed to be randomized, only 2.4% (4/167) reported the methods for allocation concealment, compared with 17% trials reporting allocation concealment in periodontology [16]. The reason why the rate of randomization concealment (and or the reporting of it) is so low is unknown, though some researchers believe that with blinding—and especially double blinding—allocation concealment is not necessary. But blinding and allocation concealment are different [22]. Blinding concentrates on preventing study participants and personnel from determining the group to which participants have been assigned (which leads to ascertainment bias); it safeguards the sequence after allocation. In contrast, allocation concealment concentrates on preventing selection and confounding biases; it safeguards the assignment sequence before and until allocation. Some researchers may not understand the two concepts. Whatever the reason, RCT investigators need to understand these two principles and the rationales for using them. Commonly used methods to conceal allocation include calling a central, coordinating office for each patient assignment at the time that the patient presents for study inclusion; using sequentially numbered, opaque sealed envelopes; and using numbered bottles or containers [21].

4.5 Blinding

Why is blinding necessary for RCTs? Briefly, the reason is related with the aim of clinical trial: to find out the objective efficacy of the target intervention. In order to maintain this objectivity, we should make sure that the results were not diluted or misled by the subjective preferences (bias) from the
participants, investigators, or assessors. If not, a clinical trial cannot generate accurate results and thus cannot truly advance our knowledge of health care. From this very important reason, blinding is the gold standard for clinical trial design and should be carried out even for those trials with objective indexes for assessment. Otherwise it will damage the validity of the results they generate.

Previous reports have shown that trials that were not double blinded yielded larger estimates of treatment effect than trials in which authors reported double blinding (odds ratios exaggerated, on average, by 17%) . Therefore it is necessary to execute the blinding test. In fact, the importance of blinding was recognized for a long time, and many researchers worldwide understand blinding terminology, but application of proper blinding procedures in RCTs must be improved. Our data showed that only 13.2% (22/167) of trials were described as blinded, while 86.8% (145/167) were open trials in which the patients or doctors (or both) were aware of the assignment to treatment or control group. Of these 22 blinded trials, 13 claimed to be double blinded but failed to provide details about how this was implemented. In the 9 single blinded trials, 5 provided a clear explanation of blinding methodology. This observation was similar to that in trials of conventional medicine, where only 10.4% of 173 RCTs published from 1985 to 2000 in the Journal of Intensive Care Medicine described the blinding procedure. In order to change this situation, more attentions should be given to blinding in clinical trials. This is especially true of trials in CHM assessment where efficacy is mainly subjective.

Another challenge in CHM trials is the preparation of the herbs used as an intervention, which may be administered as a tea, tablet, capsule, or decoction (extract of a crude drug made by boiling or simmering herbs in water, usually has stronger effect than a tea or infusion). Due to characteristic odors, flavors, and colors, these may easily be identified by participants and/or clinicians. Ideally, interventions for different treatment groups should be given in the same form with the same route of administration, as is common in conventional medicine drug trials. A double dummy trial design could be used to overcome this challenge, whereby each would receive two preparation forms (e.g., tea and capsule), only one of which would contain the active treatment while the other would contain placebo; this topic will be discussed further in Part II of this four-part series of articles on RCTs in CHM.

In summary, blinding should be strengthened in RCTs of CHM. In the process of blinding, investigators should decide who should be blinded, who will be in charge of the process, and what preparation form of the intervention should be taken to help the blinding.

5 LIMITATION

The main limitation of this study is that results are applicable only to those RCTs identified from the Cochrane Library for quality assessment. Though these studies cannot represent all RCTs in CHM, they are believed to form a representative sample.

6 CONCLUSION

In general, the quality of clinical trial design and methodology of RCTs with CHM is low. More attentions should be given to the design and methodology of these trials. Specifically, items such as selection of participants, randomization sequence generation, allocation concealment, blinding, sample size calculation, and compliance of participants and investigators should be implemented in all RCTs to improve their overall quality and ensure the validity and usefulness of their results.

7 RECOMMENDATIONS

Following our review of the quality of RCTs in CHM, we have developed some recommendations for improving their methodology.

1. Investigators who want to execute clinical trials in CHM should thoroughly understand the concepts and procedures involved in clinical trials through formal training, including basic concepts of clinical trials as well as tools to deal with problems during their conduct and implementation.

2. Using a flow chart of the clinical trial design is helpful to ensure all essential steps are being included. Based on the CONSORT list, we proposed one flow chart for methodology design in Figure 1.

3. Conducting a pilot study prior to a RCT is helpful for testing the proposed design prior to devoting considerable resources to a RCT, and will likely result in methods being modified prior to undertaking a larger study. Feasibility of trial design cannot be completely ascertained on paper.

4. Registering clinical trials and publishing the protocol will help to improve the quality of RCTs. In 2004, a group of editors from leading health journals worked together to discuss about the necessity of
clinical trial registration. Subsequently, a paper published in JAMA reported the importance of registration. Member journals of the International Committee of Medical Journal Editors (ICMJE) will require registration in a public trials registry prior to patient enrollment as a condition of consideration for publication; this policy is set for implementation on July 1, 2005[27].

(5) Collaboration with traditional academic research centers interested in integrative medicine is encouraged by investigators interested in conducting RCTs in CHM. They may benefit from consultation and partnership with experienced researchers such as methodologists, epidemiologists, biostatisticians and others from academic research centers who are interested in integrative medicine. This may also provide an avenue for training investigators in modern trial methodology.

![Flow chart of design of clinical trial](image)

**Figure 1** Flow chart of design of clinical trial

**REFERENCES**


4. Linde K, Jonas WB, Melchart D, et al. The methodological quality of randomized controlled trials of