Critical appraisal of clinical studies in Chinese herbal medicine

Simon DAGEN A I S \(^{1,2,3}\), Andrea C. TRICCO \(^{1,4}\), Zhao-Xiang BIAN \(^{5}\), Wen-Hua HUANG \(^{6}\), David MOHER \(^{1,2,3}\)

(1. Chalmers Research Group, Children’s Hospital of Eastern Ontario Research Institute, Ottawa, Canada; 2. Department of Pediatrics, Faculty of Medicine, University of Ottawa, Ottawa, Canada; 3. Department of Epidemiology and Community Medicine, Faculty of Medicine, University of Ottawa, Ottawa, Canada; 4. Institute of Population Health, University of Ottawa, Ottawa, Canada; 5. School of Chinese Medicine, Hong Kong Baptist University, Hong Kong SAR, China; 6. Department of Traditional Chinese Medicine, Shanghai Hospital, Second Military Medical University, Shanghai 200433, China)

ABSTRACT The use of complementary and alternative medicine (CAM) is currently widespread and appears to be growing. As an increasing proportion of the population turns to CAM therapies, whether singly or in combination with allopathic medicine, the need for quality research in this area is reinforced. Much of this research consists of clinical studies aimed primarily at clinicians, yet challenges arising from poor methodological quality will occur when interpreting study findings and their implications. For clinicians to be effective consumers of the scientific literature, familiarization with the principles of evidence-based medicine (EBM) is essential. The goal of this review is to introduce clinicians to the concept of critical appraisal of clinical studies and foster critical thinking when reading research articles in order to best evaluate and incorporate study findings into their daily practice. Topics discussed in this article include: (1) fundamentals of EBM; (2) types of clinical studies; (3) hierarchy of evidence; (4) Consolidated Standard of Randomized Trials (CONSORT) statement to evaluate the quality of reporting in randomized controlled trials (RCTs);

Correspondence to: Simon DAGEN A I S, DC, PhD. Chalmers Research Group, Children’s Hospital of Eastern Ontario Research Institute, 401 Smyth Road, Ottawa ON K1H 8L1, Canada. E-mail: SDagenais@cheo.on.ca
(5) methodologic quality rating scales for RCTs; and (6) issues specific to evaluating studies of Chinese herbal medicine.

KEY WORDS evidence-based medicine; research design; randomized controlled trial; Chinese herbal medicine, CONSORT statement


1 INTRODUCTION

The use of complementary and alternative medicine (CAM) in the Western world is currently widespread and appears to have grown in the past decade[1,2]. As an increasing proportion of the population turns to CAM, whether singly or in combination with allopathic medicine, there is a greater need for quality research in this area. Academic CAM practitioners and other researchers have responded to this demand by increasing the number of clinical studies and journals related to CAM.

As the complexity and quantity of the health literature related to CAM evolves, it may prove challenging for clinicians — broadly defined as healthcare professionals directly responsible for patient care, regardless of specific training or discipline — to fully appreciate and use this knowledge without the benefit of additional training in this area. Empirical evidence suggests that results of clinical studies may be biased when their methodology is poor or poorly reported[3,4]. If clinicians are unable to critically appraise clinical studies, they may unnecessarily alter their practice and potentially jeopardize patient care based on weak, and possibly false, evidence.

In order for clinicians to be effective and educated end-users of clinical research literature, it is important for them to be familiar with the fundamentals of evidence-based medicine (EBM). This knowledge will help them to actively question and engage in critical appraisal of clinical studies to better assess the quality of what is being read. Through this process, clinicians will thereby have a greater appreciation of the context in which study findings should be interpreted based on strengths and limitations.

The primary objective of this article is to provide an overview of various aspects of EBM as they pertain to helping clinicians to critically appraise clinical research studies. A secondary objective is to discuss special requirements for critical appraisal of clinical studies involving Chinese herbal medicine (CHM).

2 EVIDENCE-BASED MEDICINE

Evidence-based medicine (EBM) is a principle to which clinicians have long aspired, but which was only relatively recently concretized and defined as, "The conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients. The practice of evidence-based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research"[5].

Historically, clinicians have attempted to integrate information from their formal education, with experience gained in practice, and findings from new research studies to provide the best care to their patients. These sources of knowledge all contribute to what is formally termed "evidence", and serve to inform and improve patient care. Although clinicians may believe that EBM is overly complex, at its core EBM is simply an attempt to learn from the past and apply what we already know to improve patient care. With this in mind, EBM is accessible to everyone involved in healthcare and has recently burgeoned as clinicians, researchers, policy makers, and patients embrace this concept as an important tool in decision-making. An understanding of the basic principles of EBM is crucial in order to critically appraise clinical studies.

3 TYPES OF CLINICAL STUDIES

Health research is a broad field and there are several different types of studies being conducted and reported in journals and elsewhere. Clinical studies (also called clinical trials, though this is sometimes used to refer only to randomized controlled trials — RCTs) may be broadly defined as research to assess the efficacy and/or safety of a treatment in human participants[6]. Within the field of clinical studies are several common study designs. Identifying the type of clinical study being reported in an article is the first component of critical appraisal. Some of the most commonly conducted and reported clinical study designs are; (1) descriptive (case report/case series);
(2) cohort (retrospective/prospective); (3) controlled (non-randomized, quasi-randomized, and randomized).

Basic features of these study designs are summarized in Table 1, which will help readers recognize and evaluate reports of these study designs in the literature. Determining the type of clinical study being reported can be challenging, and the study design reported by authors may not always be adequately described within their research methodology. Furthermore, designs may overlap and some design aspects of studies may share properties with other study designs. For example, a study presenting itself as a prospective cohort that fails to describe the study population, inclusion/exclusion criteria, and study timeline, could therefore be interpreted simply as a case series. The final decision as to the type of study being reported in a journal article may therefore be based on overall impressions rather than the presence or absence of specific criteria common to each study design. A general rule to consider is that there should be little ambiguity for a reader to determine the study design in a well conducted and reported clinical study. Furthermore, it is important to note that the study design should be described in a reproducible manner so that another investigator could apply the same methods and obtain similar results if applied to an equivalent population.

4 STRENGTHS AND LIMITATIONS OF STUDY DESIGNS

There are inherent strengths and weaknesses to each study design. For example, a case report provides detailed information on a single patient that may be representative of a new condition and/or treatment approach. When several patients are involved this is considered a case series. It is difficult to generalize results to other patients based on such limited observations, but these studies can alert clinicians to new disease entities or methods of treatment. Cohort studies build on case reports/series by defining a broader patient population to increase the generalizability of their results. They are an excellent starting point for assessing study methods to be used in future studies, gathering preliminary evidence of efficacy to justify further studies, and providing data that can be used for determining sample size in RCTs. Although the focus here is on clinical studies, cohort studies are also used in epidemiological studies to observe the potential health effects of exposure to various substances or phenomena over time (e.g. lung cancer and asbestos). However, when examining a particular intervention, it is difficult to attribute the results of a cohort study without randomization, blinding, and a control group. Controlled trials are often cited as the gold standard for determining the efficacy of an intervention, but methods developed for pharmaceutical intervention studies may not always be suitable for trials of CAM interventions. For example, the use of Western medicine diagnoses and standardized treatment regimens for all participants may not reflect CAM practice. This is discussed further below.

5 HIERARCHY OF EVIDENCE

A single clinical study can rarely be thought of as definitive and must be taken in the broader context of all available evidence for a particular intervention. Gathering research into the efficacy of an intervention typically requires multiple studies conducted over several years by different investigators and study sites. As each clinical study is published, others may be conducted by using slightly different study designs or methods to answer different research questions. When multiple studies have been conducted on the same topic, they, as a whole, constitute the body of evidence for that intervention.

5.1 Bias and confounding

Within that body of healthcare literature is a hierarchy of evidence that has been established based on the strengths and limitations inherent to specific study designs as mentioned above (see Figure 1)\(^{(1)}\). This hierarchy of evidence also indicates a study design’s susceptibility to bias and confounding, which are two important concepts in EBM. Bias is the distortion of study data, whereas confounding is the inability to distinguish the separate impacts of multiple variables on a single outcome\(^{(1)}\).

Examples of bias include selection bias (i.e. participants not representative of population), measurement bias (i.e. outcome measure not accurate), and responder bias (i.e. participants completing study differ from others). An example of confounding could be a clinical study of acupuncture for back pain whereby participants in both the intervention and control groups report improvement. However, if both groups had to first lie down on a comfortable table for 30 minutes in a quiet room, the improvement in pain could be attributed to
reducing stress rather than the intervention. Clinicians reading clinical studies should therefore ask themselves whether a study’s results could have been influenced by bias and confounding as part of the critical appraisal process.

In addition to the 3 common primary (i.e. patient level data) clinical study designs described in Table 1, other non-experimental study designs, and secondary (i.e. study level data) study designs are ranked within this hierarchy of evidence and are briefly described below.

<table>
<thead>
<tr>
<th>Study design</th>
<th>Descriptive</th>
<th>Cohort</th>
<th>Controlled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subtypes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case report</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- 1 patient</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case series</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- several patients</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Purpose
- Provide extensive information on case(s) selected due to unusual nature of intervention or disease
- Question current thinking and share new clinical observations
- Assess the efficacy of a treatment compared to a control group in a defined population using rigorous scientific methods
- Confirm or refute findings from other clinical study designs

Typical article format
- Presenting complaint
- Examination findings
- Description of treatment
- Results
- Discussion
- Conclusion

Advantages
- Extensive information provided on single (few) case(s)
- More reflective of day-to-day practice focused on individual patients
- May foster new theories for further investigation
- More standardized approach than case series
- Good starting point for clinical research involving new intervention
- Relatively inexpensive and simple to conduct
- Control group minimizes non-treatment effects observed
- Defined inclusion/exclusion criteria help generalizability
- Randomization ensures baseline comparability

Disadvantages
- Few participants
- Non-experimental
- Higher likelihood and more sources of bias
- May not be generalizable
- Patient selection criteria often unclear
- Patients not blinded to treatment
- No control group to compare new treatment to standard of care or natural history
- Often requires numerous participants with long recruitment and screening period
- Randomization and blinding difficult to implement in certain clinical settings
- Requires extensive support from research team and infrastructure

Cost/complexity
- Low
- Medium
- High

Approximate number of participants
- 1-10
- 100 or more
- 50 to 100 or more
5.2 Non-experimental studies
Expert opinion is often written as an editorial by those with strong opinions on a topic, whether they are proponents or opponents of an intervention. These studies are non-experimental, rarely report data, have no formal study design, often selectively cite references to confirm their viewpoints, and are subject to many forms of bias and confounding. Opinionated reviews attempt to summarize literature without following a systematic search approach, and are thus subject to bias[11]. These are the weakest forms of evidence and should rarely be used to influence practice without additional evidence.

5.3 Secondary studies
Systematic reviews conduct a thorough search of the health literature focused on a particular intervention and condition, and describe the search strategy in detail to allow for independent replication[12,13]. Specific inclusion and exclusion criteria are applied to the search results and relevant studies are often assessed for methodologic quality. Results of included studies are summarized to provide an overview of their findings. Meta-analyses are based on systematic reviews and perform quantitative statistical pooling of results from included studies when they are sufficiently homogeneous (alike). Combining participants from several studies increases statistical power to detect differences between the intervention and the control groups that may not be apparent from individual studies.

Clinical guidelines typically begin with a systematic review and/or meta-analysis, whose results are then reviewed, discussed, and evaluated by multiple stakeholders, including clinical, research, and policy experts, and occasionally patient representatives. The purpose of clinical guidelines is to translate research findings into daily practice by making specific recommendations based on the best evidence available. Guidelines are often widely distributed and form the basis for the accepted standard of care. Readers interested in evaluating the methodological quality of guidelines are referred to the Appraisal of Guidelines for Research and Evaluation (AGREE) instrument, a well-respected international initiative[14]. Health technology assessments (HTAs) are similar to clinical guidelines, with additional components such as economic evaluations and societal implications for adopting or rejecting a particular intervention/technology for use within a particular jurisdiction or organization.

6. ASSESSING THE RISK OF BIAS

6.1 Conflicting evidence
Clinical studies may present discordant results, even when examining a similar intervention, indication, and population[15]. When readers of the health literature discover such discrepancies, they should first identify the study design to place it within the hierarchy of evidence. The reader should then evaluate the study's validity and methodological quality. Since RCTs are one of the most common and important clinical study designs, methods for critically appraising RCTs are described below.

6.2 Randomized controlled trial
Clinicians will encounter numerous RCTs as they read scientific health journals. Although such designs account for only about 2% of the total healthcare literature found in Medline, the number of published RCTs is growing at 11% annually[14]. Results from RCTs are often touted by proponents or opponents of a particular intervention as "definitive proof" of efficacy, or lack thereof, of an intervention. However, the discerning reader will first appraise the methodology of a particular study before considering its implications. This is a crucial step since empirical evidence suggests that bias exists in RCTs with poor methodological quality and/or RCTs that do not fully report their methods[16,17]. This latter point is important since, in most cases, the only way for readers to judge how the RCT was conducted is by examining its report. Surveys of the current literature indicate that although there is definite improvement in how RCTs are reported, many of them are still not optimally reported[17].

This situation was addressed in the mid 1990s by a group of scientists, including editors, who established standards for the reporting of RCTs. Their efforts resulted in the Consolidated Standards of
Reporting Trials (CONSORT) Statement, an international collaborative effort led by one of the authors of this paper[18]. The CONSORT Statement provides clinical trialists with tools to improve their reporting practices and also provides a framework for the critical appraisal of the reporting of RCT publications. Journal editors, systematic reviewers, clinicians or anyone wishing to make an evidence-informed decision can use this framework. The two central features of the CONSORT Statement are a checklist of 22 items that should be reported for a RCT (Table 2), and a diagram that should be included to portray the flow of participants through the various stages of a trial (Figure 2). Recent evidence suggests that the quality of reporting of RCTs is better in journals endorsing CONSORT compared to those not doing so[19].

### Table 2 CONSORT checklist for reporting simple 2-group parallel randomized controlled trials (to be continued)

<table>
<thead>
<tr>
<th>Item</th>
<th>Section</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TITLE &amp; ABSTRACT</td>
<td>Clearly state the population used, intervention studied, trial design, and outcome of interest (e.g. a 20-week randomized controlled trial of estradiol replacement therapy for women aged 70 years and older).</td>
</tr>
<tr>
<td>2</td>
<td>INTRODUCTION</td>
<td>Explain past scientific evidence in the field relevant to the trial (i.e. conduct a literature search). The rationale should justify the need to conduct the trial and importance of the trial.</td>
</tr>
<tr>
<td>3</td>
<td>METHODS</td>
<td>Description of the trial methods and how they were administered.</td>
</tr>
<tr>
<td>4</td>
<td>Objectives</td>
<td>report if training was provided to assess depression using the Hamilton Depression Scale in the trial.</td>
</tr>
<tr>
<td>5</td>
<td>Outcomes</td>
<td>Describe the primary and secondary outcomes and how they were measured (e.g. the primary outcome is remission from depression noted by a score of 7 or less on the Hamilton Depression Scale and the secondary outcome includes clinical response defined as a 50% decrease on Hamilton Depression scores from baseline). Report methods used to increase the quality of measurements. For example, report if training was provided to assess depression using the Hamilton Depression Scale in the trial.</td>
</tr>
<tr>
<td>6</td>
<td>Sample size</td>
<td>Report the calculations used to decide the number of participants required for the trial.</td>
</tr>
<tr>
<td>7</td>
<td>Randomization sequence generation</td>
<td>Randomization minimizes bias associated with assigning patients to treatment. Clearly describe the method used to generate random allocation (e.g. Participants had an equal probability of assignment to the groups. The randomization code was developed using a computer random number generator). Describe details of restriction, if any were used (e.g. Random permuted blocks were selected. The block lengths were 4, 8, and 10 varied randomly).</td>
</tr>
<tr>
<td>8</td>
<td>Allocation concealment</td>
<td>Report all methods pertaining to allocation concealment methods. Allocation concealment is the process used to prevent any prior knowledge of group assignment (e.g. capsules were placed in identical-looking, numbered bottles and administered sequentially; sequentially numbered opaque, sealed envelopes were used to conceal treatment allocation). concealment prior to intervention assignment is preferred.</td>
</tr>
<tr>
<td>9</td>
<td>Implementation</td>
<td>Provide details about how the trial was executed. This includes who generated the allocation sequence, who enrolled participants, and who assigned participants to their groups.</td>
</tr>
<tr>
<td>10</td>
<td>Blinding (masking)</td>
<td>Report all blinding processes. Blinding refers to the process in which participants, treatment administrators, and outcome assessors are unaware of the group to which participants are assigned. For example, our study was double-blinded as the participants and the investigators were unaware of which treatment group that the participants were in.</td>
</tr>
<tr>
<td>11</td>
<td>Statistical methods</td>
<td>Report all methods used to statistically analyze trial results. This includes adjustments, subgroup analyses, and any additional analyses.</td>
</tr>
</tbody>
</table>
Table 2 (continuation) CONSORT checklist for reporting simple 2-group parallel randomized controlled trials

<table>
<thead>
<tr>
<th>Item</th>
<th>Section</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>Participant flow</td>
<td>The participant flow is a description of the flow of participants through each stage of the trial. A diagram is strongly recommended. For each treatment group report the number of participants that were randomly assigned, received the intended treatment, completed study protocol, and analyzed for the primary outcome. Describe and explain any differences between the methods outlined in the protocol and the methods that were actually used in the trial.</td>
</tr>
<tr>
<td>14</td>
<td>Recruitment</td>
<td>Report the timing of enrollment and the length of patient follow-up. For example, patients were enrolled on September 1, 2005 and followed up until December 23, 2007.</td>
</tr>
<tr>
<td>15</td>
<td>Baseline data</td>
<td>State the demographics of participants in the trial (e.g., age, gender) and clinical characteristic data collected at the beginning of the study (e.g., blood pressure at the first patient visit with the physician).</td>
</tr>
<tr>
<td>16</td>
<td>Numbers analyzed</td>
<td>State the total number of participants in each group. Describe whether intention-to-treat analysis was done (i.e., all patients who began the trial are used in the analysis, whether or not treatment was received for the full time period of the trial). State results in absolute numbers if feasible and not in percentages.</td>
</tr>
<tr>
<td>17</td>
<td>Outcomes and estimation</td>
<td>Report the results of primary and secondary outcomes by treatment group. State the difference between treatments (i.e., effect size; e.g., relative risk) and the degree of uncertainty surrounding the effect size (i.e., 95% confidence interval).</td>
</tr>
<tr>
<td>18</td>
<td>Ancillary analyses</td>
<td>Describe any additional analyses that were performed (e.g., subgroup analyses, adjusted analyses). Provide information about whether the pre-specified analysis in the trial protocol was followed. Report any additional analyses done for exploratory purposes.</td>
</tr>
<tr>
<td>19</td>
<td>Adverse events</td>
<td>State all undesirable or unwanted effects of the treatment, even if they were not harmful.</td>
</tr>
<tr>
<td>20</td>
<td>Interpretation</td>
<td>Describe the interpretation of the results and take into account any trial hypotheses. Report limitations of the trial such as sources of bias and multiple analyses, which increases your chances of rejecting the null hypothesis when it was in fact true (i.e., you find that a difference between the interventions exist when they were in fact the same. This is otherwise known as a “type I” error).</td>
</tr>
<tr>
<td>21</td>
<td>Generalizability</td>
<td>Generalizability relates to the degree for which the trial results can be inferred to a larger population beyond the trial participants. Another term for generalizability is external validity. Describe the extent of external validity of the trial.</td>
</tr>
<tr>
<td>22</td>
<td>Overall evidence</td>
<td>Interpret and report the results within the context of the current evidence.</td>
</tr>
</tbody>
</table>

7 RATING SCALES

The CONSORT Group is committed to continuous improvement through the renewal and updates of their recommendations. The first set of recommendations was published in 1996 and updated in 2001. It is now endorsed by international editorial groups, including the International Committee of Medical Journal Editors, and used by several hundred biomedical journals. The CONSORT Statement has also led to recommendations on the reporting of cluster trials, equivalence and non-inferiority trials, and harms (e.g., safety and adverse events) in RCTs. Recently, CONSORT has been extended to reporting herbal interventions. Although not intended as an assessment tool, clinicians evaluating RCTs can use the CONSORT statement checklist to assess the number of reported criteria against a standard benchmark. This can help clinicians to determine whether a study report is in fact of high quality.

In addition to the CONSORT reporting checklists for RCTs, researchers have also developed methodological quality assessment scales for clinical studies. Methodological quality is a quantitative or qualitative measure of the internal validity of a research study, and is reflective of the confidence that can be placed in study results. There are many methodological quality scales for clinical studies. These instruments are commonly used in systematic reviews and meta-analyses to assign a quality score to a reviewed study. Those reading clinical studies may also use these rating scales to supplement the information gained by identifying the study design, placing it in the hierarchy of evidence, and evaluating its reporting quality.
Moher et al reported in 1995 that at least 22 scales had been developed to assess the methodological quality of RCTs. These scales vary in their size, complexity, and level of development. In a review published in 2001, the authors discussed 60 scales to discuss the reliability of the resultant assessments. Some scales (e.g., Jadad scale, Chalmers TC scale) were designed to assess the quality of any trial, whereas others were for specific trials (e.g., Beckerman scale, PEDro scale). The number of quality assessment scales continues to increase, with little agreement as to which should be used. Among these different scales, the Jadad scale is the most well-known and has been validated to grade the quality of RCTs. It consists of a 5-point scale focusing on adequate generation of the allocation sequence, double-blinding, and description of follow-up (see Table 3). It may also be supplemented with checklists for other important components (e.g., allocation concealment).

### Table 3 Jadad scale

<table>
<thead>
<tr>
<th>Item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Was the study described as randomized (this includes the use of words such as randomly, random, and randomization)?</td>
</tr>
<tr>
<td>2</td>
<td>Was the study described as double blind?</td>
</tr>
<tr>
<td>3</td>
<td>Was there a description of withdrawals and dropouts?</td>
</tr>
</tbody>
</table>

Scoring the items: Score 1 point for each “yes” or 0 point for each “no”. Give 1 additional point if: (1) For question 1, the method to generate the sequence of randomization was described and it was appropriate (table of random numbers, computer generated, etc.); and/or (2) For question 2, the method of double blinding was described and it was appropriate (identical placebo, active placebo, dummy, etc.). Deduct 1 point if: (1) For question 1, the method to generate the sequence of randomization was described and it was inappropriate (patients were allocated alternately, or according to date of birth, hospital number, etc.); and/or (2) For question 2, the study was described as double blind but the method of blinding was inappropriate (e.g., comparison of tablet vs injection with no double dummy).

Although methodological quality scales are important to critically appraise the quality of RCTs, readers should be aware of certain inherent weaknesses. It is important to note that scoring of RCTs
using methodological scales is based on whether the essential elements of the study were reported in the published journal article, and not whether those elements were actually done appropriately in the trial. Recent evidence suggests that the report is a reasonable representation of how the study was conducted for clinical trials. In this respect, methodological quality goes hand in hand with reporting quality. Also, some of the items included in some of the scales are not directly related to the internal validity of a trial. Moreover, the scales are less likely to be transparent when encountered in a systematic review, and different scales can lead to discordant reviews. For example, a study in which 25 different scales were used to assess 17 trials comparing low molecular weight heparin with standard heparin for thromboprophylaxis, reported that these scales seemed unable to provide more reliable assessments of validity.

Furthermore, most scales for assessing the quality of RCTs have not been developed with sufficient rigor, and derive a summary score by simply adding the number of responses without differential weighting for each item. While this approach offers appealing simplicity, it is not supported by empirical evidence and overlooks the importance of specific elements such as random allocation concealment in RCTs. Future efforts at developing quality scales may be best focused on gathering more empirical evidence to identify trial characteristics that are directly related to bias. The Cochrane Handbook for Systematic Reviews of Interventions also remarks that the scales with multiple items and complex scoring systems often take more time to complete than simple approaches.

It should be noted that methodological quality rating scales are not sufficient to conduct critical appraisal of a clinical study, and should simply be viewed as one of the tools available to help clinicians determine whether the results of a study are valid. Being aware of the limitations of methodological quality rating scales will help clinicians determine their role in critical appraisal, if any.

8 ISSUES SPECIFIC TO HERBAL MEDICINE

The research methods commonly used in clinical studies such as using an inert sugar pill as a placebo or blinding the clinician, outcome assessor and patient were originally developed by researchers to assess the efficacy of oral pharmaceuticals. Although these methods are valid in that context, it is difficult to ascertain whether they can (or should) be used in research involving non-pharmaceutical interventions such as Chinese herbal medicine (CHM), which is based on entirely different foundations and principles. These differences must be considered when appraising clinical studies. Some of the methodological quality issues that are unique to clinical studies involving CHM, for example, include: defining the targeted population and disease of interest; administering a standardized intervention to all participants; and quality control of the intervention.

Each is briefly discussed below to provide clinicians with another perspective when performing critical appraisal of the quality of clinical study reports they read related to CHM.

8.1 Defining the targeted population and disease of interest

Participants in RCTs of CHM are often selected based on the diagnosis of a disease of interest according to Western medicine. However, CHM clinicians often assess their patients differently, viewing them holistically and diagnosing general syndromes (e.g., liver stagnation), which may not fit into specific (Western) disease categories. A clinical study of an herbal remedy typically prescribed for a syndrome according to CHM may overlap certain characteristics of a formal Western diagnosis. However, it is unclear whether this overlap is always sufficient to draw generalizations between syndromes and diagnoses. Another related concern is the clinical heterogeneity inherent to the syndromes assessed by CHM clinicians in individual patients. The issue of participant inclusion/exclusion criteria is therefore crucial when considering the interpretation of results. Subsequent comparisons between different trials as performed in systematic reviews can also be problematic. Clinicians should therefore consider whether the targeted population in a trial reflects their patient population before adopting a new approach based on a published study.

8.2 Administering a standardized intervention to all participants

Clinicians in CHM typically perform individualized assessments resulting in multifactorial syndromes rather than specific diagnoses. The treatments recommended for patients are also therefore highly individualized in CHM. A typical holistic treatment regimen in CHM may encompass herbal preparations containing multiple ingredients, as well as acupuncture, moxibustion, dietary modifications, and lifestyle changes. Each of these
components could then be altered by a clinician to address specific concerns according to a patient’s overall health. This individual approach to health practice is one of the aspects that distinguish a traditional healing system (such as CHM) apart from Western medicine, which often prescribes specific remedies based on explicit diagnoses, regardless of overall health. This distinction raises the issue of whether RCTs in which a standard intervention is given to all participants is an accurate representation of CHM practice. Clinicians should therefore consider whether the restrictions imposed on a clinical study because of its research nature is reflective of their usual approach to a particular patient when interpreting study results.

8.3 Quality control of herbal medicine

This issue is rarely discussed in reports of trial methods yet it is crucial to the validity of the study’s results\(^7\). Unlike Western pharmaceutical products that are standardized by adhering to strict quality control measures for manufacturing and batch release testing such as the Good Manufacturing Practices (GMP), CHM products rarely undergo quality control. The primary reasons for executing quality control of CHM products include consistency and safety. In order to assess the efficacy of a specific product in a clinical study, all participants should be given exactly the same intervention in terms of product identity, purity, dosage, formulation, etc. This should be accomplished by testing all stages of production, from the raw material to the final products, and could also extend to other aspects of production such as the species and subspecies of plant used, the location of growth, method of harvest, transportation, and processing, etc., in order to ensure product consistency.

Products should also be tested to address safety concerns surrounding herbal products, which should not be assumed to be innocuous. Studies examining herbal products have previously reported contaminants such as heavy metals, herbicides, pesticides, microorganisms, mycotoxins, insects, pharmaceuticals, and other undeclared herbal constituents\(^8\). Readers of clinical studies should therefore look for a discussion of quality control when assessing a trial.

8.4 Reporting guidelines for trials with herbal interventions

It should be noted that the original CONSORT statement was recently amended in order to incorporate additional items that should be reported in trials involving herbal interventions\(^9\). Readers are encouraged to become familiar with these new guidelines when submitting relevant manuscripts for publication. The process of developing CONSORT guidelines for trials of Chinese herbal medicine has also been initiated by some of the authors, though publications of these guidelines are not expected until 2007.

9 SUMMARY

Clinical research is an essential component of practicing evidence based medicine. Clinicians who read reports of clinical studies will be better able to critically appraise them with knowledge of clinical epidemiology and clinical trial methodology. Such knowledge will facilitate a more informed user and decision maker regarding the merits of the healthcare literature. These principles include basic study designs and the hierarchy of evidence based on strengths and weaknesses inherent to different study designs. Clinicians can use these basic principles as a starting point to critically appraise research articles using tools such as the CONSORT checklist and various tools developed to assess the validity of RCTs and other study designs. Results from clinical studies pertaining to CHM also need to be evaluated with respect to specific considerations, such as participant inclusion/criteria, standardization of treatment, and quality control of the intervention. As these methods are applied to reading the health literature, clinicians will obtain new insight and greater appreciation for what reports of clinical studies can (and cannot) bring to their daily practice.

REFERENCES


辅酶 Q_{10}

张晶编译，刘建平审校
（1. 上海中医药大学中医循证医学中心，上海 201203；2. 北京中医药大学循证中医药临床研究与评价中心，北京 100029）

辅酶 Q_{10}是经体内合成的类维生素物质，有助于发挥细胞内某些重要酶的功能。尽管目前尚不清楚人体发生疾病时是否会影响其含量，但临床建议将其作为一种必需的补充物广泛地使用。辅酶 Q_{10}虽有抗癌、保护心脏免受化化疗药物毒性损伤，抗疲劳等功效而用于癌症治疗，但目前证据并不支持以上效果，考虑其不良反应及药物间的相互作用，故化疗期间应避免使用辅酶 Q_{10}。

1 干预措施

辅酶 Q_{10}亦称作 CoQ_{10}、维生素 Q_{10}、泛醌 Q_{10}等，"辅酶"是维持酶正常功能必需的分子，"Q"代表化学基团苯醌，"10"是苯醌的一种特殊类型。在肉类、全谷、豆制品及蔬菜中含量丰富，可由甜菜和甘蔗发酵而成。大多数人体组织中存在 CoQ_{10}，肝、心、肾和胰腺中的含量最高而肺中最低，研究人员也指出心脏病和肿瘤患者体内的 CoQ_{10}含量较低。通常将其制成丸剂（胶囊或片剂）口服，也可通过静脉给药，每天用量为 50～400 mg。

在日本其运用较广，且政府批准用于心血管病的治疗，但心律失常或出血的危险。因此供应商不能宣传其医疗功效，而将其界定为营养品。尽管目前尚不清楚人体发生疾病时是否会影响其含量，但临床建议将其作为一种必需的补充物广泛地使用。辅酶 Q_{10}虽有抗癌、保护心脏免受化化疗药物毒性损伤，抗疲劳等功效而用于癌症治疗，但目前证据并不支持以上效果，考虑其不良反应及药物间的相互作用，故化疗期间应避免使用辅酶 Q_{10}。

2 酶 Q_{10}的疗效

2.1 抗癌 动物实验表明，辅酶 Q_{10}的同型物具抗癌功效。丹麦一项 32 例乳腺癌的研究显示，在常规抗癌治疗的同时或治疗后给予包括辅酶 Q_{10}在内的补充剂，随访 18 个月后，所有患者均存活，生活质量有所改善，6 例得到缓解，需用止痛药及体重下降的病例如数有所减少。但由于未设立对照，同时运用传统抗癌疗法、补充剂不止辅酶 Q_{10}，故其抗癌功效的证据不足。

2.2 抑制放化疗毒性 动物实验支持辅酶 Q_{10}可保护心脏免受化疗药物的损害。Takimoto 等对乳腺癌的小样本（n=40）随机对照试验表明，辅酶 Q_{10}可改善心功能，但心律或其他心功能障碍的症状并无显著差别。Kihama 等对白血病的小样本安慰剂双盲对照研究显示试验组与对照组间比较差异无统计学意义。Okuma 等对各种癌症的大样本（n=80）研究发现辅酶 Q_{10}对心脏有很好的保护效果。1986年，Lucarelli 等对 30 例血癌患者临床对照试验及 Iarussi 等进行的临床试验，因没有严格的对照组，因此尚不能说明辅酶 Q_{10}有效。Lastiyh 和 Thibault 等对服用洛伐他汀和辅酶 Q_{10}的各种癌症患者进行了较大样本的对照临床试验，发现辅酶 Q_{10}能缓解化疗引起的心肌毒副作用。Roffé 等按当前证据撰写的综述表明，研究方法学的缺陷及目前阴性的试验结果，使辅酶 Q_{10}的抗癌功效缺乏说服力。

2.3 抗疲劳 至今没有临床研究验证这一特定功效。

3 安全性

已证实的不良反应有：恶心、腹泻、皮疹、眩晕、光敏反应，过敏，疲倦、头痛、烧心，夜间用量大于 100 mg 时，会轻微失眠。怀孕和哺乳期女性禁用，糖尿病及高血压患者在用药期间不能过量运动。

运用辅酶 Q_{10}可能会增加化疗的效果。因会增加凝血或出血的危险，故可能会增加华法林的疗效，但辅酶 Q_{10}与抗凝血药联合运用则不受影响，同时服用降胆固醇药（他汀类）和降血糖的药物时会分解天然水平的辅酶 Q_{10}。

4 结论

缺乏支持辅酶 Q_{10}抗癌、保护心脏免受化疗药毒性损害、抗疲劳的证据，因此尚不宜作为常规化疗的辅助药物。目前尚无充分证据支持辅酶 Q_{10}对临床实践的影响，肿瘤放化疗期间应避免使用辅酶 Q_{10}。

（声明：因受字数限制，本译文对原文有所删节，省略了参考文献。）