

Statistical Considerations for Traditional Chinese Medicine Clinical Trials

Author's Details: ⁽¹⁾ Shein-Chung Chow-

Duke University School of Medicine, Durham, North Carolina, USA ⁽²⁾ Fuyu Song-

Peking University Clinical Research Institute, Peking University Health Science Center, Beijing, China

Abstract

In recent years, the use of complementary and alternative medicine including botanical drug product and traditional Chinese (herbal) medicine (TCM) in humans for treating critical and/or life-threatening diseases has received much attention. In pharmaceutical/clinical development of a given TCM, one of the major criticisms is lack of objectively scientific evidence (documents) of clinical safety and efficacy. Unlike the Western medicines (WM), TCM often consists of multiple components (active ingredients) whose pharmacological activities are often unknown or are not fully characterized or understood. Thus, standard methods for WM clinical trials may not be appropriately applied directly to TCM clinical trials. In this article, some statistical considerations including selection of study design, preparation of matching placebo, development of study endpoint, validation of an instrument, calibration of the validated instrument, and power calculation for sample size estimation are discussed. These considerations have an impact on effectively and scientifically evaluation of clinical safety and efficacy of TCM in clinical trials. In addition, some practical issues regarding test for consistency in raw materials, stability of drug substance, and animal studies are also discussed.

Key Words: Herbal medicine; Botanical drug products; Matching placebo; Calibration; QOL-like instrument.

INTRODUCTION

In recent years, the search for complementary and alternative medicine such as botanical drug product and traditional Chinese (herbal) medicine (TCM) for treating critical and/or life-threatening diseases has received much attention. This has led to the study of the potential use of promising TCMs. As indicated by Chow et al. (2006) [1] and Chow (2015) [2], TCM originated in ancient China has evolved over several thousands of years, which usually refers to a broad range of Chinese medicine practice including various forms of herbal medicine, acupuncture, massage (*Tui-Na*), exercise (*Qi-Gong*), and dietary therapy. In pharmaceutical/clinical development of a test treatment, one of the major criticisms for the development of TCM is lack of objectively scientific evidence (documents) of clinical safety and efficacy. Unlike the Western medicines (WM), TCM often consists of multiple components (active ingredients) whose pharmacological activities are often unknown or are not fully characterized or understood. Thus, standard methods for evaluation of WM clinical trials (see, e.g., [1-3]) may not be appropriately applied directly to TCM clinical trials.

In TCM clinical trials, it is a concern whether a TCM can be scientifically evaluated the Western way due to some fundamental differences between a WM and a TCM. These fundamental differences include differences in formulation and drug administration, medical theory/practice, diagnostic procedure, and criteria for evaluation [1]. As an example, the Chinese diagnostic procedure for patients with certain diseases consists of four major techniques, namely, inspection, auscultation and olfaction, interrogation, and pulse taking and

palpation (see also, [2]). Under these differences, it is of interest to the investigators regarding how to design and conduct a scientifically valid (i.e., an adequate and well-controlled) clinical trial for evaluation of the clinical safety and efficacy of the TCM under investigation. In addition, it is also of particular interest to the investigators as to how to translate an observed significant difference detected by the Chinese diagnostic procedure to a clinically meaningful difference based on some well-established clinical study endpoint. The purpose of this article is not only to describe some perspectives regarding TCM development, but also to provide some basic considerations regarding practical issues that are commonly encountered during the conduct of clinical trials in the development of TCMs the Western way. These statistical considerations include selection of study design, preparation of matching placebo, development of study endpoint, validation of an instrument, calibration of a validated instrument, and power calculation for sample size estimation.

In the next section, some perspectives regarding TCM development are described. Section 3 provides some basic (statistical) considerations for TCM clinical trials. Some practical issues that are commonly encountered during the process of development of a TCM are reviewed in Section 4. Some concluding remarks are given in the last section.

1. PERSPECTIVES OF TCM CLINICAL DEVELOPMENT

For the development of TCMs, before a TCM clinical trial is conducted, the following questions are necessarily asked.

(1) Will the TCM clinical trial be conducted by Chinese doctors alone, Western clinicians alone, Western

clinicians who have some background and experience of Chinese herbal medicine alone, or both Chinese doctors and Western clinicians?

- (2) Will traditional Chinese diagnostic procedure and/or trial procedures be used throughout the TCM clinical trial?
- (3) Upon approval, is the TCM intended for use by Chinese doctors or Western clinicians?

For the first two questions, if the TCM clinical trial is to be conducted by Chinese doctors alone, the following questions arise. First, should the Chinese diagnostic procedure be validated in order to provide an accurate and reliable assessment of the TCM? In addition, it is of interest to determine how an observed difference obtained from the Chinese diagnostic procedure can be translated to the clinical endpoint commonly used in similar WM clinical trials with the same indication. These two questions can be addressed statistically by the calibration and validation of the Chinese diagnostic procedure with respect to some well-established clinical endpoints for evaluation of Western medicines. If the TCM clinical trial is to be conducted by Western clinicians or Western clinicians who have some background of Chinese herbal medicine, the standards and consistency of clinical results as compared to those WM clinical trials are ensured. However, the good characteristics of TCM may be lost during the process of the conduct of the TCM clinical trials. On the other hand, if the TCM clinical trial is to be conducted by both Chinese doctors and Western clinicians, difference in medical practice and/or possible disagreement regarding the diagnosis, treatment, and evaluation are major concerns.

For the third question, if the TCM is intended for use of

Chinese doctors but it is conducted by Western clinicians, difference in perception regarding how to prescribe the TCM is of great concern. The preparation of a package insert based on the clinical data could be a major issue, not only to the sponsor but also to regulatory authorities. Similar comments apply to the situation where the TCM is intended for use of Western clinicians, but the trial is conducted by Chinese doctors.

As a result, it is suggested that the intention of use (i.e., labeling for the indication) be clearly evaluated when planning a TCM clinical trial. In other words, the sponsor needs to determine whether the TCM is intended for use of Western clinician only, Chinese doctors only, or both Western clinicians and Chinese doctors at the planning stage of a TCM clinical trial, for an adequate package insert of the target diseases under study.

2. BASIC CONSIDERATIONS

To effectively and scientifically evaluate clinical safety and efficacy of a TCM under investigation, a valid study design with some basic considerations for an intended clinical trial is necessarily considered for achieving study objectives with a desired power under the perspectives described in the previous section (see also [4]).

3.1 Selection of study design

Let WW and CW denote conducting a clinical trial the Western way (i.e., based on Western well-established clinical endpoints) and the Chinese way (i.e., based on traditional Chinese diagnostic procedures), respectively. Depending upon the study endpoints (or diagnostic procedures), clinical trials can be classified into the following three different types (see Table 1).

Table 1.Types of WM and TCM Clinical Trials

Pre-treatment (Screening)	Post-treatment	
	WW	CW
WW	type I = WM ₁	type II = TCM ₁
CW	type III = WM ₂	type IV = TCM ₂

Note: WW = the Western way; CW = the Chinese way.

Type I clinical trials are classic clinical trials for Western medicines (WM) which utilize well-established Western clinical endpoints before and after treatment. Type II (TCM₁) clinical trials usually refer to as those TCM clinical trials utilizing well-established Western clinical endpoints at screen (pre-treatment) and evaluating the test treatment the Chinese way (i.e., based on Chinese diagnostic/evaluating procedures) post-treatment. For type III (WM₂) clinical trials usually refer to as those WM clinical trials utilizing Chinese four diagnostic procedures at screening (pre-treatment) and evaluating the test treatment the Western way post-treatment .Type IV clinical

trials are typical TCM clinical trials which adopt Chinese diagnostic and evaluating procedures pre- and post-treatment. As it can be seen from Table 1, for evaluation of a TCM under investigation, one may consider conducting a clinical trial either type I (WM₁), type II (TCM₁), type III (WM₂) or type IV (TCM₂). In practice, however, WM type of clinical trials (type I and type II) and TCM type of clinical trials (type II and type IV) are very likely arriving different conclusions on the test treatment under investigation due to some fundamental differences between WMs and TCMs. In the interest of testing for consistency between the Western way and the Chinese way

for evaluation of the TCM under investigation, the following study designs are useful.

Design A – Subjects will first be screened by using the well-established Western study endpoints. Qualified subjects will then be randomly assigned to receive either a test treatment (T) or a control (C). In each treatment group, subjects are further randomly split into two subgroups: one subgroup will be evaluated the Western way (WW) and the other subgroup will be evaluated the Chinese way (CW). Design A is illustrated in Figure 1. Under this study design, not only that the treatment effect can be evaluated by means of either the Western way or the Chinese way, but also the consistency between WW and CW can be evaluated.

<Figure 1 goes about here>

Design B–Subjects will first be screened by using the Chinese diagnostic procedures. Qualified subjects will then be randomly assigned to receive either a test treatment (T) or a control (C). In each treatment group, subjects are further randomly split into two subgroups: one subgroup will be evaluated the Western way (WW) and the other subgroup will be evaluated the Chinese way (CW). Design B is illustrated in Figure 2. Under this study design, similarly, not only that the treatment effect can be evaluated by means of either the Western way or the Chinese way, but also the consistency between WW and CW can be evaluated.

<Figure 2 goes about here>

Design C – This design is a combination of Design A and Design B. Subjects will be first randomly split into two subgroups. Subjects in one subgroup will be screened by using the well-established Western study endpoints, while subjects in the other subgroup will be screened by means of the Chinese diagnostic procedure. In each subgroup, qualified subjects will then be randomly assigned to receive either a test treatment (T) or a control (C). In each treatment group, subjects are further randomly split into two subgroups: one subgroup will be evaluated the Western way (WW) and the other subgroup will be evaluated the Chinese way (CW). Design C is illustrated in Figure 3. Under this study design, the treatment effect in each subgroup can be evaluated by means of either the Western way or the Chinese way. In addition, the consistency between WW and CW across subgroups can also be evaluated.

<Figure 3 goes about here>

Alternative Design – In practice, it may not be ethical if the disease under study is critical and/or life-threatening provided that a WM is available. Thus, alternatively, Hsiao et al. (2009) proposed randomized placebo-control crossover clinical trial or a parallel-group design consisting of three arms (i.e., the TCM under study, a WM as an active control, and a placebo) [5]. The three-arm, parallel-group design allows the establishment of non-inferiority/equivalence of the TCM as compared to the active control (WM) and the demonstration of the superiority of the TCM with respect to the placebo. One of the advantages of a crossover clinical trial is that a comparison within each individual can be made, although it will take a longer time to complete the study. Although a crossover design requires a smaller sample size as compared to a parallel-group design, there are some limitations for the use of crossover design. First, baselines prior to dosing may not be the same. Second, when a significant sequence effect is observed, we would not be able to isolate the effects of period effect, carry-over effects, and subject-by-treatment effect which are confounded to one another.

3.2 Preparation of matching placebo

In clinical development, double-blind, placebo-control randomized clinical trials are often conducted for evaluation of the safety and effectiveness of a test treatment under investigation. To maintain blindness, a matching placebo should be identical to the active drug in all aspects of, size, color, coating, taste, texture, shape, and order except that it contains no active ingredient. In clinical trials, as advanced technique available for formulation, a matching placebo is not difficult to make because most Western medicines contain single active ingredient. Unlike Western medicines, TCMs usually consist of a number of components, which often have different taste. In TCM clinical trials, the TCM under investigation is often encapsulated. However, the test treatment will be easily unblinded if either the patient or Chinese doctor breaks the capsule. As a result, the preparation of matching placebo in TCM clinical trials plays an important role for the success of the TCM clinical trials.

Fai et al. (2011) pointed out that in many TCM clinical trials, it is very difficult to make a quality matching placebo to achieve the purpose of blinding [6]. Ideally, the characteristics of the test drug and matching placebo should be identical in color, appearance, smell and taste. The quality matching placebo should be identical to the test drug in physical form, sensory perception, packaging, and labeling, and it should

have no pharmaceutical activity. For this purpose, Fai et al. (2011) developed a placebo capsule to match an herbal medicine in terms of its physical form, chemical nature, appearance, packaging and labeling. Based on the assessment results, the developed placebo capsule assessment results suggested that the placebo was found satisfactory in these aspects [6]. Thus, Fai et al. (2011) concluded that a matching placebo could be created for a RCT involving herbal medicine [6]. In addition, Fai et al. (2011) also discussed the means to acquire patent for a developed matching placebo [6].

It should be noted that the preparation of matching placebo is extremely important to maintain the integrity of blinding for avoid any possible operational biases that may be introduced due to the knowledge of the treatment assignment. The oral dosage form of capsule is often considered for preparation of matching placebo for clinical trials involving Chinese herbal medicines as it may remove the strong smell and taste of the herbal medicines. However, one of the major challenges is that patients or clinicians will reveal the treatment assignments if they break the capsules. Thus, standard operating procedures (SOP) for preventing patients and clinicians from breaking the capsules are necessary developed.

3.3 Development of clinical endpoint

Unlike WMs, the primary study endpoints for assessment of safety and effectiveness of a TCM are usually assessed subjectively by a quantitative instrument by experienced Chinese doctors. Although the quantitative instrument is developed by the community of Chinese doctors and is considered a gold standard for assessment of safety and effectiveness of the TCM under investigation, it may not be accepted by the Western clinicians due to fundamental differences in medical theory, perception and practice. In practice, it is very difficult for a Western clinician to conceptually understand the clinical meaning of the difference detected by the subjective Chinese quantitative instrument. Consequently, whether the subjective quantitative instrument can accurately and reliably assess the safety and effectiveness of the TCM is always a concern to Western clinicians.

As an example, for assessment of safety and efficacy of a drug product for treatment of ischemic stroke, a commonly considered primary clinical endpoint is the functional status assessed by the so-called Barthel index. The Barthel index is an ordinal scale used to measure performance in activities of daily living, which was introduced by [7]. The Barthel index is a weighted functional assessment scoring technique composed

of 10 items with a minimum total score of 0 (functional incompetence) and a maximum total score of 100 (functional competence). The Barthel index is a weighted scale measuring performance in self-care and mobility, which is widely accepted in ischemic stroke clinical trials. A patient may be considered a responder if his/her Barthel index is greater than or equal to 60. On the other hand, Chinese doctors usually consider a quantitative instrument developed by the Chinese medical community as the standard diagnostic procedure for assessment of ischemic stroke. The standard quantitative instrument is composed of six domains, which capture different information regarding patient's performance, functional activities, and signs and symptoms and status of the disease.

In practice, it is of interest to both Western clinicians and Chinese doctors how an observed clinically meaningful difference by the Chinese quantitative instrument can be translated to that of the primary study endpoint assessed by the Barthel index. To reduce the fundamental differences in medical theory/perception and practice, it is suggested that the subjective Chinese quantitative instrument be calibrated and validated with respect to that of the clinical endpoint assessed by the Barthel index before it can be used in TCM ischemic stroke clinical trials.

3.4 Validation of a QOL-like instrument

In TCM medical practice, a Chinese doctor usually collects information from the patient with a certain disease through the four subjective diagnostic procedures as described earlier. The purpose of these subjective approaches is to collect information on various aspects of the disease under study such as signs, symptoms, patient's performance and functional activities. In practice, a quality of life like (QOL-like) instrument with a number of questions/items is often considered to quantitatively assess treatment effect. These items are usually grouped to form subscales, composite scores (domains) or overall score for a simple and easy interpretation of the treatment effect. Since the items (subscales) in each subscale (composite score) are correlated, the structure of responses to a quantitative instrument is usually multidimensional, complex and correlated.

Guilford (1954) [8] discussed several methods such as Cronbach's α for measuring the reliability of internal consistency of a quantitative instrument [8]. Guyatt et al. (1989) indicated that a quantitative instrument should be validated in terms of its validity, reproducibility, and

responsiveness [9]. Hollenberg et al. (1991) discussed several methods for validation of a quantitative instrument, such as consensual validation, construct validation and criterion-related validation [10]. There is, however, no gold standard as to how a quantitative instrument should be validated. As indicated in Chow and Liu (2013), the validity of a quantitative instrument is the extent to which the instrument measures what is designed to measure. It is a measure of biasedness of the instrument [3]. The biasedness of a quantitative instrument reflects the accuracy of the instrument. The reliability of a quantitative instrument measures the variability of the instrument, which directly relates to the precision of the instrument. On the other hand, the responsiveness of a quantitative instrument is usually referred to as the ability of the instrument to detect a difference of clinical significance within a treatment.

Hsiao, et al. (2009) considered a specific design for calibration/validation of the Chinese diagnostic procedure [5]. In their proposed study design, qualified subjects are randomly assigned to receive either a TCM or a WM. Each patient will be evaluated by a Chinese doctor and a Western clinician independently, regardless which treatment group he/she is in. As a result, there are four groups of data, namely (i) patients who receive TCM and evaluated by a Chinese doctor, (ii) patients who receive TCM but evaluated by a Western clinician, (iii) patients who receive WM but evaluated by a Chinese doctor, and (iv) patients who receive WM and evaluated by a Western clinician. Groups (iii) and (iv) are used to establish a standard curve for calibration between the TCM and the WM. Groups (i) and (ii) are then used to validate the Chinese diagnostic procedure based on the established standard curve.

3.5 Calibration of a validated instrument

Unlike WMs, the primary study endpoints for assessment of safety and effectiveness of a TCM are usually assessed by a quantitative instrument or the four diagnostic procedures by experienced Chinese doctors. The assessment by a quantitative instrument has been criticized in many ways. First, it may not capture the true health of status of the patients with diseases under study (e.g., by asking wrong questions). Second, it may not detect the effect of the test treatment under investigation. As an example, consider a quantitative instrument with possible scores from 0 (perfect health) to 100 (worst possible disease status). Suppose the scores can be classified into the following categories of health status: *Health* (0-25), *Mild* (26-50), *Moderate* (51-75), and *Severe* (76-100). In this case, there is significant difference between a patient with a score of

25 (*Health*) and a patient with a score of 26 (*Mild*) despite they only differ by one point. On the other hand, a patient with a score of 26 and a patient with a score of 50 are both considered having *Mild* disease status although they differ by 24 points. Thus, the assessment based on a quantitative instrument by experienced Chinese doctors is not only subjective, but also lack of validity. Consequently, the reliability of the assessment is a concern, especially when there is evidence of large rater-to-rater variability.

Thus, although the quantitative instrument is developed by the community of Chinese doctors and is considered a gold standard for assessment of safety and effectiveness of the TCM under investigation, it may not be accepted by the Western clinicians not only due to the lack of validity and reliability, but also the interpretation of the assessment (or translation of assessment to well-established and widely accepted clinical endpoints). In practice, it is very difficult for a Western clinician to conceptually understand the clinical meaning of the difference detected by the subjective Chinese quantitative instrument due to fundamental differences in medical theory, perception and practice.

Thus, for modernization or Westernization of TCMs, whether the subjective quantitative instrument can accurately and reliably assess the safety and effectiveness of the TCM is a concern for development of TCM. In practice, it is then suggested that a clinical trial be conducted to calibrate the subjective quantitative assessment against either life events or well-established clinical endpoints that are commonly used in assessment of Western medicines. The clinical trials should consist of two arms: one arm will include subjects with diseases under study diagnosed by the subjective quantitative instrument and the other arm will include subjects diagnosed by Western diagnostic or testing procedures. Each subject post-treatment will be assessed by both Chinese doctors using the quantitative instrument and Western clinicians based on the well-established and widely accepted study endpoints [5].

3.6 Power calculation for sample size estimation

In clinical trials, sample size is usually selected to achieve a desired power for detecting a clinically meaningful difference in one of the primary study endpoints for the intended indication of the treatment under investigation [11]. As a result, sample size calculation depends upon the primary study endpoint and the clinically meaningful difference that one would like to detect. Different primary study endpoints may result in very different sample sizes.

For illustration purpose, consider the example concerning a TCM for treatment of ischemic stroke, which was developed with more than 30 years clinical experience with humans. Suppose a sponsor would like to conduct a clinical trial to scientifically evaluate the safety and efficacy of the TCM the Western way as compared to an active control (e.g., aspirin). Thus, the intended clinical trial is a double-blind, parallel-group, placebo-control, randomized trial. The primary clinical endpoint is the response rate (a patient is considered a responder if his/her Barthel index is greater than or equal to 60) based on the functional status assessed by the Barthel index. Sample size calculation is performed based on the response rate after 4 weeks of treatment under the hypotheses of testing for superiority. As a result, a sample size of 150 patients per treatment group is required for achieving an 80% power for establishment of superiority of the TCM over the active control agent. Alternatively, we may consider the quantitative instrument developed by experienced Chinese doctors as the primary study endpoint for sample size calculation. Based on a pilot study, about 80% (79 out of 122) of ischemic stroke patients were diagnosed by one domain of the quantitative instrument. A patient is considered a responder if his/her domain score is greater than or equal to 7. Based on this primary study endpoint, a sample size of 90 per treatment group is required to achieve an 80% power for establishment of superiority.

The difference in sample size leads to the question of whether the use of the primary endpoint of response rate based on one domain of the Chinese quantitative instrument could provide substantial evidence of safety and effectiveness of the TCM under investigation.

3. PRACTICAL ISSUES

Before the test treatment under investigation can be used in human, some practical issues regarding sufficient information of chemistry, manufacturing, and control (CMC), clinical pharmacology, and toxicology are necessary considered [12]. However, since most TCMs consist of multiple components with unknown pharmacological activities, information regarding CMC, clinical pharmacology, and toxicology are often difficult to obtain. In what follows, these difficulties are briefly described.

3.1 Test for consistency

Unlike most western medicines (WM), TCMs usually consist of a number of components, which are extracted from herbal samples. The herbal samples are normally dried at 60°C to a constant weight, followed by grinding in a mortar and storing in a desiccator. For water soluble substances, an appropriate

amount of water is first added to the dried material and boil for about one hour. For alcohol soluble substances, 60% ethanol is added and the mixture is extracted at 60°C for one hour in an ultrasonic bath. After cooling to room temperature, the extract can be cleared by filtration through net or centrifugation at 12,000 g for 10 min at 20°C, and the supernatant is used for further applications. The pharmacological activities, interactions, and relative proportions of these components are usually unknown. In practice, TCM is usually prescribed subjectively by an experienced Chinese doctor. As a result, the actual dose received by each individual varies depending upon the signs and symptoms as perceived by the Chinese doctor. Although the purpose of this medical practice is to reduce the within-subject (or intra-subject) variability, it could also introduce non-negligible variability such as variations from component-to-component and from rater-to-rater (a Chinese doctor to another). Consequently, reproducibility or consistency of clinical results is questionable. Thus, how to ensure the reproducibility or consistency of the observed clinical results has become a great concern to regulatory agencies in the review and approval process. It is also a great concern to the sponsor of the manufacturing process. To address the question of reproducibility or consistency, a valid statistical quality control process on the raw materials and final product is essential (see, e.g., [13, 14]).

3.2 Stability analysis

Most regulatory agencies require that the expiration dating period (or shelf-life) of a drug product must be indicated in the immediate container label before it can be released for use. To fulfill this requirement, stability studies are usually conducted in order to characterize the degradation of the drug product. For drug products with a single active ingredient, statistical methods for determination of drug shelf-life are well established (e.g., [15,16]). However, regulatory requirements for estimation of drug shelf-life for drug products with multiple components are not available.

Following the concept of estimating shelf-life for drug products with single active ingredient, two approaches are worth considering. First, we may (conservatively) consider the *minimum* of the shelf-lives obtained from each component of the drug product. This approach is conservative, and yet may not be feasible due to the fact that (i) not all of the components of a TCM can be accurately and reliably quantitated, and (ii) the resultant shelf-life may be too short to be useful. Alternatively, we may consider a two-stage approach for determination of drug shelf-life. At the first stage, an attempt should be made to identify the most active component(s)

whenever possible. A shelf-life can then be obtained based on the method suggested in the FDA and ICH guidelines [15-16]. At the second stage, the obtained shelf-life is adjusted based on the relationship and/or interactions of the most active ingredient(s) and other components. As an alternative, Chow and Shao (2007) proposed a statistical method for determining the shelf-life of a TCM following a similar idea suggested by the FDA, assuming that the components are linear combinations of some factors [17].

3.3 Animal studies

The purpose of animal studies is not only to study possible toxicity in animals, but also to suggest an appropriate dose for use in humans, assuming that the established animal model is predictive of the human model. For a newly developed drug product, animal studies are necessary. However, for some well-known TCMs, which have been used in humans for years and have a very mild toxicity profile, it is questionable whether animal studies are necessary. It is suggested that all components of TCMs as described in Chinese Pharmacopoeia (CP) be classified into several categories depending upon their potential toxicities and/or safety profiles as a basis for regulatory requirements for animal studies. In other words, for some well-known TCM components such as Ginseng, animal studies for testing toxicity may be waived depending upon past experience of human use, although health risks or side effects following the proper administration of designated therapeutic dosages were not recorded in human use. Note that the German regulatory Authority's herbal watchdog agency, commonly called Commission E, has conducted an intensive assessment of the peer-reviewed literature on some 300 common botanicals with respect to the quality of the clinical evidence and the uses for which the herb can be reasonably considered effective [18].

5. CONCLUDING REMARKS

Although TCM has a long history of being used in humans, little or no scientifically valid documentations are available for demonstration of clinical safety and efficacy of the TCM. In the interest of modernization or Westernization of TCM development, as indicated by the FDA, substantial evidence regarding safety and effectiveness of the test treatment under investigation can only be obtained by conducting adequate and well-controlled clinical trials [4]. In TCM clinical trials, however, it is a concern whether a TCM can be scientifically evaluated the Western way due to some fundamental differences between a WM and a TCM.

The purpose of this article is to provide some basic considerations for providing substantial evidence of clinical safety and efficacy of a TCM under investigation during the conduct of TCM clinical trials in the development of TCMs the Western way. These statistical considerations include selection of study design, preparation of matching placebo, development of study endpoint, validation of an instrument, calibration of a validated instrument, and power calculation for sample size estimation.

In addition, as discussed earlier, before a TCM under investigation can be used in human, sufficient information regarding CMC, clinical pharmacology, and toxicology are necessarily provided. In practice, these information, which have impact on the scientific validity for the assessment of the TCM under investigation, are difficult, if not impossible, to obtain. Thus, it is suggested that some practical issues such as test for consistency, stability analysis for shelf-life estimation, and animal studies for toxicity be evaluated before the conduct of the intended TCM clinical trials.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests regarding the publication of this article.

REFERENCES

- [1]. Chow, S.C., Pong, A. and Chang, Y.W. (2006). On traditional Chinese medicine clinical trials. *Drug Information Journal*, 40, 395-406.
- [2]. Chow, S.C. (2015). *Quantitative Methods for Traditional Chinese Medicine Development*. Chapman and Hall/CRC Press, Taylor & Francis, New York.
- [3]. Chow, S.C. and Liu, J.P. (2013). *Design and Analysis of Clinical Trials – Revised and Expanded, Third Edition*, John Wiley & Sons, New York.
- [4]. FDA (1988). *Guideline for the Format and Content of the Clinical and Statistical Sections of New Drug Applications*, The United States Food and Drug Administration, Rockville, Maryland, USA.
- [5]. Hsiao, C.F., Tsou, H.H., Pong, A., Liu, J.P., Lin, C.H., Chang, Y.J., and Chow, S.C. (2009). Statistical validation of traditional Chinese diagnostic procedure. *Drug Information Journal*, 43 83-95.
- [6]. Fai, C.K., Qi, G.D., Wei, D.A., and Chung, L.P. (2011). Placebo preparation for the proper clinical trial of herbal medicine--requirements, verification and quality

control. *Recent Pat Inflamm Allergy Drug Discov*, 5, 169-174.

[7]. Mahoney, F. and Barthel, D. (1965). Functional evaluation: the Barthel Index. *Md. Med. J.*, 14, 61–65.

[8]. Guilford, J.P. (1954). *Psychometric Methods*, 2nd edition. McGraw-Hill, New York, New York.

[9]. Guyatt, G.H., Veldhuyen Van Zanten S.J.O., Feeny, D.H., Patric, D.L. (1989). Measuring quality of life in clinical trials: A taxonomy and review. *Canadian Medical Association Journal*, 140, 1441-1448.

[10]. Hollenberg, N.K., Testa, M., Williams, G.H. (1991). Quality of life as a Therapeutic endpoint – an analysis of therapeutic trials in hypertension. *Drug Safety*, 6, 83-93.

[11]. Chow, S.C., Shao, J. and Wang, H. (2007). *Sample Size Calculation in Clinical Research*. Second Edition, Taylor & Francis, New York, New York.

[12]. FDA (2004). *Guidance for Industry – Botanical Drug Products*. The United States Food and Drug Administration, Rockville,

Maryland, USA.

[13]. Tse, S.K., Chang, J.Y., Su, W.L., Chow, S.C., Hsiung, C., and Lu, Q. (2006). Statistical quality control process for traditional Chinese medicine. *Journal of Biopharmaceutical Statistics*, 16, 861-874.

[14]. Lu, Q., Chow, S.C., and Tse, S.K. (2007). Statistical quality control process for traditional Chinese medicine with multiple correlative components. *Journal of Biopharmaceutical Statistics*, 17, 791-808.

[15]. FDA (1987). *Guideline for Submitting Documentation for the Stability of Human Drugs and Biologics*. Center for Drugs and Biologics, Office of Drug Research and Review, Food and Drug Administration, Rockville, Maryland, USA.

[16]. ICH (1993). *Stability Testing of New Drug Substances and Products*. Tripartite International Conference on Harmonization Guideline Q1A, Geneva, Switzerland.

[17]. Chow, S.C. and Shao, J. (2007). Stability analysis for drugs with multiple ingredients. *Statistics in Medicine*, 26, 1512-1517.

[18]. PDR (1998). *Physicians’ Desk Reference for Herbal Medicines*. Medical Economics Company, Montvale, New Jersey.

Design A

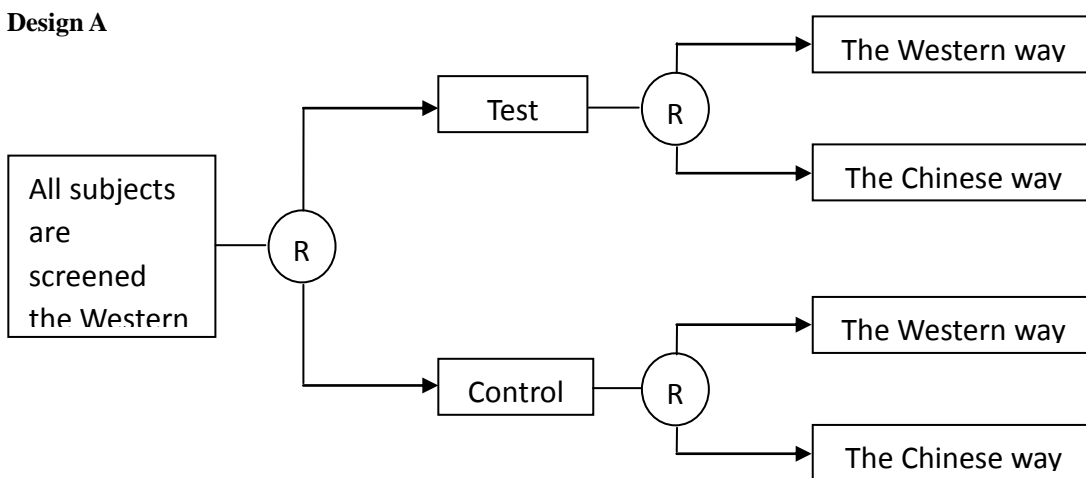


Figure 1. Design for WM Clinical Trials comparing the Western way and the Chinese way

Design B

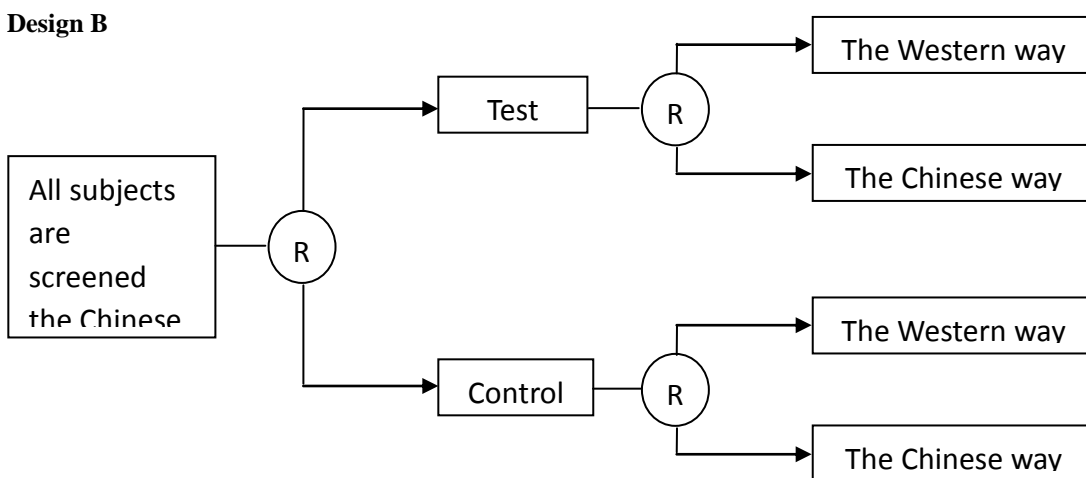


Figure 2. Design for TCM Clinical Trials comparing the Western way and the Chinese way

Design C

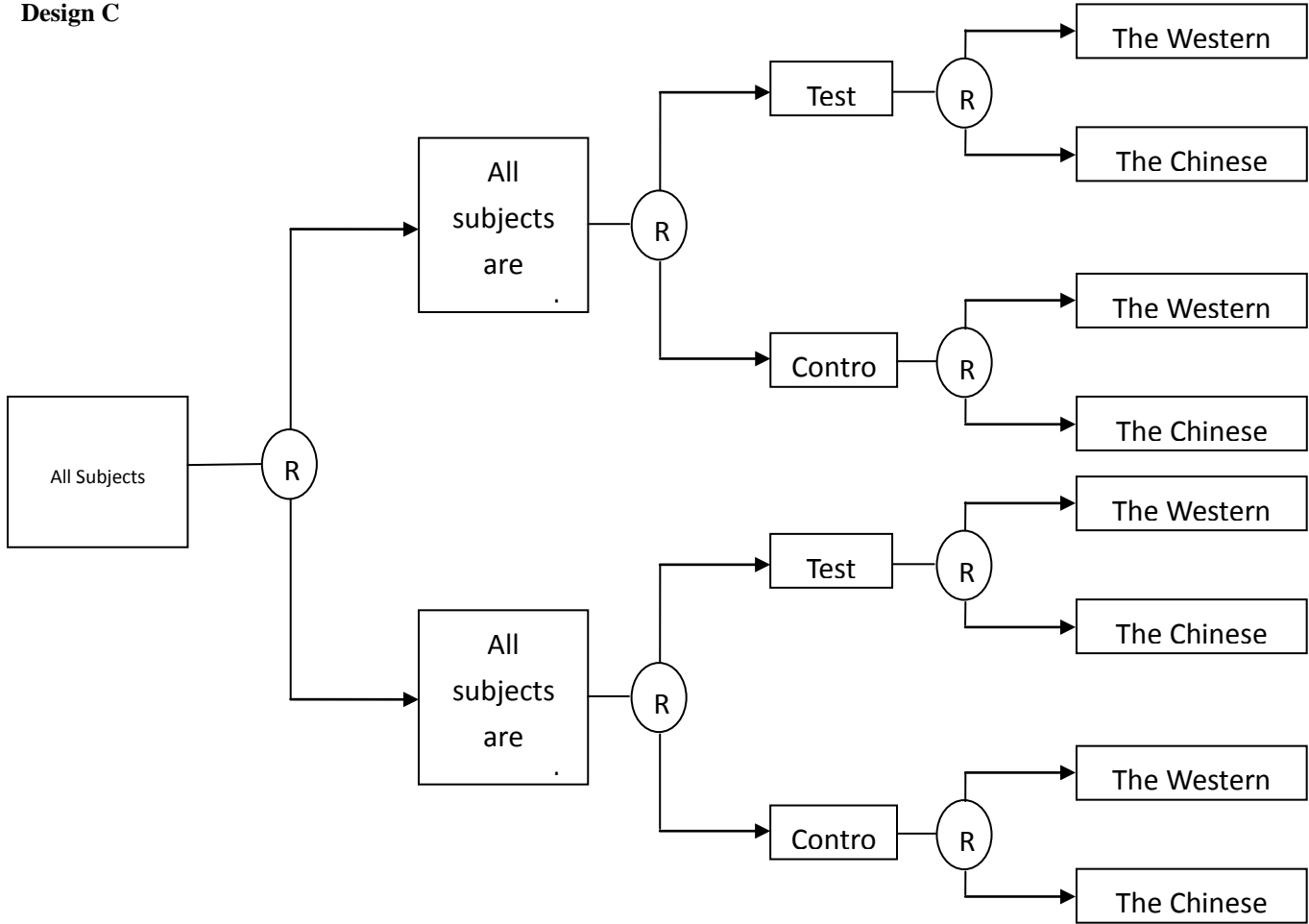


Figure 3. Design for Combining WM and TCM Clinical Trials comparing the Western way and the Chinese way