



Research Article

Copyright@ Dezhao Fu

# Novel Design for Validation of Study Endpoints in TCM Clinical Trials

Dezhao Fu<sup>1\*</sup> and Shein Chung Chow<sup>1</sup>

<sup>1</sup>Department of Biostatistics and Bioinformatics, Duke University School of Medicine, USA

\*Corresponding author: Dezhao Fu, Department of Biostatistics and Bioinformatics, Duke University School of Medicine, USA

To Cite This Article: Dezhao Fu and Shein Chung Chow. Novel Design for Validation of Study Endpoints in TCM Clinical Trials. Am J Biomed Sci & Res. 2023 18(1) AJBSR.MS.ID.002422, DOI: [10.34297/AJBSR.2023.18.002422](https://doi.org/10.34297/AJBSR.2023.18.002422)

Received: 📅 December 12, 2022; Published: 📅 February 09, 2023

## Abstract

In recent years, the search for traditional Chinese medicines (TCM) for treating patients with critical and/or life-threatening diseases has attracted much attention in the pharmaceutical industry. However, the evaluation of safety and efficacy of a TCM under investigation has been criticized due to its subjectivity (i.e., experience based rather than scientific based). Thus, statistical validation of study endpoints (usually based on a quality of-life-like instrument) is essential to have an accurate and reliable clinical assessment of the performance of the TCM under study. Hsiao et al. [1] proposed some statistical methods for calibration/validation of Chinese study endpoints against Western clinical endpoints in terms of the performance characteristics of validity, reliability, and ruggedness under a valid study design. In this article, we proposed some innovative study designs for calibration/validation of Chinese study endpoints against Western clinical endpoints. under certain considerations of study design.

**Keywords:** Validity, Reliability, Calibration, Validation

## Introduction

In recent years, the search for traditional Chinese (herbal) medicine (TCM) for treating patients with critical and/or life-threatening diseases has attracted much attention in the pharmaceutical industry [2]. However, there are some fundamental differences between Western medicine (WM) and TCM (see Table 1). For example, most WMs contain a single active ingredient, while TCMs consist of multiple active components. For assessment of treatment effect of the test drug under investigation, WMs utilize well-established/validated and objective clinical endpoints which

are considered evidence-based clinical endpoints, while TCMs adopt four subjective Chinese diagnostic procedures (techniques) (CDPs) which are known as experience-based study endpoints [3]. From statistical point of view, WMs with fixed dose aim to achieve the ultimate goal of precision medicine (which minimizes inter-subject variability), while TCMs with flexible dose lean toward individualized or personalized medicine (which minimize intra-subject variability) [2]. Thus, for development of TCMs, some basic considerations are necessarily considered for providing accurate and reliable assessment of safety and efficacy of the TCMs under investigation [4].

**Table 1:** Fundamental Differences Between a WM and a TCM.

Description	WM	TCM
Active Ingredient	Single	Multiple
Dose	Fixed	Flexible
Study Endpoint	Objective validated	Subjective not validated
Therapeutic Index	Well-established	Not well-established
Medical Mechanism	Specific organs	Global dynamic among Organs
Medical Perception	Evidence-based	Experience-based
Primary focus	Population	Individual
Statistical assessment	Minimize inter-subject variability	Minimize intra-subject variability
Ultimate Goal	Precision medicine	Individualized or Personalized medicine

WM: Western Medicine; TCM: Traditional Chinese Medicine

In this article, we will focus on comparison between WM clinical endpoints and Chinese study endpoints in clinical development of TCMs. For development of a test treatment or drug product, clinical trials are often conducted to demonstrate the safety and efficacy of the drug product under investigation in terms of some pre-specified study endpoints. A clinical endpoint is defined as an event or outcome that can be measured objectively to determine whether the treatment being studied is beneficial or harmful. Thus, endpoint selection plays an important role in clinical trials for providing an accurate and reliable assessment of treatment effect in terms of safety and efficacy of the test drug under study. For development of a Western medicine (WM), objective and valid study endpoints which are acceptable to the regulatory agencies such as United States (US) Food and Drug Administration (FDA) are often used for regulatory submission, review, and approval [4].

Unlike Western clinical endpoints, in practice, the assessment of the safety and efficacy of a TCM under investigation is often conducted based on some Chinese study endpoints, which are often derived from some Chinese diagnostic procedures (CDP). Traditional CDP consists of four major techniques (categories), namely, looking (i.e., inspection), listening and smelling (i.e., auscultation and olfaction), asking (i.e., interrogation), and touching (i.e., pulse taking and palpation). The Chinese endpoints for assessment of the safety and efficacy of TCMs under study have been criticized being subjective and yet not reliable [3]. In practice, Western clinical endpoints are considered evidence-based study endpoints which can provide accurate and reliable assessment of drug product under investigation, while the Chinese endpoints are experience-based study endpoints which are subjective with expected higher rater-to-rater variability [5].

Moreover, it is not clear how the Western endpoints are translated to Chinese endpoints in terms of disease status (diagnosis) and treatment effect (clinically meaningful difference). Thus, it is of interest to determine how accurate and reliable of the subjective CDP for evaluation of patients with certain diseases under study. Also, it is of interest to determine how an observed change in the CDP is translated to a change in a well-established Western clinical endpoint [6]. In this article, we will examine the above mentioned two questions by studying the calibration and

validation of the Chinese endpoint (in terms of GDP) of a TCM with respect to a well-established Western clinical endpoint for evaluation of a TCM. The calibration and validation of Chinese study endpoints (GDP) against Western clinical endpoint is discussed in Section 3. Several innovative study designs for TCM clinical trials are described in Section 4. Some concluding remarks are given in the last section.

## End Point Selection

In clinical trials, endpoint selection plays an important role which has an impact on power calculation for sample size requirement. Appropriate endpoint selection cannot only ensure that the intended trial will achieve the study objectives at a pre-specified level of significance, but also increase the probability of the success of the intended study [5,6].

## WM Clinical Endpoints

In WM clinical trials, a study endpoint is defined as an event or outcome that can be measured objectively to determine whether the treatment being studied is beneficial. The endpoints of a clinical trial are usually included in the study objectives. Some examples of endpoints are survival, response rate, and time-to-disease progression, improvements in quality of life, and relief of symptoms in cancer clinical trials [2]. WM clinical endpoints are considered objective and scientifically validated which can be used to accurately and reliably inform disease status and treatment effect of the test treatment under investigation.

## TCM Study Endpoints

Unlike WM clinical endpoints, in TCM clinical trials, a Chinese study endpoint is considered an event or outcome that is measured by four Chinese diagnostic procedures, namely, inspection, auscultation and olfaction, interrogation, and pulse taking and palpation. Inspection involves observing the patient's general appearance (strong or weak, fat or thin), mind, complexion (skin color), five sense organs (eye, ear, nose, lip, and tongue), secretions, and excretions. Auscultation involves listening to the voice, expression, respiration, vomit, and cough. Olfaction involves smelling the breath and body odor [2]. Interrogation involves asking questions about specific symptoms and the general condition

including history of the present disease, past history, personal life history, and family history. Pulse taking and palpation can help to judge the location and nature of a disease according to the changes of the pulse [7].

After these four diagnostic techniques have been performed, the TCM doctor has to configure a syndrome diagnosis describing the fundamental substances of the body and how they function in the body based on the eight principles, five element theory, five Zang and six Fu, and information regarding channels and collaterals. Eight principles consist of Yin and Yang (i.e., negative and positive), cold and hot, external and internal, and Shi and Xu (i.e., weak and strong). Eight principles can help the TCM doctors to differentiate syndrome patterns. For instance, Yin people will develop disease in a negative, passive, and cool way (e.g., diarrhea and back pain), while Yang people will develop disease in an aggressive, active, progressive, and warm way (e.g., dry eyes, tinnitus, and night sweats). The five elements (earth, metal, water, wood, and fire) correspond to particular organs in the human body. Each element operates in harmony with the others [1,3].

Although these diagnostic techniques aim to provide objective diagnosis/assessment of the disease under study by collecting symptoms and signs from the patient, the Chinese study endpoints have been criticized being subjective and not scientifically valid. For example, the smallest detail can have a major impact on the treatment scheme as well as on the prognosis. While the pulse diagnosis and examination of the tongue receive much attention due to their frequent mention, the other aspects of diagnosis cannot be ignored [1].

### A Comparison

There are fundamental differences between WM clinical endpoints and Chinese study endpoints. First, as indicated earlier, WM clinical endpoints are considered well established/validated and objective study endpoints, while the Chinese study endpoints are considered subjective and may not be scientifically valid. Second, power calculation for sample size requirement based on either WM clinical endpoints or Chinese study endpoints may be different. Third, WM clinical endpoints and Chinese study endpoints may not be translated each other in terms of clinically meaningful difference [6].

### Calibration/Validation Against Western Clinical Endpoints

Before a Chinese endpoint (or a quantitative instrument) in TCM clinical trials can be validated with respect to a well-established clinical endpoint for a Western medicine, the four Chinese diagnostic procedure is necessarily calibrated against measurements obtained from the well-established clinical endpoint. A common approach to the calibration of a quantitative instrument with respect to some well-established clinical endpoints is to have several groups of patients with known measurements of the well-established clinical endpoint. The quantitative instrument is then applied to these patients to obtain the corresponding responses (scores). On the

basis of these measurements of clinical endpoint (standards) and their corresponding responses (scores), an estimated calibration curve can be obtained by fitting an appropriate statistical model between these standards and their corresponding responses. The estimated calibration curve is also known as the standard curve. For a given patient, his/her unknown measurement of well-established clinical endpoint can be determined based on the standard curve by replacing the dependent variable with its response.

### Calibration

Let  $N$  and  $x_i$  be the number of patients and measurement of the well-established WM clinical endpoint of the  $j$ th patient. For simplicity, we assume that the measurement of well-established clinical endpoint is continuous. Suppose that the TCM diagnostic procedure consists of  $K$  items. Let  $z_{ij}$  denote the TCM diagnostic score of  $j$ th patient from the  $i$ th item,  $i = 1, \dots, K$ ,  $j = 1, \dots, N$ . Let  $y_j$  represent the scale (or score) of the  $j$ th patient summarized from the  $K$  TCM diagnostic items. For simplicity, we assume that

$$ky_i = XZ_{ij} \\ i = 1$$

Similar to calibration of an analytical method [8], the following candidate models are often considered for calibration:

$$\text{Model 1: } y_i = \alpha + \beta x_j + \varepsilon_j$$

$$\text{Model 2: } y_j = \beta x_j + \varepsilon_j$$

$$\text{Model 3: } y_j = \alpha + \beta_1 x_j + \beta_2 x_j^2 + \varepsilon_j$$

$$\text{Model 4: } y_j = \alpha x_j^\beta \varepsilon_j$$

$$\text{Model 5: } y_j = \alpha \varepsilon^{\beta x_j} \varepsilon_j$$

where  $\alpha$ ,  $\beta$ ,  $\beta_1$ , and  $\beta_2$  are unknown parameters and  $\varepsilon$ 's are independent random errors with  $E(\varepsilon_j) = 0$  and finite  $V \text{ ar}(\varepsilon_j)$  in models 1-3 and  $E(\log(\varepsilon_j)) = 0$  and finite  $V \text{ ar}(\log(\varepsilon_j)) = 0$  in models 4-5.

Model 1 is a simple linear regression model which is probably the most commonly used statistical model for establishment of standard curves for calibration. When the standard curve passes through the origin, model 1 reduces to model 2. Model 3 indicates that the relationship between  $y$  and  $x$  is quadratic. When there is a nonlinear relationship between  $y$  and  $x$ , models 4 and 5 are useful. Note that both models 4 and 5 are equivalent to simple linear regression model after logarithm transformation [9].

### Validation

Since the diagnostic procedure of a TCM could vary from a Chinese doctor to another, a standardized diagnostic procedure is usually developed prior to the conduct of a clinical trial. As indicated earlier the standardized TCM diagnostic instrument usually contains the four categories or domains, which in turn consist of

a number of questions agreed by the community of the Chinese doctors. For validation of such an instrument, [1] considered the following validation performance characteristics (parameters): validity (or accuracy) and reliability (or precision), which are briefly outlined below [10,11].

### Validity

The validity itself is a measure of biasedness of the TCM instrument. As mentioned earlier a TCM instrument usually consists of a number of questions agreed by the community of the Chinese doctors. It is a great concern that the questions may not be the right questions to capture the information regarding patient's activity/function, disease status, and disease severity. Let  $X$  be the unobservable measurement of the well-established clinical endpoint which can be quantified by the TCM items,  $Z_{i,i} = 1, \dots, K$  based on the estimated standard curve in previous section.

Suppose that  $X$  is distributed as a normal distribution with mean  $\theta$  and variance  $\tau^2$ . Let  $Z = (Z_1, \dots, Z_K)'$ . Again, suppose  $Z$  follows a distribution with mean  $\mu = (\mu_1, \dots, \mu_K)'$  and variance  $P$ . To assess the validity, it is desired to see whether the mean of  $Z_{i,i} = 1, \dots, K$  is close to  $(\alpha + \beta\theta)/K$ .

$$\text{Let } \bar{\mu} = \frac{1}{K} \sum_{i=1}^K \mu_i$$

Then  $\theta = (\bar{\mu} - \alpha)/\beta$ .

Consequently, we can claim that the instrument is validated in terms of its validity if

$$|\mu_i - \mu| < \delta, \forall i = 1, \dots, K, \quad (1)$$

for some small pre-specified  $\delta$ . To verify [1], we can consider construct a simultaneous confidence interval for  $\mu_i - \mu$ . Assume that the TCM instrument is administered to  $N$

patients from subgroup 1. Let

$$-\hat{\mu} - \frac{1}{N} \sum_{j=1}^N Z_j = Z$$

Then the  $(1 - \alpha)100\%$  simultaneous

confidence interval for  $\mu_i - \mu$  are given by

$$\mu - \sqrt{\frac{1}{N} a_i S a_i T(\alpha, k, N - k)} \leq \mu_i - \hat{\mu} \leq \hat{\mu} + \sqrt{\frac{1}{N} a_i S a_i T(\alpha, k, N - k)},$$

$$i = 1, \dots, k$$

$$\text{Where } a_i = \begin{pmatrix} -\frac{1}{k} 1_{i-1} \\ 1 - \frac{1}{k} \\ -\frac{1}{k} 1_{i-1} \end{pmatrix}, 1_i = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}_{(i-1) \times 1}, 1_{k-i} = S = \frac{1}{N-1} \sum_{j=1}^N (Z_j - \bar{Z})(Z_j - \bar{Z})'$$

$$\text{and } P(T^2(\alpha, K, N - K) \leq T^2(\alpha, K, N - K)) = 1 - \alpha$$

The Bonferroni adjustment of an overall  $\alpha$  level might be conducted as follows:

$$a_i \hat{\mu} - \sqrt{\frac{1}{N} a_i S a_i T(\frac{\alpha}{2K}, N - 1)} \leq \mu_i - \bar{\mu} \leq \hat{\mu} + \sqrt{\frac{1}{N} a_i S a_i T(\frac{\alpha}{2K}, N - 1)}$$

Thus, we can reject the null hypothesis that

$$H_0 : |\mu_i - \mu| \geq \delta, \forall i = 1, \dots, K, \quad (2)$$

if any confidence interval falls completely within  $(-\delta, \delta)$ .

### Reliability

The calibrated well-established clinical endpoints derived from the estimated standard curve are considered reliable if the variance of  $X$  is small. In this regard, we can test the hypothesis:

$$H_0 : \tau^2 \geq \Delta \quad \text{Vs} \quad H_A : \tau^2 < \Delta \quad (3)$$

for some fixed  $\Delta$  to verify the reliability of estimating  $\theta$  by  $X$ . We can then verify the reliability based on the previously established standard curve for calibration. Based on the estimated standard curve, we can derive that

$$\tau^2 = \frac{1}{\beta^2} \text{Var}(\sum_{i=1}^K Z_i)$$

$$= \frac{1}{\beta^2} 1' \Sigma 1$$

Thus, we may reject the null hypothesis of (3) at the  $\alpha$  level of significance if,

$$Q = \frac{\sum_{j=1}^N (X_j - \bar{X})^2}{\Delta} < \chi^2(1 - \alpha, N - 1),$$

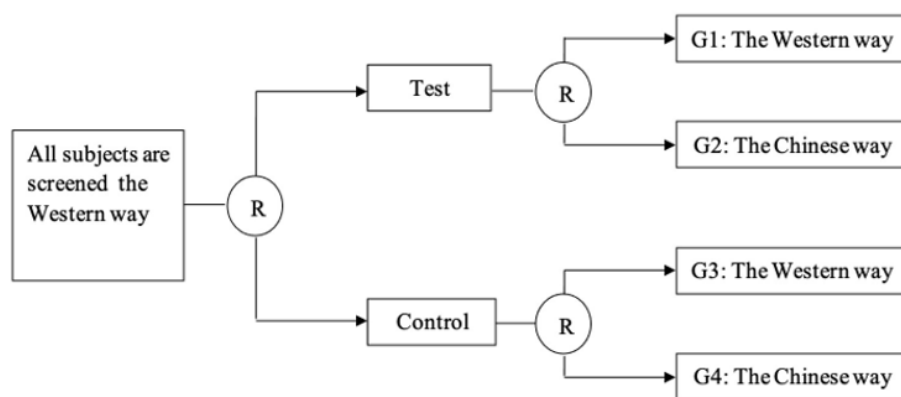
where  $\chi^2(1 - \alpha, N - 1)$  is the  $(1 - \alpha)^{\text{th}}$  upper quantiles of a central chi-square distribution with  $N - 1$  degrees of freedom.

### Innovative Study Designs

In the interest of calibration between the Western clinical endpoints and the Chinese study endpoints for evaluation of the TCM under investigation, in addition to the design proposed by Hsiao, et al. [1], the following innovative designs are useful [7].

#### Design A

In this design, 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).

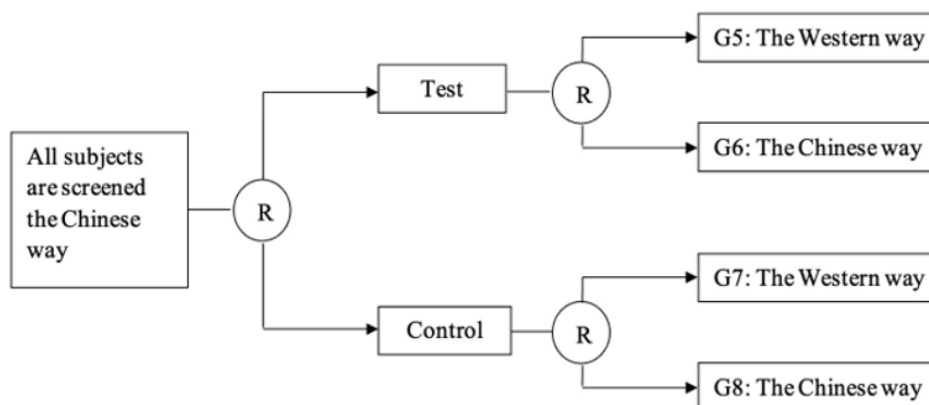


**Figure 1:** Design for WM Clinical Trials comparing the Western way and the Chinese way.

### Design B

For 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.



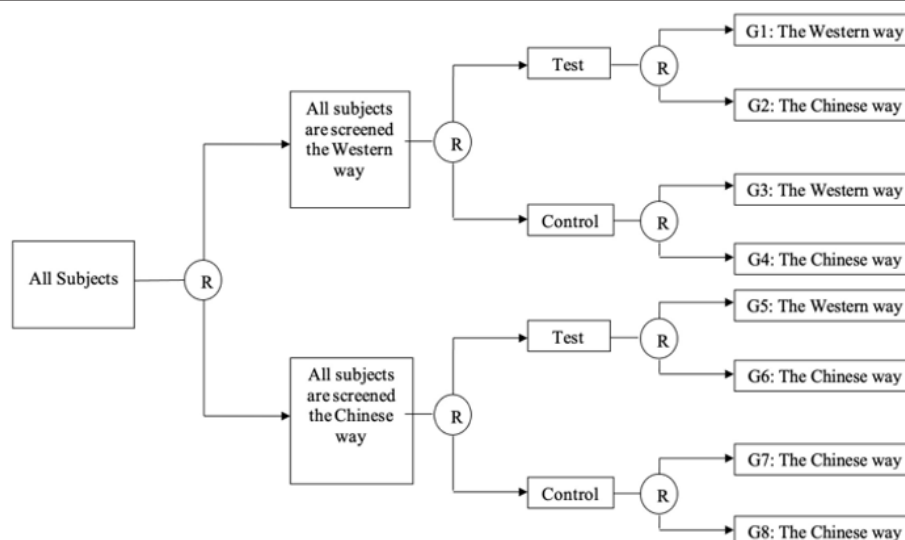
**Figure 2:** Design for TCM Clinical Trials comparing the Western way and the Chinese way.

### 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).





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

In order to evaluate consistency between the outcome groups, we conducted two-sample t-test to determine the treatment effect of two groups. For each outcome group, the treatment effect is evaluated by mean of the group as:

$$\mu_{ijk}$$

where  $i$  is the determinant of the screened type of the group, where  $i = 1$  means that all subjects are screened the Western way, and  $i = 2$  means that all subjects are screened the Chinese way;  $j$  is the determinant of the type of received treatment,

where

**Table 2:** Comparison of Different Outcome Groups.

Design Type		Comparison Of Interest	Hypothesis Of Interest	Description
Design C	Design A	G1 vs G2	$H_0: \mu_{121} = \mu_{122}$ $H_A: \mu_{121} \neq \mu_{122}$	compare the diagnosis between test group evaluated under The Western Way and test group evaluated under The Chinese Way when all subjects are screened The Western Way
		G3 vs G4	$H_0: \mu_{111} = \mu_{112}$ $H_A: \mu_{111} \neq \mu_{112}$	compare the diagnosis between control group evaluated under The Western Way and control group evaluated under The Chinese Way when all subjects are screened The Western Way
		G1 vs G3	$H_0: \mu_{121} = \mu_{111}$ $H_A: \mu_{121} \neq \mu_{111}$	compare the diagnosis between test group evaluated under The Western Way and control group evaluated under The Western Way when all subjects are screened The Western Way
		G2 vs G4	$H_0: \mu_{122} = \mu_{112}$ $H_A: \mu_{122} \neq \mu_{112}$	compare the diagnosis between test group evaluated under The Chinese Way and control group evaluated under The Chinese Way when all subjects are screened The Western Way
		G1+G2 vs G3+G4	$H_0: \mu_{11} = \mu_{12}$ $H_A: \mu_{11} \neq \mu_{12}$	compare the diagnosis between groups evaluated under The Western Way and groups evaluated under The Chinese Way when all subjects are screened The Western Way
		G5 vs G6	$H_0: \mu_{221} = \mu_{222}$ $H_A: \mu_{221} \neq \mu_{222}$	compare the diagnosis between test group evaluated under The Western Way and test group evaluated under The Chinese Way when all subjects are screened The Chinese Way

$j = 1$  means that the group received control treatment (C)

$j = 2$  means that the group received test treatment (T);

$k$  is the determinant of the subgroup type,

where

$k = 1$  means that the subgroup is evaluated through the Western way

$k = 2$  means that the subgroup is evaluated through the Chinese way (Table 2).

		G7 vs G8	H0: $\mu_{211} = \mu_{212}$ HA: $\mu_{211} \neq \mu_{212}$	compare the diagnosis between control group evaluated under The Western Way Chinese Way when all subjects are screened The Chinese Way
		G5 vs G7	H0: $\mu_{221} = \mu_{211}$ HA: $\mu_{221} \neq \mu_{211}$	compare the diagnosis between test group evaluated under The Western Way and control group evaluated under The Western Way when all subjects are screened The Chinese Way
		G6 vs G8	H0: $\mu_{212} = \mu_{222}$ HA: $\mu_{212} \neq \mu_{222}$	compare the diagnosis between test group evaluated under The Chinese Way and control group evaluated under The Chinese Way when all subjects are screened The Chinese Way
		G5+G6 vs G7+G8	H0: $\mu_{21} = \mu_{22}$ HA: $\mu_{21} \neq \mu_{22}$	compare the diagnosis between groups evaluated under The Western Way and groups evaluated under The Chinese Way when all subjects are screened The Chinese Way

As shown in the table, we use a two-sample t-test to test the hypothesis of interest. We assume that each group ( $G_{m,m} = 1,2,3,\dots,8$ ) is independent to each other and normally distributed with the mean of  $\mu_{ijk}$  ( $i = 1,2; j = 1,2; k = 1,2$ ).

If we assume the variances for all groups ( $G_{m,m} = 1,2,3,\dots,8$ ) are equal, by applying the test statistics under null hypothesis:

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t_{(n_1+n_2-2)}$$

If we assume the variances are not equal and the sample sizes are large for all groups, applying the test statistics under null hypothesis:

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t_{v,v} = \frac{v_1^2 - v_2^2}{\sqrt{\frac{v_1^2}{n_1-1} + \frac{v_2^2}{n_2-1}}}$$

Thus, we can reject the null hypothesis of interest shown in the Table 2 at the  $\alpha$  level of significance if:

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} < t_{(n_1+n_2-2)}(1-\alpha)$$

## Concluding Remarks

In TCM clinical trials, the validation of Chinese study endpoint against WM clinical endpoint is critical not only to provide an accurate and reliable assessment of the safety and effectiveness of the TCM under investigation, but also translate the interpretation of clinically meaningful difference from Chinese doctor's perspective to Western clinician's perspective. The calibration of the quantitative instrument with respect to a well-established clinical endpoint provides the clinicians (both Chinese doctors and Western clinicians) a better understanding whether the observed significant difference from the quantitative instrument is clinically meaningful [2].

In clinical development of TCMs, however, validated quantitative instruments for the diseases under study may not be available. In this case, it is suggested that a small-scale validation pilot study be conducted to validate the quantitative instrument against Western clinical endpoint for a valid assessment of the safety and efficacy of the TCM under investigation. If such a small-scale pilot study is not feasible, concurrent validation using a valid study design (e.g., Design A, Design B or Design C) as described in Section 3 may be useful. In many case, retrospective validation may also be considered. Based on a well-calibrated and validated quantitative instrument, sample size calculation for achieving a desired power of detecting a clinically meaningful difference can then be accurately performed.

## References

- Hsiao CF, Tsou HH, Pong A, Jen-pei Liu, Chien-Hsiung Lin, et al. (2009) Statistical validation of traditional chinese diagnostic procedures. Drug information journal: DIJ/Drug Information Association 43(1): 83-95.
- Chow SC (2015) Quantitative methods for Traditional Chinese Medicine Development, 83.
- Chow SC, Pong A, Chang YW (2006) On traditional chinese medicine clinical trials. Drug information journal 40(4): 395-406.
- FDA (2016) Botanical Drug Development; Guidance for Industry. The United States Food and Drug Administration, Rockville, Maryland, USA.
- Chow SC, Shao J, Wang H (2017) Sample size calculations in clinical research. chapman and hall/CRC.
- Chow SC, Shao J, Ho HT (2000) On statistical analysis for placebo-challenging designs in clinical trials. Stat Med 19(8): 1029-1037.
- Shein Chung C, Fuyu S (2015) Statistical considerations for traditional chinese medicine clinical trials. Case Studies Journal 4(5): 134-142.
- Chow S, Liu J (1995) Validation, process controls and stability, statistical design and analysis in pharmaceutical science.
- Tsex SK, Chow SC (1995) On model selection for standard curve in assay development. J Biopharm Stat 5(3): 285-296.
- Shein-Chung C, Fanny YK (1996) Statistical issues in quality-of-life assessment. J Biopharm Stat 6(1): 37-48.
- Chow SC, Ki FY (1994) On statistical characteristics of quality-of-life assessment. Journal of Biopharmaceutical Statistics 4(1): 1-17.