

Establishing an evaluation mode with multiple primary outcomes based on combination of diseases and symptoms in TCM clinical trials

Jing Hu¹, Shuo Liu², Weihong Liu¹, Huina Zhang¹, Jing Chen³, Hongcai Shang²

¹Beijing Traditional Chinese Medicine Hospital, Capital Medical University, Beijing Institute of Traditional Chinese Medicine, Beijing 100010, China; ²Key Laboratory for Internal Chinese Medicine of Ministry of Education and Beijing, Dongzhimen Hospital, Beijing University of Chinese Medicine, Beijing 100700, China; ³Tianjin University of Traditional Chinese Medicine, Tianjin 300193, China

Contributions: (I) Conception and design: J Hu, H Shang, J Chen; (II) Administrative support: S Liu; (III) Provision of study materials or patients: W Liu, H Zhang; (IV) Collection and assembly of data: W Liu, H Zhang; (V) Data analysis and interpretation: W Liu, H Zhang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Hongcai Shang. Key Laboratory for Internal Chinese Medicine of Ministry of Education and Beijing, Dongzhimen Hospital, Beijing University of Chinese Medicine, Beijing 100700, China. Email: shanghongcai@foxmail.com; Jing Chen. Tianjin University of Traditional Chinese Medicine, Tianjin 300193, China. Email: cjshcsyc@126.com.

Abstract: The development of a scientific evaluation tool is a prerequisite to reflect the efficacy of traditional Chinese medicine (TCM). At present, separate analysis of each outcome is the most commonly used method in randomized controlled trials (RCTs) of TCM, while this method has been reported to have some limitations. Combination evaluation of multiple primary outcomes including diseases and symptoms can reflect the efficacy of TCM comprehensively and objectively. Now, several multivariate statistical methods have been proposed to analyze RCTs with multiple clinical outcomes. In this article, we highlighted the two commonly used methods, global statistical test (GST) methodology and alpha-adjustment procedures, and introduced an evaluation mode. First, we used one stage adaptive adjustment strategy to estimate the correlation among multiple primary outcomes in blindness, which can resolve the inappropriate correlation estimation among multiple primary outcomes, then we introduced an adaptation of the Bonferroni procedure which can account for correlated data to calculate the alpha of individual outcome. GST using O'Brien ranking procedure and the corresponding global treatment effect (GTE) measure were used to assess the treatment's global impact. We will conduct a randomized, double-blind, controlled trial with compound Danshen dripping pills for stable angina to explore establishing the evaluation mode. We hope that the introduction of this approach can provide methodological aid for the assessment of therapeutic effect of TCM.

Keywords: Combination of diseases and syndromes; multiple primary outcomes; randomized controlled trials (RCTs); traditional Chinese medicine (TCM)

Submitted Dec 14, 2016. Accepted for publication May 26, 2017.

doi: 10.21037/atm.2017.06.06

View this article at: <http://dx.doi.org/10.21037/atm.2017.06.06>

Introduction

Randomized controlled trials (RCTs) are widely accepted as the “gold standard” for comparing different therapeutic modalities. Since the first RCT of traditional Chinese medicine (TCM) was published in 1983 (1), RCTs have been generally used to assess the clinical efficacy of TCM. While, as many researchers indicated, RCTs in TCM come with some challenges (2-4). For TCM, the human body is ideally understood as an interconnected dynamical network of mental, physical, and spiritual processes, each of which is constantly affected by the other. Health is understood as an intricate and ongoing balance of these multiple processes, and disease is understood to be a manifestation of imbalance at many levels of the self, which is known as holism (5). The theory makes TCM for diseases are multidimensional; multiple patient-reported, laboratory test, clinician-rated, TCM syndrome outcomes are often used in evaluations of treatment impact of TCM. Differences like these can make it difficult to research TCM with current conventional RCT. The conventional design of RCTs often selects a single primary outcome that provides a complete characterization of the disease and permits an efficient evaluation of the effect of intervention (6), while TCM looks to many outcomes at once to understand effectiveness. So selecting a single primary outcome may be inappropriate because a single measure may not sufficiently characterize the effect of a TCM intervention on a broad set of domains (7,8). However, at present, this strategy still is the most commonly used in TCM trials. Therefore, the innovation in TCM evaluation methodology is highly demanded.

In this article, we analyze the limitations with the most commonly used outcome assessments method in TCM RCTs, and present an evaluation mode with multiple primary outcomes based on combination of diseases and symptoms. We will conduct a randomized, double-blind, controlled trial with compound Danshen dripping pills for stable angina to explore establishing this evaluation mode.

Limitations with the separate analysis of each outcome in TCM RCTs

Like the recommendation of ICH E9 guideline on biostatistics (6), RCTs of TCM are commonly designed

with a single primary outcome, leaving all others as secondary outcomes. The primary outcome is usually Western Medicine (WM)-specific outcome such as physiological or biochemical parameters. TCM-specific outcomes such as tongue and pulse characteristics, symptoms and signs are often listed as secondary ones. When analyzing these multiple outcomes, the common method is separate testing of each individual outcome (9), most often without adjustment for multiple testing.

There are two major drawbacks of this method: (I) because the multiple outcomes are not accounted for in the statistical analysis appropriately, the probability of obtaining statistically significant results by chance may increase (10,11); (II) TCM and WM often hold different viewpoints on diseases because their theories are distinct. TCM emphasizes the improvement in symptoms, while WM may focus on the improvement of objective medical indicators like physiological and biochemical indexes. If we use this strategy in TCM RCTs, the primary WM-specific outcome maybe not show treatment benefit, but all the secondary TCM-specific outcomes show benefit, the results could be difficult to interpret.

Moreover, for TCM clinical trials, the goal is to determine whether one TCM intervention is preferred over one WM intervention based on multiple important outcomes. Separate analysis of each outcome is not accordance with the multidimensional characteristic of TCM for diseases.

Multiple primary outcomes with diseases and symptoms in TCM RCTs

Since TCM for diseases is multidimensional, it is difficult to identify a single most important outcome as the primary outcome to summarize the efficacy of TCM. Multiple WM-specific and TCM-specific outcomes are often required to be analyzed jointly to determine whether a TCM intervention should be recommended. So we present a clinical efficacy evaluation system with multiple primary outcomes based on combination of diseases and symptoms in TCM clinical trials.

The multiple primary outcomes should include three core domains: (I) WM-specific outcome (e.g., physiological and biochemical indicators); (II) TCM

syndrome outcome (e.g., tongue and pulse characteristics); (III) quality of life. This has been widely accepted by TCM researchers (12-14).

Several multivariate statistical methods have been proposed to analyze clinical trials with multiple clinical outcomes, including the use of a linear combination of several outcomes, comprehensive evaluation method, alpha-adjustment procedures, omnidirectional tests, hierarchical models using latent parameters or hyperparameters, and global statistical test (GST) (15-17). This article highlights the two commonly used methods, GST methodology and alpha-adjustment procedures.

GST

GST can combine information from multiple outcomes into a single test of treatment effectiveness and take into account the correlations among outcomes (18). The strength of GST is it can test a treatment's global benefit across different outcomes and determine whether a treatment is preferred to use. When a treatment shows improvement on all target outcomes, the GST often has a higher power than tests of single outcomes or other multiple test procedures. While the weakness of GST is they generally permit only global, not component-specific, conclusions, leading to difficulties in interpretation (19).

O'Brien proposed a nonparametric GST procedure, a rank-sum-type test, which is based on the rank of each individual outcome among the combined observations from two samples (20). It does not require a common treatment effect assumption and can be applied to outcomes measured in different scales (suppose X_{jv} is the observation of the v outcome from subject j in group i , let R_{jv} be the rank of X_{jv} , the patient's ranks for each outcome are summed, and then assess whether outcome measures from one group are consistently larger than outcome measures from the other group).

$$T = \frac{\bar{R}_2 - \bar{R}_1}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}, \bar{R}_i = \sum_{j=1}^{n_i} \sum_{v=1}^k R_{jv} / n_i \quad [1]$$

A unified interpretation of nonparametric GST can be provided through the use of global treatment effect (GTE). GTE is defined as an average of probabilities of treatment benefit on multiple outcomes; it plays a

similar role as the traditionally used effect size in study design (21). The interpretation of GTE is uniform; no matter what measurement scales are used, the GTE is unchanged. GTE takes value between -1 and 1 , when $GTE = 0$, there is no global preference between two groups, when $GTE = 1$, the treatment is most preferred, when $GTE = -1$, the treatment is least preferred. Larger positive GTE values correspond to higher degrees of treatment preference (22). The GST utilizing GTE can compare treatments based on a treatment's multidimensional performance and provide a single test for global interpretations on whether a new treatment should be advocated.

$$\theta = \sum_{v=1}^k \theta_v / k, \theta_v = (1/n_1 n_2) \left\{ 2 \sum_{j=1}^{n_1} R_{jv} - n_2(n_1 + n_2 + 1) \right\} \quad [2]$$

Alpha-adjustment procedures

Alpha-adjustment procedures are multiple tests with adjustment to the overall significance level (23). The advantage of these methods is that they can test whether there is any treatment difference on any single outcome and control the family-wise Type I error rate (FWER). While clinical interpretations can be difficult in the presence of multiple conflicting results, and the methods can't give a global assessment of a treatment's benefit on multiple outcomes, especially when treatment demonstrates both beneficial and detrimental effects on different outcomes (24).

A number of methods have been proposed to adjust significance levels for the analysis of multiple outcomes, including Bonferroni test, Simes, James and Hochberg procedures (25). Of all the methods in practice, Bonferroni test is the most well known and has strong appeal because of its ease of use (26), it is an approximate method based on the probability of obtaining a false positive and compares each single outcome's P value with the adjusted level of a/K rather than a , where K is the total number of outcomes.

The Bonferroni test assumes that the multiple outcomes are independent, it may suffer from poor statistical performance when outcomes are highly correlated, which is a major drawback of the Bonferroni procedure (27). While in practice, it is often not an

appropriate assumption, these multiple outcomes are usually correlated because they measured related quantities in the same patients. By ignoring these correlations, we will obtain a less precise estimate of the treatment effect.

Qian Shi introduced an adaptation of the Bonferroni procedure (28), a correction factor based on intraclass correlation (ICC) is applied to the Bonferroni test to account for the correlation of multiple outcomes, and this method can overcome the shortcomings of the standard Bonferroni adjustment yet maintains its advantages.

$$\alpha^* = \alpha / K^*, \quad K^* = (K + 1) - [1 + (K - 1)r] \quad [3]$$

For traditionally designed RCTs with multiple outcomes, the correlation of outcomes is always estimated according to clinical experience or published researches, maybe the estimation is often inappropriate in a certain extent, and the evaluation of treatment effect will be less precise. It is helpful to prespecify the correlation among multiple outcomes when we adopt the adaptive design adjustment method in the design of RCTs (29).

Using adaptive design to calculate correlation of multiple primary outcomes

An adaptive design is defined as a clinical trial design that uses accumulating data to decide on how to modify trial and/or statistical aspects of the study as it continues, without undermining the validity and integrity of the trial (30,31).

In a TCM clinical trial with multiple primary outcomes, the strategy of one stage adaptive design can be used, when all patients completed the trial and data collection, we can calculate ICC of multiple primary outcomes with blind adjustment (29), apply the adaptation of Bonferroni procedure introduced by Qian Shi (28) to adjust alpha of each outcome, then unblinding the data and do statistic analysis.

The adaptive design can resolve the inappropriate correlation estimation among multiple primary outcomes when started RCTs. In addition, the correlation calculation is under blindness and can control the FWER.

Establishing an evaluation mode with multiple primary outcomes based on combination of diseases and symptoms in TCM clinical trials

As we mentioned above, in TCM RCTs, the most commonly used outcome assessment method, separate analysis of each outcome, has been reported to have some limitations. So we introduced the combination evaluation of multiple primary outcomes including diseases and symptoms outcomes, which can reflect the efficacy of TCM comprehensively and objectively. We used one stage adaptive adjustment strategy to estimate the correlation among multiple primary outcomes in blindness, then introduced an adaptation of the Bonferroni procedure which accounts for correlated data to calculate the alpha of individual outcome. GST using the O'Brien ranking procedure and the corresponding GTE measure were used to assess the treatment's global impact.

The adaptive design can resolve the inappropriate correlation estimation among multiple primary outcomes. The nonparametric GST proposed by O'Brien can provide an overall test of multiple outcomes, with separate reports of individual outcome using an adaptation of the Bonferroni procedure, can provide useful additional information. We hope that the introduction of this approach will provide methodological aid for the assessment of holistic therapeutic effect of TCM.

Example

We will conduct a randomized, double-blind, controlled trial to explore establishing an evaluation mode with multiple primary outcomes based on combination of diseases and symptoms. The total target sample size is planned at 60 participants with stable angina, with a balanced (1:1) treatment allocation. In the intervention group, patients will take compound Danshen dripping pills plus simulated isosorbide dinitrate, patients in the control group will take isosorbide dinitrate plus simulated compound Danshen dripping pills. The treatment period for the trial drugs was 8 weeks. The primary outcomes will be electrocardiogram (ECG) efficiency, TCM syndrome score and quality of life (*Figure 1*).

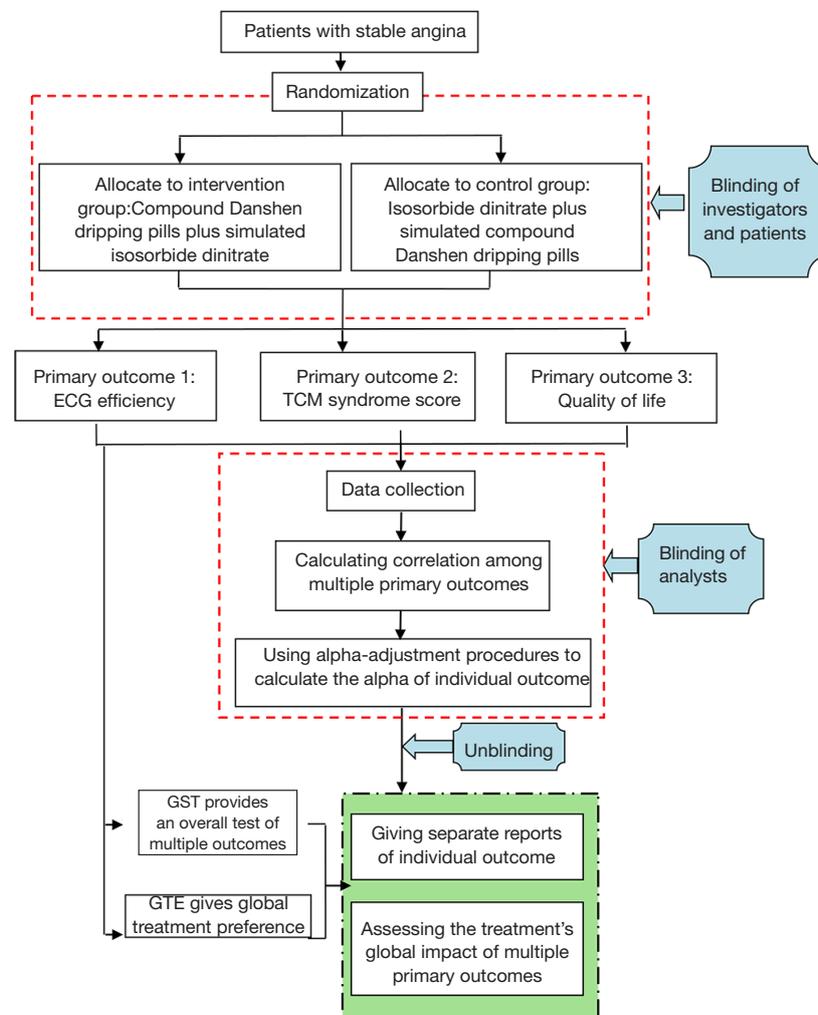


Figure 1 Flow chart of a clinical trial design with multiple primary outcomes based on combination of diseases and symptoms

Acknowledgements

Funding: The study was supported by the Natural Science Foundation of China Project (No. 81403283; No. 81273935). The funders had no role in study design, decision to publish, or preparation of the manuscript.

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

References

1. Shan P, Mao RB, Xu JM, et al. A doubled-blind clinical trial of 110 patients with Huangyangning for coronary heart disease. *J Tradit Chin Med* 1983;5:37-40.
2. Pritzker S, Hui KK. Building an Evidence-Base for TCM and Integrative East-West Medicine: A Review of Recent Developments in Innovative Research Design. *J Tradit Complement Med* 2012;2:158-63.
3. Ma Y, Zhou K, Fan J, et al. Traditional Chinese medicine: potential approaches from modern dynamical complexity theories. *Front Med* 2016;10:28-32.
4. Fung FY, Linn YC. Developing traditional chinese medicine in the era of evidence-based medicine: current evidences and challenges. *Evid Based Complement Alternat Med* 2015;2015:425037.
5. Hui KK, Zhang WJ. Innovative clinical scientific

- researches promotes the chinese medicine development grounding on globalization. *Zhongguo Zhong Xi Yi Jie He Za Zhi* 2010;30:789-92.
- 6 ICH harmonised tripartite guideline: statistical principles for clinical trials E9. Available online: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf
 - 7 Shang HC, Zhang BL, Li YP. Thinking and methods in practical assessment of TCM chinal therapeutic effect. *Zhongguo Zhong Xi Yi Jie He Za Zhi* 2008;28:266-8.
 - 8 Zhang L, Zhang J, Chen J, et al. Clinical research of traditional Chinese medicine needs to develop its own system of core outcome sets. *Evid Based Complement Alternat Med* 2013;2013:202703.
 - 9 Pocock SJ, Geller NL, Tsiatis AA. The analysis of multiple endpoints in clinical trials. *Biometrics* 1987;43:487-98.
 - 10 Tukey JW. Some thoughts on clinical trials, especially problems of multiplicity. *Science* 1977;198:679-84.
 - 11 Ludbrook J. Multiple comparison procedures updated. *Clin Exp Pharmacol Physiol* 1998;25:1032-7.
 - 12 Wang XL, Mao JY, Hou YZ. Preliminary study of establishing clinical effect evaluation methods of Chinese Medicine based on combination of disease and syndrome, systematic staging, and multi-dimension index. *Zhongguo Zhong Xi Yi Jie He Za Zhi* 2013;33:270-3.
 - 13 Gao FZ, Xie YM, Wang YY. Complex intervention and comprehensive evaluation of Traditional Chinese Medicine. *Chinese Journal of Basic Medicine in Traditional Chinese Medicine* 2010;16:527-9.
 - 14 Li JS, Yu XQ. Thinking on development of curative effect evaluation index system based on mode of combination of disease and TCM syndrome. *China Journal of Traditional Chinese Medicine and Pharmacy* 2011;26:1666-70.
 - 15 Neuhäuser M. How to deal with multiple endpoints in clinical trials. *Fundam Clin Pharmacol* 2006;20:515-23.
 - 16 Rauch G, Jahn-Eimermacher A, Brannath W, et al. Opportunities and challenges of combined effect measures based on prioritized outcomes. *Stat Med* 2014;33:1104-20.
 - 17 Pocock SJ. Clinical trials with multiple outcomes: a statistical perspective on their design, analysis, and interpretation. *Control Clin Trials* 1997;18:530-45.
 - 18 Lefkopoulou M, Ryan L. Global tests for multiple binary outcomes. *Biometrics* 1993;49:975-88.
 - 19 Baraniuk S, Seay R, Sinha AK, et al. Comparison of the global statistical test and composite outcome for secondary analyses of multiple coronary heart disease outcomes. *Prog Cardiovasc Dis* 2012;54:357-61.
 - 20 O'Brien PC. Procedures for comparing samples with multiple endpoints. *Biometrics* 1984;40:1079-87.
 - 21 Huang P, Woolson RF, O'Brien PC. A rank-based sample size method for multiple outcomes in clinical trials. *Stat Med* 2008;27:3084-104.
 - 22 Huang P, Goetz CG, Woolson RF, et al. Using global statistical tests in long-term Parkinson's disease clinical trials. *Mov Disord* 2009;24:1732-9.
 - 23 Sankoh AJ, D'Agostino RB Sr, Huque MF. Efficacy endpoint selection and multiplicity adjustment methods in clinical trials with inherent multiple endpoint issues. *Stat Med* 2003;22:3133-50.
 - 24 Alosch M, Bretz F, Huque M. Advanced multiplicity adjustment methods in clinical trials. *Stat Med* 2014;33:693-713.
 - 25 Sankoh AJ, Huque MF, Dubey SD. Some comments on frequently used multiple endpoint adjustment methods in clinical trials. *Stat Med* 1997;16:2529-42.
 - 26 Hochberg Y, Tamhane AC. Multiple comparison procedures. New York: Wiley; 1987:72-109.
 - 27 Leon AC, Heo M. A comparison of multiplicity adjustment strategies for correlated binary endpoints. *J Biopharm Stat* 2005;15:839-55.
 - 28 Shi Q, Pavey ES, Carter RE. Bonferroni-based correction factor for multiple, correlated endpoints. *Pharm Stat* 2012;11:300-9.
 - 29 Wang L. The study of multiplicity adjustment in clinical trials with correlated endpoints and its application. Doctoral dissertation. Fourth Military Medical University 2011. Available online: <http://www.doc88.com/p-9753784907568.html>
 - 30 Chow SC, Chang M. Adaptive design methods in clinical trials-a review. *Orphanet J Rare Dis* 2008;3:11.
 - 31 Mi MY, Betensky RA. An analysis of adaptive design variations on the sequential parallel comparison design for clinical trials. *Clin Trials* 2013;10:207-15.

Cite this article as: Hu J, Liu S, Liu W, Zhang H, Chen J, Shang H. Establishing an evaluation mode with multiple primary outcomes based on combination of diseases and symptoms in TCM clinical trials. *Ann Transl Med* 2017;5(21):420. doi: 10.21037/atm.2017.06.06