JOURNAL OF BIOPHARMACEUTICAL STATISTICS
https://doi.org/10.1080/10543406.2020.1726368

Taylor & Francis
Taylor & Francis Group

Check for updates

Unified approaches to assessing treatment effect of traditional Chinese medicine based on health profiles

Bin Cheng[a], Bingzhi Zhang[b], and Shein-Chung Chow[c]

[a]Department of Biostatistics, Columbia University, New York, New York, USA; [b]Department of Biostatistics and Programming, Sanofi, Bridgewater, New Jersey, USA; [c]Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, North Carolina, USA

ABSTRACT

Two dissimilarity indices are introduced to measure the disharmony of a human body system by mimicking the population bioequivalence and the individual bioequivalence concepts. Hypotheses for the treatment effect of a traditional Chinese medicine are formulated based on the two indices and then tested under the proposed designs by reverting an approximate confidence upper bound. The proposed methods can also be used when a drug product has multiple components or a trial has multiple endpoints.

ARTICLE HISTORY
Received 8 February 2018
Accepted 27 January 2020

KEYWORDS
Profile analysis; principal component analysis; bioequivalence; biosimilarity; multiple endpoints

1. Introduction

In recent years, as more and more innovative drug products are going off patent protection, the search for new medicines that treat critical and/or life-threatening diseases such as cardiovascular diseases and cancer has become the center of attention of many pharmaceutical companies and research organizations such as National Institute of Health (NIH). This leads to the study of the potential use of promising traditional Chinese medicines (TCM), especially for critical and/or life-threatening diseases. Bensoussan et al. (1998) used randomized clinical trial (RCT) to assess the effect of Chinese herb medicine in treating the Irritable Bowel Syndrome. However, RCT is not in common use when studying TCM. There are fundamental differences between Western medicines and TCM in terms of diagnostic procedures, therapeutic indices, medical mechanism, medical theory, and practice (Chow et al. 2006; Chow 2015; Zhou et al. 2012). Besides, TCM often consists of multiple components with flexible dose.

Chinese doctors believe that all of the organs within a healthy subject should reach the so-called *global dynamic balance and harmony* among organs. Once the global balance is broken at certain sites such as heart, liver, or kidney, some signs and symptoms will appear to reflect the imbalance at these sites. The collective signs and symptoms are then used to determine what disease that the individual has. An experienced Chinese doctor usually assesses the causes of global imbalance before a TCM with flexible doses is prescribed to fix the problem. This approach is sometimes referred to as a personalized (or individualized) medicine approach. In practice, TCM consider inspection, auscultation and olfaction, interrogation, and pulse taking and palpation as the primary diagnostic procedures. The scientific validity of these subjective and experience-based diagnostic procedures has been criticized due to lack of reference standards and anticipated large evaluator-to-evaluator (i.e., Chinese doctor-to-Chinese doctor) variability. For a systematic discussion of the statistical issues of TCM, see Chow (2015).

In this paper, we attempt to propose a unified approach to developing two composite dissimilarity indices under the concept of global dynamic balance among organs. Dynamic balance among organs are formulated as hypotheses using the dissimilarity indices. Following the concept of testing

CONTACT Bin Cheng bc2159@cumc.columbia.edu Department of Biostatistics, Columbia University, New York, NY 10032
© 2020 Taylor & Francis

bioequivalence or biosimilarity, if the 95% confidence upper bound is less than some health limit, we conclude that the treatment achieves dynamic balance among the organs of the subject hence is considered as efficacious. If we fail to reject the null hypothesis, we conclude that the treatment is not efficacious since there is still a signal of illness, i.e., some of signs and symptoms are still out of the health limit. Although developed with TCM applications in mind, the proposed methods can be utilized in any clinical trials where multiple drug components and/or multiple endpoints are involved.

The article is organized as follows. Two approaches are proposed to assessing the effect of a TCM. Section 2 introduces the population similarity approach where the population dissimilarity index is defined, and a population similarity test is constructed. Section 3 defines the individual dissimilarity and based on which an individual similarity test is constructed. The performances of the approximate tests are evaluated via simulation studies and reported in Section 4, followed by a brief discussion in Section 5.

2. Population similarity approach

TCM cares about the harmony among different parts of the human body. The measure of these different parts consists of the *health profile* of a subject, denoted as $\boldsymbol{X} = (X_1, \ldots, X_k)$. The components X_i of a health profile could be either continuous or ordinal. The dimension of the health profile k is potentially high.

One approach to assessing the treatment effect of a TCM is to enroll a group of patients who will undergo the TCM treatment, follow them for a prespecified period of time, and compare the health profiles of these patients at the end of the study with those of a group of healthy subjects, possibly matched by age, gender, and other characteristics. If the health profiles of the two groups are reasonably close, then the TCM is considered as efficacious. This approach is called the *population similarity approach* as it involves both the treatment group and a healthy control group, and is often used in the development of precision medicines.

2.1. *Population dissimilarity index*

Let $\boldsymbol{X}_T = (X_{T1}, \ldots, X_{Tk})$ and $\boldsymbol{X}_H = (X_{H1}, \ldots, X_{Hk})$ be the k-dimensional health profiles of a subject who has undergone treatment of TCM and of a subject who is a healthy control, respectively. Under this formulation, a TCM treatment is considered efficacious if the health profile of subjects who receive treatment is not significantly different from the one of a healthy subject, possibly age and gender matched if needed. To this end, we define an index θ as a pseudo distance between the profile of a treated subject and the profile of a healthy subject. Specifically, let $\boldsymbol{\mu}_T$ and $\boldsymbol{\mu}_H$ be the mean health profiles of a treated subject and a healthy subject, respectively, and let $\boldsymbol{\Sigma}_T$ and $\boldsymbol{\Sigma}_H$ be the covariance matrices of a treated subject and a healthy subject, respectively. Let $\lambda_1(\boldsymbol{\Sigma})$ be the largest eigenvalue of a symmetric matrix $\boldsymbol{\Sigma}$. We define the *population dissimilarity index* as

$$\theta = \frac{(\boldsymbol{\mu}_T - \boldsymbol{\mu}_H)^T(\boldsymbol{\mu}_T - \boldsymbol{\mu}_H) + \lambda_1(\boldsymbol{\Sigma}_T) - \lambda_1(\boldsymbol{\Sigma}_H)}{\max\{\sigma_0^2, \lambda_1(\boldsymbol{\Sigma}_H)\}}, \tag{1}$$

where σ_0^2 is a known constant.

We point out that the above index is invariant under scale change, a prerequisite for an index to be reasonable. In addition, the above dissimilarity index has an interesting connection with principal component analysis. To see this, image that $\boldsymbol{X}_T$ and $\boldsymbol{X}_H$ are multivariate normal. Then, $(\boldsymbol{\mu}_T - \boldsymbol{\mu}_H)^T(\boldsymbol{\mu}_T - \boldsymbol{\mu}_H)$ is the squared mean of the first principal component of $\boldsymbol{X}_T - \boldsymbol{X}_H$, $\lambda_1(\boldsymbol{\Sigma}_T)$ is the variance of the first principal component of $\boldsymbol{X}_T$, and $\lambda_1(\boldsymbol{\Sigma}_H)$ is the variance of the first principal component of $\boldsymbol{X}_H$. A small θ implies two things. First, it implies a small difference between the mean health profiles $\boldsymbol{\mu}_T$ and $\boldsymbol{\mu}_H$, relative to the maximal variance of a healthy profile X_H. Second, it implies that the maximal variance of the diseased profile

$\boldsymbol{X}_T$ is at least not much greater than the maximal variance of the healthy profile $\boldsymbol{X}_H$. Therefore, a small dissimilarity index can be viewed as evidence for efficacy of a treatment. Chervoneva, Hyslop, and Hauck (2007) propose a testing procedure on multivariate population bioequivalence using trace of the covariance instead of the largest eigenvalue as we do.

The efficacy of a treatment could be formulated as a test for equivalence with the health profile of a healthy control. Let X_T denote the health profile of a diseased subject at the completion of the treatment. The efficacy is claimed if a test rejects the following H_0

$$H_0 : \theta \geq \epsilon \quad \text{versus} \quad H_1 : \theta < \epsilon, \tag{2}$$

where ϵ is a given threshold.

2.2. *Assessing treatment effect through population dissimilarity index*

Let

$$\gamma = (\boldsymbol{\mu}_T - \boldsymbol{\mu}_H)^T(\boldsymbol{\mu}_T - \boldsymbol{\mu}_H) + \lambda_1(\boldsymbol{\Sigma}_T) - \lambda_1(\boldsymbol{\Sigma}_H) - \epsilon \max\{\sigma_0^2, \lambda_1(\boldsymbol{\Sigma}_H)\}. \tag{3}$$

Then, testing (2) is equivalent to testing

$$H_0 : \gamma \geq 0, \quad \text{versus} \quad H_1 : \gamma < 0. \tag{4}$$

We construct an asymptotic test via the duality between hypothesis test and confidence interval. Specifically, we construct a 95% approximate confidence upper bound for γ based on two independent random samples $\boldsymbol{X}_{Ti} = (X_{T1i}, \ldots, X_{Tki})^T, i = 1, \ldots, n_T$, $\mathbf{X}_{Hj} = (X_{H1j}, \ldots, X_{Hkj})^T, j = 1, \ldots, n_H$, and reject the H_0 if this 95% confidence upper bound for γ is smaller than 0.

Let $\boldsymbol{B}$ be a $k \times k$ orthogonal matrix such that

$$\boldsymbol{B}^T(q\boldsymbol{\Sigma}_T + \mathbf{S}_H)\boldsymbol{B} = \text{diag}\{\eta_1, \ldots, \eta_k\},$$

where $q = n_H/n_T$ and define

$$v = (v_1, \ldots, v_k)^T = \boldsymbol{B}(\boldsymbol{\mu}_T - \boldsymbol{\mu}_H).$$

Let $\hat{\boldsymbol{\mu}}_a$ and $\widehat{\boldsymbol{\Sigma}}_a$ be the sample means and sample covariances for $a = T, H$, respectively. Let $\hat{B}$ be a $k \times k$ orthogonal matrix such that

$$\hat{B}^T\left(q\widehat{\boldsymbol{\Sigma}}_T + \widehat{\boldsymbol{\Sigma}}_H\right)\hat{B} = \text{diag}\{\hat{\eta}_1, \ldots, \hat{\eta}_k\}.$$

Define

$$\widehat{v} = (\hat{v}_1, \ldots, \hat{v}_k) = \hat{B}(\hat{\boldsymbol{\mu}}_T - \hat{\boldsymbol{\mu}}_H).$$

Then, parameter γ can be rewritten as

$$\gamma = \sum_{i=1}^{k} v_i^2 + \lambda_1(\boldsymbol{\Sigma}_T) - \lambda_1(\boldsymbol{\Sigma}_H) - \epsilon \max\{\sigma_0^2, \lambda_1(\boldsymbol{\Sigma}_H)\}.$$

It is easily seen that as $n_H \to \infty$, $\hat{v}_i$, $i = 1, \ldots, k$, are asymptotically independent and normally distributed as $N(v_i, \eta_i)$. A 95% confidence upper bound for $\sum_{i=1}^{k} v_i^2$ is

$$\sum_{i=1}^{k} \hat{v}_i^2 + \sqrt{\sum_{i=1}^{k}\left[\left(|\hat{v}_i| + z_{0.05}\sqrt{\frac{\hat{\eta}_i}{n_H}}\right)^2 - \hat{v}_i^2\right]^2},$$

where $z_{0.05}$ is the 95%th percentile of the standard normal distribution.

Let $l_{1,T}$ be the largest eigenvalue of $\widehat{\mathbf{\Sigma}}_T$. Then, by Anderson (2003), an asymptotic 95% confidence upper bound for $\lambda_1(\mathbf{\Sigma}_T)$ is

$$\frac{l_{1,T}}{1 - z_{0.05}\sqrt{2/n_T}}.$$

Similarly, an asymptotic 95% confidence lower bound for $\lambda_1(\mathbf{\Sigma}_H)$ is

$$\frac{l_{1,H}}{1 + z_{0.05}\sqrt{2/n_H}}.$$

If $\lambda_1(\mathbf{\Sigma}_H) \geq \sigma_0^2$, then γ in (3) reduces to

$$\gamma = \sum_{i=1}^{k} v_i^2 + \lambda_1(\mathbf{\Sigma}_T) - (1+\epsilon)\lambda_1(\mathbf{\Sigma}_H). \tag{5}$$

Since $\hat{v}_i$'s, $l_{1,T}$, and $l_{1,H}$ are independent, then, using the idea in Howe (1974) and Graybill and Wang (1980), we construct an approximate 95% confidence upper bound for γ as

$$\hat{\gamma}_{U,1} = \sum_{i=1}^{k} \hat{v}_i^2 + l_{1,T} - (1+\epsilon)l_{1,H} + \sqrt{\Delta_1}, \tag{6}$$

where

$$\Delta_1 = \sum_{i=1}^{k}\left[\left(|\hat{v}_i| + z_{0.05}\sqrt{\frac{\hat{\eta}_i}{n_H}}\right)^2 - \hat{v}_i^2\right]^2 + \left(\frac{l_{1,T}}{1-z_{0.05}\sqrt{2/n_T}} - l_{1,T}\right)^2 + (1+\epsilon)^2\left(\frac{l_{1,H}}{1+z_{0.05}\sqrt{2/n_H}} - l_{1,H}\right)^2.$$

If $\lambda_1(\mathbf{\Sigma}_H) < \sigma_0^2$, then γ in (3) reduces to

$$\gamma = \delta + \lambda_1(\mathbf{\Sigma}_T) - \lambda_1(\mathbf{\Sigma}_H) - \epsilon\sigma_0^2. \tag{7}$$

An approximate 95% confidence upper bound for γ is

$$\hat{\gamma}_{U,2} = \sum_{i=1}^{k} \hat{v}_i^2 + l_{1,T} - l_{1,H} - \epsilon\sigma_0^2 + \sqrt{\Delta_2}, \tag{8}$$

where

$$\Delta_2 = \sum_{i=1}^{k}\left[\left(|\hat{v}_i| + z_{0.05}\sqrt{\frac{\hat{\eta}_i}{n_H}}\right)^2 - \hat{v}_i^2\right]^2 + \left(\frac{l_{1,T}}{1-z_{0.05}\sqrt{2/n_T}} - l_{1,T}\right)^2 + \left(\frac{l_{1,H}}{1+z_{0.05}\sqrt{2/n_H}} - l_{1,H}\right)^2.$$

To test (4) or (2) at level 0.05, we reject H_0 if and only if $l_{1,H} \geq \sigma_0^2$ and $\hat{\gamma}_{U,1} < 0$, or $l_{1,H} < \sigma_0^2$ and $\hat{\gamma}_{U,2} < 0$. Specifically, we claim that testing rule

$$\phi = I(\hat{\gamma}_{U,1} < 0, l_{1,H} \geq \sigma_0^2) + I(\hat{\gamma}_{U,2} < 0, l_{1,H} < \sigma_0^2) \tag{9}$$

has an approximate level of 0.05 for testing H_0. This test will be referred to as the *population similarity test* subsequently in this paper.

3. Individual similarity approach

3.1. *Individual dissimilarity index*

The approach we adopt in the previous section is a population bioequivalence approach. It is therefore most relevant when the main purpose is to establish treatment effect of a TCM at the population level. It is of interest to develop methods to assess the TCM effect when no healthy controls are required or available. In such cases, each subject acts as his or her own control by comparing with the health profiles when he or she was in good health. This approach is called the *individual similarity approach*. Since it is often true that the between subject variations are large in TCM studies, the individual similarity approach is particularly appealing in the development of individualized (or personalized) TCM.

Specifically, consider a design where each patient i has three healthy profiles $\boldsymbol{X}_{Hi} = (X_{H1i}, \ldots, X_{Hki})$, $\boldsymbol{X}_{H'i} = (X_{H'1i}, \ldots, X_{H'ki})$, and $\boldsymbol{X}_{H''i} = (X_{H''1i}, \ldots, X_{H''ki})$, measured at three different time points prior to the TCM treatment when the subject is known as in good health, and one post treatment profile $\boldsymbol{X}_{Ti} = (X_{T1i}, \ldots, X_{Tki})$ measured at the completion of treatment, $i = 1, \ldots, n$. We consider the following random effects model

$$\boldsymbol{X}_{ai} = \boldsymbol{E}_i + \boldsymbol{W}_{ai}, a = H, H', H'', T,$$

where $\boldsymbol{E}_i, i = 1, \ldots, n$, are the independent and normally distributed subject random effects with zero mean and covariance $\boldsymbol{\Sigma}_E$, $\boldsymbol{W}_{Hi}, \boldsymbol{W}_{H'i}, \boldsymbol{W}_{H''i}, i = 1, \ldots, n$, are independent and normally distributed random vectors with mean $\boldsymbol{\mu}_H$ and covariance $\boldsymbol{\Sigma}_H$, and $\boldsymbol{W}_{Ti}, i = 1, \ldots, n$, are independent and normally distributed random vectors with mean $\boldsymbol{\mu}_T$ and covariance $\boldsymbol{\Sigma}_T$.

Mimicking the conception of individual bioequivalence, we define the *individual dissimilarity index* as

$$\vartheta = \frac{(\boldsymbol{\mu}_T - \boldsymbol{\mu}_H)^T(\boldsymbol{\mu}_T - \boldsymbol{\mu}_H) + \lambda_1(\boldsymbol{\Sigma}_T + \boldsymbol{\Sigma}_H) - 2\lambda_1(\boldsymbol{\Sigma}_H)}{\max\{\sigma_0^2, \lambda_1(\boldsymbol{\Sigma}_H)\}},$$

and consider testing the following

$$H_0 : \vartheta \geq \varepsilon \quad \text{versus} \quad H_1 : \vartheta < \varepsilon, \tag{10}$$

where $\varepsilon > 0$ is a prespecified threshold. Individual efficacy of the TCM is established if H_1 in (10) is rejected.

3.2. *Assessing treatment effect through individual dissimilarity index*

Define

$$\xi = (\boldsymbol{\mu}_T - \boldsymbol{\mu}_H)^T(\boldsymbol{\mu}_T - \boldsymbol{\mu}_H) + \lambda_1(\boldsymbol{\Sigma}_T + \boldsymbol{\Sigma}_H) - 2\lambda_1(\boldsymbol{\Sigma}_H) - \varepsilon \max\{\sigma_0^2, \lambda_1(\boldsymbol{\Sigma}_H)\}. \tag{11}$$

Then, testing (10) is equivalent to testing

$$H_0 : \xi \geq 0, \quad \text{versus} \quad H_1 : \xi < 0. \tag{12}$$

Let $\boldsymbol{\mu}_T \widehat{-} \boldsymbol{\mu}_H$ and $\boldsymbol{\Sigma}_T \widehat{+} \boldsymbol{\Sigma}_H$ be the sample mean and sample variance of $\boldsymbol{X}_{Ti} - \boldsymbol{X}_{Hi}, i = 1, \ldots, n$. Let $\widehat{\boldsymbol{C}}$ be a $k \times k$ orthogonal matrix such that

$$\hat{C}^T\left(\boldsymbol{\Sigma}_T \widehat{+} \boldsymbol{\Sigma}_H\right)\hat{C} = \text{diag}\{\hat{\zeta}_1, \ldots, \hat{\zeta}_k\},$$

and

$$\hat{\phi} = (\hat{\phi}_1, \ldots, \hat{\phi}_k) = \hat{C}(\boldsymbol{\mu}_T \widehat{-} \boldsymbol{\mu}_H).$$

Similarly, let C be a $k \times k$ orthogonal matrix such that

$$C^T(\boldsymbol{\Sigma}_T + \boldsymbol{\Sigma}_H)C = \text{diag}\{\zeta_1, \ldots, \zeta_k\},$$

and define

$$\phi = (\phi_1, \ldots, \phi_k) = \boldsymbol{C}(\boldsymbol{\mu}_T - \boldsymbol{\mu}_H).$$

Rewrite ξ as

$$\xi = \sum_{i=1}^{k} \phi_i^2 + \lambda_1(\boldsymbol{\Sigma}_T + \boldsymbol{\Sigma}_H) - 2\lambda_1(\boldsymbol{\Sigma}_H) - \varepsilon \max\{\sigma_0^2, \lambda_1(\boldsymbol{\Sigma}_H)\}.$$

It is easily seen that $\hat{\phi}_i$, $i = 1, \ldots, k$, are asymptotically independent and normally distributed as $N(\phi_i, \zeta_i)$. Then, a 95% confidence upper bound for $\sum_{i=1}^{k} \phi_i^2$ is

$$\sum_{i=1}^{k} \hat{\phi}_i^2 + \sqrt{\sum_{i=1}^{k} \left[\left(|\hat{\phi}_i| + z_{0.05} \sqrt{\frac{\hat{\zeta}_i}{n_H}} \right)^2 - \hat{\phi}_i^2 \right]^2},$$

where $z_{0.05}$ is the 95%th percentile of the standard normal distribution.

Let $l_{1,TH} = \max_i\{\hat{\zeta}_i\}$ be the largest eigenvalue of $\widehat{\boldsymbol{\Sigma}_T + \boldsymbol{\Sigma}_H}$. Then, an asymptotic 95% confidence upper bound for $\lambda_1(\boldsymbol{\Sigma}_T + \boldsymbol{\Sigma}_H)$ is

$$\frac{l_{1,TH}}{1 - z_{0.05}\sqrt{2/n}}.$$

Similarly, an asymptotic 95% confidence lower bound for $\lambda_1(\boldsymbol{\Sigma}_H)$ based on $X_{H'i} - X_{H''i}, i = 1, \ldots, n$, is

$$\frac{l_{1,H}}{1 + z_{0.05}\sqrt{2/n_H}}.$$

If $\lambda_1(\boldsymbol{\Sigma}_H) \geq \sigma_0^2$, then ξ reduces to

$$\xi = \sum_{i=1}^{k} \phi_i^2 + \lambda_1(\boldsymbol{\Sigma}_T + \boldsymbol{\Sigma}_H) - (2 + \varepsilon)\lambda_1(\boldsymbol{\Sigma}_H). \tag{13}$$

Since $\hat{\phi}_i$'s, $l_{1,TH}$, and $l_{1,H}$ are independent, we construct an approximate 95% confidence upper bound for ξ as

$$\hat{\xi}_{U,1} = \sum_{i=1}^{k} \hat{\phi}_i^2 + l_{1,TH} - (2 + \varepsilon)l_{1,H} + \sqrt{\Psi_1}, \tag{14}$$

where

$$\Psi_1 = \sum_{i=1}^{k} \left[\left(|\hat{\phi}_i| + z_{0.05}\sqrt{\tfrac{\hat{\zeta}_i}{n_H}} \right)^2 - \hat{\phi}_i^2 \right]^2 + \left(\frac{l_{1,TH}}{1 - z_{0.05}\sqrt{2/n}} - l_{1,TH} \right)^2 + (2 + \varepsilon)^2 \left(\frac{l_{1,H}}{1 + z_{0.05}\sqrt{2/n}} - l_{1,H} \right)^2.$$

If $\lambda_1(\boldsymbol{\Sigma}_H) < \sigma_0^2$, then ξ reduces to

$$\xi = \sum_{i=1}^{k} \phi_i^2 + \lambda_1(\boldsymbol{\Sigma}_T + \boldsymbol{\Sigma}_H) - 2\lambda_1(\boldsymbol{\Sigma}_H) - \varepsilon\sigma_0^2. \tag{15}$$

An approximate 95% confidence upper bound for ξ is

$$\hat{\xi}_{U,2} = \sum_{i=1}^{k} \hat{\phi}_i^2 + l_{1,TH} - 2l_{1,H} - \varepsilon\sigma_0^2 + \sqrt{\Psi_2}, \tag{16}$$

where

$$\Psi_2 = \sum_{i=1}^{k} \left[\left(|\hat{\phi}_i| + z_{0.05}\sqrt{\frac{\hat{\zeta}_i}{n_H}} \right)^2 - \hat{\phi}_i^2 \right]^2 + \left(\frac{l_{1,TH}}{1 - z_{0.05}\sqrt{2/n}} - l_{1,TH} \right)^2 + \left(\frac{l_{1,H}}{1 + z_{0.05}\sqrt{2/n}} - l_{1,H} \right)^2.$$

We propose to reject H_0 in (12) or (10) at level 0.05. If and only if $l_{1,H} \geq \sigma_0^2$ and $\hat{\xi}_{U,1} < 0$, or $l_{1,H} < \sigma_0^2$ and $\hat{\xi}_{U,2} < 0$. This test will be referred to as the *individual similarity test* subsequently in this paper.

4. Simulation

It is easily seen that the constructions of the two tests in Sections 2.2 and 3.2 are very similar. Therefore, without loss of generality, we only conduct simulation for the population similarity test in Section 2.2.

In our simulation, we set $k = 10$, and since γ depends on $\boldsymbol{\mu}_T$ and $\boldsymbol{\mu}_H$ only through their difference $\boldsymbol{\mu}_T - \boldsymbol{\mu}_H$, we assume

$$\boldsymbol{\mu}_T = (a, ar, ar^2, \ldots, ar^9), \boldsymbol{\mu}_H = (0, 0, 0, \ldots, 0),$$

where $a \neq 0, r \geq 0$. For the covariance matrix, we set

$$\boldsymbol{\Sigma}_T = \begin{pmatrix} \sigma_1 \boldsymbol{I_5} & \mathbf{0} \\ \mathbf{0} & \sigma_2 \boldsymbol{I_5} \end{pmatrix} \cdot \boldsymbol{R}_T \cdot \begin{pmatrix} \sigma_1 \boldsymbol{I_5} & \mathbf{0} \\ \mathbf{0} & \sigma_2 \boldsymbol{I_5} \end{pmatrix},$$

where $\boldsymbol{I}_5$ is a 5×5 identity matrix, σ_1 is the standard deviation of the first 5 components of $\boldsymbol{X}_T$ and σ_2 is the standard deviation of the last 5 components of $\boldsymbol{X}_T$, and $\boldsymbol{R}_T$ is the 10×10 correlation matrix. Similarly, we set

$$\boldsymbol{\Sigma}_H = \begin{pmatrix} \tau_1 \boldsymbol{I_5} & \mathbf{0} \\ \mathbf{0} & \tau_2 \boldsymbol{I_5} \end{pmatrix} \cdot \boldsymbol{R}_H \cdot \begin{pmatrix} \tau_1 \boldsymbol{I_5} & \mathbf{0} \\ \mathbf{0} & \tau_2 \boldsymbol{I_5} \end{pmatrix},$$

where τ_1 is the standard deviation of the first 5 components of $\boldsymbol{X}_H$ and τ_2 is the standard deviation of the last 5 components of $\boldsymbol{X}_H$, and $\boldsymbol{R}_H$ is the 10×10 correlation matrix.

For illustration, we choose $\epsilon = 2.0, \sigma_0 = 0.5$. Two types of correlation patterns are considered: compound symmetric (CS) correlation and the autoregressive of order 1 (AR(1)) correlation, where the parameter for $\boldsymbol{R}_T$ and $\boldsymbol{R}_H$ are ρ_T and ρ_H, respectively. All estimations are based on 10,000 simulation runs and $z_{0.025}$ is used in the test.

The simulated sizes of the test under CS and AR(1) correlations are summarized in Tables 1 and 2, respectively. The size is controlled in all the CS and AR(1) scenarios.

The simulated powers of the test under CS and AR(1) correlations are summarized in Tables 3 and 4, respectively. A balance design, i.e., $n_T = n_H$, typically yield a higher power than the case where more healthy controls are used, i.e., $n_H > n_T$. The simulation indicates that for an effect size of $\gamma = -2.00$, a sample size of 120 in each group would be adequate to achieve 80% power.

In Tables 5 and 6, more scenarios were considered to investigate how the γ value changes and what is the impact on the power under various total sample sizes and ratio q choices.

Table 1. Estimated sizes based on 10,000 simulation runs, assuming $\epsilon = 2.0, \sigma_0 = 0.5$ and compound symmetric correlation.

(a, r)	$(\sigma 1, \sigma 2, \rho 1)$	$(\tau 1, \tau 2, \rho 2)$	γ	(n_T, n_H)	Size
(1.45, 0.25)	(1.10, 1.00, 0.10)	(1.00, 1.00, 0.05)	0.00	(50, 100)	0.0019
(1.45, 0.25)	(1.10, 1.00, 0.10)	(1.00, 1.00, 0.05)	*0.00*	(50, 100)	0.0019
(1.45, 0.25)	(1.10, 1.00, 0.10)	(1.00, 1.00, 0.05)	0.00	(75, 75)	0.0116
(1.45, 0.25)	(1.10, 1.00, 0.10)	(1.00, 1.00, 0.05)	0.00	(80, 160)	0.0018
(1.45, 0.25)	(1.10, 1.00, 0.10)	(1.00, 1.00, 0.05)	0.00	(120, 120)	0.0086
(0.00, 0.00)	(1.71, 1.00, 0.10)	(1.00, 1.00, 0.05)	0.00	(50, 100)	0.0001
(0.00, 0.00)	(1.71, 1.00, 0.10)	(1.00, 1.00, 0.05)	0.00	(75, 75)	0.0017
(0.00, 0.00)	(1.71, 1.00, 0.10)	(1.00, 1.00, 0.05)	0.00	(80, 160)	0.0002
(0.00, 0.00)	(1.71, 1.00, 0.10)	(1.00, 1.00, 0.05)	0.00	(120, 120)	0.0027

Table 2. Estimated sizes based on 10,000 simulation runs, assuming $\epsilon = 2.0, \sigma_0 = 0.5$ and AR(1) correlation.

(a, r)	$(\sigma 1, \sigma 2, \rho 1)$	$(\tau 1, \tau 2, \rho 2)$	γ	(n_T, n_H)	Size
(1.45, 0.25)	(1.10, 1.00, 0.10)	(1.00, 1.00, 0.05)	0.00	(50, 100)	0.0019
(1.33, 0.24)	(1.10, 1.00, 0.10)	(1.00, 1.00, 0.05)	0.00	(50, 100)	0.0010
(1.33, 0.24)	(1.10, 1.00, 0.10)	(1.00, 1.00, 0.05)	0.00	(75, 75)	0.0180
(1.33, 0.24)	(1.10, 1.00, 0.10)	(1.00, 1.00, 0.05)	0.00	(80, 160)	0.0011
(1.33, 0.24)	(1.10, 1.00, 0.10)	(1.00, 1.00, 0.05)	0.00	(120, 120)	0.0143
(0.00, 0.00)	(1.67, 1.00, 0.10)	(1.00, 1.00, 0.05)	0.00	(50, 100)	0.0000
(0.00, 0.00)	(1.67, 1.00, 0.10)	(1.00, 1.00, 0.05)	0.00	(75, 75)	0.0005
(0.00, 0.00)	(1.67, 1.00, 0.10)	(1.00, 1.00, 0.05)	0.01	(80, 160)	0.0000
(0.00, 0.00)	(1.67, 1.00, 0.10)	(1.00, 1.00, 0.05)	0.01	(120, 120)	0.0016

Table 3. Estimated powers based on 10,000 simulation runs, assuming $\epsilon = 2.0, \sigma_0 = 0.5$ and compound symmetric correlation.

$(a, r$	$(\sigma 1, \sigma 2, \rho 1)$	$(\tau 1, \tau 2, \rho 2)$	γ	(nT, nH)	Power
(0.30, 0.08)	(1.10, 1.00, 0.10)	(1.00, 1.00, 0.05)	−1.99	(50, 100)	0.2895
(0.30, 0.08)	(1.10, 1.00, 0.10)	(1.00, 1.00, 0.05)	−1.99	(75, 75)	0.6309
(0.30, 0.08)	(1.10, 1.00, 0.10)	(1.00, 1.00, 0.05)	−1.99	(80, 160)	0.6193
(0.30, 0.08)	(1.10, 1.00, 0.10)	(1.00, 1.00, 0.05)	−1.99	(120, 120)	0.8378
(0.00, 0.00)	(1.19, 1.00, 0.10)	(1.00, 1.00, 0.05)	−2.00	(50, 100)	0.2399
(0.00, 0.00)	(1.19, 1.00, 0.10)	(1.00, 1.00, 0.05)	−2.00	(75, 75)	0.6075
(0.00, 0.00)	(1.19, 1.00, 0.10)	(1.00, 1.00, 0.05)	−2.00	(80, 160)	0.5957
(0.00, 0.00)	(1.19, 1.00, 0.10)	(1.00, 1.00, 0.05)	−2.00	(120, 120)	0.8486

Table 4. Estimated powers based on 10,000 simulation runs, assuming $\epsilon = 2.0, \sigma_0 = 0.5$ and AR(1) correlation.

$(a, r$	$(\sigma 1, \sigma 2, \rho 1)$	$(\tau 1, \tau 2, \rho 2)$	γ	(nT, nH)	Power
(0.00, 0.08)	(1.10, 1.00, 0.10)	(1.00, 1.00, 0.07)	−1.99	(50, 100)	0.5794
(0.00, 0.08)	(1.10, 1.00, 0.10)	(1.00, 1.00, 0.07)	−1.99	(75, 75)	0.9599
(0.00, 0.08)	(1.10, 1.00, 0.10)	(1.00, 1.00, 0.07)	−1.99	(80, 160)	0.9627
(0.00, 0.08)	(1.10, 1.00, 0.10)	(1.00, 1.00, 0.07)	−1.99	(120, 120)	0.9996
(0.00, 0.00)	(1.00, 1.00, 0.10)	(1.00, 1.00, 0.04)	−2.03	(50, 100)	0.8102
(0.00, 0.00)	(1.00, 1.00, 0.10)	(1.00, 1.00, 0.04)	−2.03	(75, 75)	0.9920
(0.00, 0.00)	(1.00, 1.00, 0.10)	(1.00, 1.00, 0.04)	−2.03	(80, 160)	0.9960
(0.00, 0.00)	(1.00, 1.00, 0.10)	(1.00, 1.00, 0.04)	−2.03	(120, 120)	0.9999

5. Discussion

Two approaches to assessing the treatment effect of a TCM have been introduced based on different bioequivalence concepts. Generally, the population similarity approach is recommended if the health profiles of healthy controls are easy to obtain and the main interest is to establish the TCM effect at the population level as in the case of a randomized clinical trial. Otherwise, if the primary interest is to establish the treatment effect at the individual level and it is feasible to obtain multiple historical

Table 5. Estimated powers based on 10,000 simulation runs, assuming $\epsilon = 2.0, \sigma_0 = 0.5$ and compound symmetric correlation.

(a, r)	$(\sigma 1, \sigma 2, \rho 1)$	$(\tau 1, \tau 2, \rho 2)$	γ	(nT, nH)	Power
(0.30, 0.09)	(1.10, 1.00, 0.10)	(1.00, 1.00, 0.05)	−1.82	(50, 100)	0.2175
(0.30, 0.09)	(1.10, 1.00, 0.10)	(1.00, 1.00, 0.05)	−1.82	(75, 75)	0.5270
(0.30, 0.09)	(1.10, 1.00, 0.10)	(1.00, 1.00, 0.05)	−1.82	(80, 160)	0.4991
(0.30, 0.09)	(1.10, 1.00, 0.10)	(1.00, 1.00, 0.05)	−1.82	(120, 120)	0.7345
(0.00, 0.00)	(0.30, 1.00, 0.10)	(1.00, 1.00, 0.05)	−2.93	(50, 100)	0.9890
(0.00, 0.00)	(0.30, 1.00, 0.10)	(1.00, 1.00, 0.05)	−2.00	(75, 75)	0.9993
(0.00, 0.00)	(0.30, 1.00, 0.10)	(1.00, 1.00, 0.05)	−2.00	(80, 160)	0.9999
(0.00, 0.00)	(0.30, 1.00, 0.10)	(1.00, 1.00, 0.05)	−2.00	(120, 120)	0.9999

Table 6. Estimated powers based on 10,000 simulation runs, assuming $\epsilon = 2.0, \sigma_0 = 0.5$ and AR(1) correlation.

(a, r)	$(\sigma 1, \sigma 2, \rho 1)$	$(\tau 1, \tau 2, \rho 2)$	γ	(nT, nH)	Power
(0.00, 0.09)	(1.10, 1.00, 0.10)	(1.00, 1.00, 0.07)	−1.99	(50, 100)	0.5794
(0.00, 0.09)	(1.10, 1.00, 0.10)	(1.00, 1.00, 0.07)	−1.99	(75, 75)	0.9599
(0.00, 0.09)	(1.10, 1.00, 0.10)	(1.00, 1.00, 0.07)	−1.99	(80, 160)	0.9627
(0.00, 0.09)	(1.10, 1.00, 0.10)	(1.00, 1.00, 0.07)	−1.99	(120, 120)	0.9996
(0.00, 0.00)	(1.00, 1.00, 0.10)	(1.00, 1.00, 0.05)	−1.97	(50, 100)	0.7594
(0.00, 0.00)	(1.00, 1.00, 0.10)	(1.00, 1.00, 0.05)	−1.97	(75, 75)	0.9878
(0.00, 0.00)	(1.00, 1.00, 0.10)	(1.00, 1.00, 0.05)	−1.97	(80, 160)	0.9895
(0.00, 0.00)	(1.00, 1.00, 0.10)	(1.00, 1.00, 0.05)	−1.97	(120, 120)	0.9999

health profiles when a patient was in good health, then the individual similarity approach is suggested.

The construction of the individual similarity test in Section 3.2 requires three healthy profiles and one post-TCM profile for each patient. This requirement is not necessary but it makes it easier to derive the test. When only one or two healthy profiles are available for each patient, it is difficult to derive a test for (10) and the bootstrap test may be utilized instead.

The health profiles are assumed to have multivariate normal distributions. If part or all the components of a health profile are not normally distributed, the proposed methods remain approximately valid as long as the sample sizes are large.

Our dissimilarity index focus on the largest eigenvalue of the covariance matrix, hence is associated with the principal component method. When the largest eigenvalue can only explain a fraction of the total variation, then the proposed method will suffer loss in power. Some other methods, for example, one similar to Chervoneva, Hyslop, and Hauck (2007), should be derived and utilized instead.

The sample size determination is quite complicated for bioequivalence test due to the fact that the null and alternative distributions are dependent on the unknown parameter and may differ under null and the alternative. Some recent work includes Chiang et al. (2014). Further research is justified on this topic.

In this article, the dimension of the health profile k is implicitly assumed to be not large compared with the sample size. If this is not the case, that is, if k is comparable to the sample size, then the methods for sparse principal component analysis, such as Zou et al. (2006) or Johnstone and Lu (2009), should be used and the corresponding 95% confidence upper bound could be constructed via the bootstrap method.

The dissimilarity indices defined in this paper measures the closeness between two health profiles using both mean and variance information. However, the causal relationship among the k components $X_1, \ldots, X_k$, are not modeled. Per TCM theory, alternative models incorporating causal information may be considered. Future research on this alternative approach is thus warranted.

Finally, we reemphasize that the proposed methods are applicable not only in TCM studies but also in any clinical trial where multiple drug components and/or multiple endpoints are involved.

References

Anderson, T. W. 2003. *An introduction to multivariate statistical analysis*. 3rd ed. New York, New York: Wiley.

Bensoussan, A., N. J. Talley, M. Hing, R. Menzies, A. Guo, and M. Ngu. 1998. Treatment of irritable bowel syndrome with Chinese herb medicine: A randomized controlled trial. *Journal of the American Medical Association* 280:1585–1589.

Chervoneva, I., T. Hyslop, and W. W. Hauk. 2007. A multivariate test for population bioequivalence. *Statistics in Medicine* 26:1208–1223. doi:10.1002/sim.2605.

Chiang, C., C. F. Hsiao, and J. P. Liu. 2014. Sample size determination for individual bioequivalence inference. *PloS One* 9:e109746. doi:10.1371/journal.pone.0109746.

Chow, S. C. 2015. *Quantitative methods for traditional Chinese medicine*. New York, New York: Chapman and Hall/CRC Press, Taylor & Francis.

Chow, S. C., A. Pong, and Y. W. Chang. 2006. On traditional Chinese medicine clinical trails. *Drug Information Journal* 40:395–406. doi:10.1177/216847900604000404.

Graybill, F., and C. M. Wang. 1980. Confidence intervals on nonnegative linear combinations of variances. *Journal of American Statistical Association* 75:869–873. doi:10.1080/01621459.1980.10477565.

Howe, W. G. 1974. Approximate confidence limits on the mean of $X+Y$ where X and Y are two tabled indepedent random variables. *Journal of American Statistical Association* 69:789–794.

Johnstone, I. M., and A. Y. Lu. 2009. On consistency and sparsity for principal components analysis in high dimensions. *Journal of the American Statistical Association* 104:682–693. doi:10.1198/jasa.2009.0121.

Zhou, X. H., S. L. Li, F. Tian, Y. M. Xie, Y. Pei, S. Kang, M. Fan, and J. P. Li. 2012. Building a disease risk model of osteoporosis based on traditional Chinese medicine symptoms and western medicine risk factors. *Statistics in Medicine* 31:643–652. doi:10.1002/sim.4382.

Zou, H., T. Hastie, and R. Tibshirani. 2006. Sparse principal component analysis. *Journal of Computational and Graphical Statistics* 15:265–286. doi:10.1198/106186006X113430.