Statistical tests for biosimilarity based on relative distance between follow-on biologics for ordinal endpoints

Myung Soo Yoo[a], Donguk Kim[1,a]

[a]Department of Statistics, Sungkyunkwan University, Korea

Abstract

Investigations of biosimilarity between reference drugs and test drugs required statistical tests; in addition, statistical tests to evaluate biosimilarity have been recently proposed. Ordinal outcome data has been observed in research; however, appropriate statistical tests to deal with ordinal endpoints for biosimilar have not yet been proposed. This paper extends existing design for ordinal endpoints. Using measure of nominal-ordinal association and relative distances between drugs are defined so that testing procedures are developed. Through simulation studies, we investigate type I error rate and power to show the performance of our suggested method. Furthermore, a comparison between the statistical tests and other designs is proviede to show significance of ordinal endpoints.

Keywords: biosimilar, nominal-ordinal association, equivalence trials, three-arm parallel design

1. Introduction

Biological products, medicines produced from living organisms, have opened the new approach to the treatment of several types of diseases such as cancer (Vulto and Jaquez, 2017). However, the use of biological products is limited due to high costs. As the patents of biological products expire, generic of biological products, which are called biosimilar (or follow-on biologics), has received interest from the pharmaceutical industry and related areas. Biosimilar is known to have similar effects with biological products, but it has considerable advantages in that it allows people to have more affordable access to drugs than biological products. Biological products have different characteristics from chemical drugs. In addition, biological products have more complex structures that are likely to have a larger variance than chemical drugs. A different approach to assess statistical similarity between reference drugs and test drugs are required due to the fundamental differences between biological products and chemical products.

Various statistical tests to assess biosimilarity have been proposed (Chen *et al.*, 2017; Shin and Kang, 2016; Lu *et al.*, 2014; Zhang *et al.*, 2014; Yang *et al.*, 2012; Kang and Chow, 2013). Kang and Chow (2013) considered statistical tests based on the ratio estimator and linearization method in a three-arm parallel design to determine biosimilarity between reference drugs and test drugs for continuous endpoints. Based on the power function, simulation to compare each methods showed that power of ratio estimator is greater than that of the linearization method. Lu *et al.* (2014) took the frequency estimator into account to assess biosimilarity in three-arm parallel design. Absolute difference between reference drugs and test drugs with an indicator function is considered to formulate frequency estimator. Power of frequency estimator is compared to that of the ratio estimator and

---

linearization method by Kang and Chow (2013), which demonstrated that the frequency estimator is more powerful than the other two methods when a suitable biosimilar margin is chosen. Kang and Shin (2015) extended three-arm parallel design to $(k + 1)$-arm parallel design. Shin and Kang (2016) extended statistical test based on the ratio estimator for binary endpoints. For the binary endpoints, the risk difference, the log relative risk, and the log odds ratio were taken for the ratio estimator. A type I error rate and power are also investigated theoretically and empirically. Chen *et al.* (2017) recently, assessed biosimilarity based on tolerance limits in a two-arm parallel design. The biosimilarity index based on tolerance limits was found to be more stringent and conservative when compared to other moment-based criterion, especially when variance of biosimilar is larger than that of reference drugs.

Yang *et al.* (2012) demonstrated that the assessment of biosimilarity in variance should be concerned, so that they considered an adapted F-test, which is an extension of traditional F-test for homogeneity of variability. To weaken the normal assumption for adapted F-test, nonparametric tests were also proposed by Zhang *et al.* (2014). The nonparametric tests were found to be robust in controlling type I error rate, especially when the underlying distribution is skewed or has a heavy tail. Research on sample size calculation for the assessment of biosimilarity have been also considered by Kang and Kim (2014) and Kang *et al.* (2015). US FDA (2015) provided a guideline to develop biosimilar in which sponsors are required to choose endpoints to demonstrate biosimilarity as well as to explain the reason why they choose such a design.

Of the proposed methods, Kang and Chow (2013) developed three-arm parallel design utilizing a relative distance between test drugs and reference drugs as a biosimilarity criterion. In this design, relative distance is defined as

$$\text{rd} = \frac{d(T, R)}{d(R_1, R_2)}, \tag{1.1}$$

where distance between test drugs and reference drugs are denoted by $d(T, R)$ and $d(R_1, R_2)$ represents the distance between reference drugs from two different batches. With this criterion, they considered the following statistical tests.

$$H_0 : \text{rd} \geq \delta \quad \text{versus} \quad H_A : \text{rd} < \delta, \tag{1.2}$$

where $\delta \, (\delta > 0)$ is prespecified margin. For the continuous endpoints, relative distance could be interpreted as absolute value of mean difference between test drugs and reference drugs scaled by the absolute value of mean difference between reference drugs from different batches. Relative distance employs distance between reference drugs from different batches as a denominator; therefore, a larger variability of biological product is considered in assessing biosimilarity, in which two-arm balanced design fails to accomplish. For more detail, refer to Kang and Chow (2013).

This work makes an extension of three-arm parallel design for ordinal endpoints based on the advantage of three-arm parallel design. An example of ordinal endpoints could be found in Doll and Pygott (1952); the change in the size of an ulcer crater after three months of treatment (larger, less than 2/3 healed, 2/3 or more healed, healed). Ordinal outcome data analyzed after being dichotomized into the form of binary data are shown to cause some loss of information in ordinal data (Sankey and Weissfeld, 1998). Roozenbeek *et al.* (2011) also demonstrated that statistical power can increase when ordinal analysis is conducted, instead of collapsing data into a binary form. Following this research, we focus on the statistical tests to assess biosimilarity, when endpoints have the ordinal form.

In the next Section, using summary measures for nominal-ordinal association, relative distance on new three-arm parallel design for ordinal endpoints are defined. Section 3 describes statistical tests for biosimilarity. Section 4 theoretically and empirically investigates as well as compares type I error rate

and power. Section 5 conducts simulation studies to show applicability and the significance of ordinal endpoints in determining biosimilarity. Finally, Session 6 provides the conclusion and suggestions for future study.

2. Distances for ordinal endpoints on three-arm parallel design

In this Section, we define relative distance on a new three-arm parallel design for ordinal endpoints. Section 1 indicated that relative distance is defined as a ratio of two distances; one between test drug (biosimilar) and reference drugs and another one between reference drugs from different batches. Two summary measures of the degree of association between a nominal variable and ordinal variable, which are denoted by Δ and α, are taken for relative distance. Δ and α are defined as follows (Agresti, 1981).

$$\Delta = P(Y_1 > Y_2) - P(Y_2 > Y_1) = \sum_{i>j} \pi_{i|1} \pi_{j|2} - \sum_{i<j} \pi_{i|1} \pi_{j|2}, \tag{2.1}$$

$$\alpha = \frac{P(Y_1 > Y_2)}{P(Y_2 > Y_1)} = \frac{\sum_{i>j} \pi_{i|1} \pi_{j|2}}{\sum_{i<j} \pi_{i|1} \pi_{j|2}}, \tag{2.2}$$

where Y_1 and Y_2 denote independent random variables representing ordinal category numbers of the response variable for subjects selected at random from group 1 and group 2, respectively. Also, $\pi_{j|i} = \pi_{ij}/\pi_{i+}$ with π_{ij} denoting cell probability for the ith row and j^{th} column and $\pi_{i+} = \sum_j \pi_{ij}$ for two-way contingency tables. In the development of biosimilar, group 1 and group 2 could be patients taking test drugs (biosimilar) and those taking reference drugs in a clinical trial. Δ can be interpreted as the difference of two probabilities; one that group 1 (patients taking test drugs) shows higher ordinal category numbers (e.g., better response in clinical trial) than group 2 (patients taking reference drugs) and another that group 2 shows higher ordinal category numbers than group 1. However, α represents the ratio between two probabilities explained in the case of Δ. These measures utilize ordering of the levels of the ordinal variables. Furthermore, note that when ordinal category number $K = 2$, Δ is equal to $\pi_{1|2} - \pi_{1|1}$ while α is equal to $(\pi_{12}\pi_{21})/(\pi_{11}\pi_{22})$ (Agresti, 1981).

The sample versions of these two measures are easily defined by substituting cell probability with sample proportion,

$$\hat{\Delta} = \sum_{i>j} \hat{\pi}_{i|1} \hat{\pi}_{j|2} - \sum_{i<j} \hat{\pi}_{i|1} \hat{\pi}_{j|2}, \tag{2.3}$$

$$\hat{\alpha} = \frac{\sum_{i>j} \hat{\pi}_{i|1} \hat{\pi}_{j|2}}{\sum_{i<j} \hat{\pi}_{i|1} \hat{\pi}_{j|2}}, \tag{2.4}$$

where $\hat{\pi}_{j|i} = \hat{\pi}_{ij}/\hat{\pi}_{i+}$ with $\hat{\pi}_{ij}$ denoting sample proportion of ith row and j^{th} column and $\hat{\pi}_{i+} = \sum_j \hat{\pi}_{ij}$. Agresti (1981) also demonstrated asymptotic normal distribution for (2.3) and (2.4).

Using Δ and $\log \alpha$, we define the two relative distances on the three-arm parallel design. Let T, R_1, and R_2 denote test drugs (biosimilar) and reference drugs from two different batches, respectively. Assume that total number of patients, N, are randomized into three groups. The number of each group is represented by n_i, $i = 1, 2, 3$; number of group 1 taking test drugs (biosimilar), those of group 2 and group 3 receiving reference drugs from different batches. We employ $1 : 1 : 1$ as a randomization ratio, $N = n_1 + n_2 + n_3$. By assuming that $Y_T \sim \text{Multi}(n_1, \boldsymbol{\pi_T})$, $Y_{R_1} \sim \text{Multi}(n_2, \boldsymbol{\pi_{R_1}})$, $Y_{R_2} \sim \text{Multi}(n_3, \boldsymbol{\pi_{R_2}})$, where Y_T, Y_{R_1}, Y_{R_2} are independent random variables denoting ordinal category

number of the response variables with $\boldsymbol{\pi}_T$, $\boldsymbol{\pi}_{R_1}$, $\boldsymbol{\pi}_{R_2}$ representing cell probabilities for each group. Distances with measure Δ and α are defined as

$$d_\Delta(T,R) = \left|\Delta^{TR}\right| = |P(Y_T > Y_R) - P(Y_R > Y_T)|$$

$$= \left|\sum_{i>j} \pi_{i|T}\pi_{j|R} - \sum_{i<j} \pi_{i|T}\pi_{j|R}\right|, \tag{2.5}$$

$$d_\Delta(R_1,R_2) = \left|\Delta^{R_1 R_2}\right|$$

$$= \left|\sum_{i>j} \pi_{i|R_1}\pi_{j|R_2} - \sum_{i<j} \pi_{i|R_1}\pi_{j|R_2}\right|, \tag{2.6}$$

$$d_\alpha(T,R) = \left|\log \alpha^{TR}\right|$$

$$= \left|\log \sum_{i>j} \pi_{i|T}\pi_{j|R} - \log \sum_{i<j} \pi_{i|T}\pi_{j|R}\right|, \tag{2.7}$$

$$d_\alpha(R_1,R_2) = \left|\log \alpha^{R_1 R_2}\right|$$

$$= \left|\log \sum_{i>j} \pi_{i|R_1}\pi_{j|R_2} - \log \sum_{i<j} \pi_{i|R_1}\pi_{j|R_2}\right|, \tag{2.8}$$

where $\Delta^{R_1 R_2}$ and $\log \alpha^{R_1 R_2}$ represent measures Δ and $\log \alpha$ from groups taking reference drugs from batches 1 and 2. With $\pi_{i|R}$ being defined as the arithmetic mean of $\pi_{i|R_1}$ and $\pi_{i|R_2}$, Δ^{TR} and $\log \alpha^{TR}$ could be seen as each measure from a group taking test drugs and one taking reference drugs.

With these two pairs of distances, relative distance, rd_Δ and rd_α, can be defined as

$$\mathrm{rd}_\Delta = \frac{d_\Delta(T,R)}{d_\Delta(R_1,R_2)} = \left|\frac{\sum_{i>j} \pi_{i|T}\pi_{j|R} - \sum_{i<j} \pi_{i|T}\pi_{j|R}}{\sum_{i>j} \pi_{i|R_1}\pi_{j|R_2} - \sum_{i<j} \pi_{i|R_1}\pi_{j|R_2}}\right|, \tag{2.9}$$

$$\mathrm{rd}_\alpha = \frac{d_\alpha(T,R)}{d_\alpha(R_1,R_2)} = \left|\frac{\log \sum_{i>j} \pi_{i|T}\pi_{j|R} - \log \sum_{i<j} \pi_{i|T}\pi_{j|R}}{\log \sum_{i>j} \pi_{i|R_1}\pi_{j|R_2} - \log \sum_{i<j} \pi_{i|R_1}\pi_{j|R_2 j}}\right|. \tag{2.10}$$

Assessment of biosimilarity is conducted based on these relative distance with prespecified margin $\delta(\delta > 0)$ as follows.

$$H_0 : \mathrm{rd} \geq \delta \quad \text{versus} \quad H_A : \mathrm{rd} < \delta. \tag{2.11}$$

Note that (2.11) with measure Δ and $\log \alpha$ could be expressed, respectively, as

$$H_0 : \left|\frac{\Delta^{TR}}{\Delta^{R_1 R_2}}\right| \geq \delta_\Delta \quad \text{versus} \quad H_A : \left|\frac{\Delta^{TR}}{\Delta^{R_1 R_2}}\right| < \delta_\Delta, \tag{2.12}$$

$$H_0 : \left|\frac{\log \alpha^{TR}}{\log \alpha^{R_1 R_2}}\right| \geq \delta_\alpha \quad \text{versus} \quad H_A : \left|\frac{\log \alpha^{TR}}{\log \alpha^{R_1 R_2}}\right| < \delta_\alpha. \tag{2.13}$$

3. Statistical tests for biosimilarity

In the last section, we define relative distance using nominal-ordinal association measures, Δ and $\log \alpha$. Statistical tests for biosimilarity based on two relative distances are provided in this section. We derive the asymptotic distribution of the test statistic along with its testing procedure will be explained. From now on, $\Delta^{TR}/\Delta^{R_1 R_2}$ and $\log \alpha^{TR}/\log \alpha^{R_1 R_2}$ will be denoted as θ_Δ and θ_α, respectively.

3.1. Statistical tests based on Δ with k-categories

Note that (2.12) can be decomposed into two one-sided tests such as

$$H_{01} : \theta_\Delta \geq \delta_\Delta \quad \text{versus} \quad H_{A1} : \theta_\Delta < \delta_\Delta, \tag{3.1}$$

and

$$H_{02} : \theta_\Delta \leq -\delta_\Delta \quad \text{versus} \quad H_{A2} : \theta_\Delta > -\delta_\Delta, \tag{3.2}$$

where $\delta_\Delta(> 0)$ is predefined margin. When both hypotheses tests are rejected at a significance level, it could be claimed that test drugs has similar effect with reference drugs. The reasonable test statistic for θ_Δ could be obtained by substituting $\pi_{j|i}$ with $\hat{\pi}_{j|i}$, and is denoted as $\hat{\theta}_\Delta$.

Before deriving the asymptotic distribution of $\hat{\theta}_\Delta$, note that if $X \sim \text{Multi}(n,\pi)$ with $\pi = (\pi_1, \ldots, \pi_k)$,

$$\sqrt{n} \left[\begin{pmatrix} \hat{\pi}_1 \\ \hat{\pi}_2 \\ \vdots \\ \hat{\pi}_k \end{pmatrix} - \begin{pmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_k \end{pmatrix} \right] \xrightarrow{d} N_k \left[\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ \Sigma \right], \tag{3.3}$$

where $\Sigma = \text{Diag}(\pi) - \pi\pi'$. We use multivariate delta method to derive the asymptotic distribution of $\hat{\theta}_\Delta$. Note that θ_Δ can be seen as g_1 function of π_{R_1}, π_{R_2}, and π_T as follows.

$$
\begin{aligned}
\theta_\Delta &= g_1(\pi_{R_1}, \pi_{R_2}, \pi_T) \\
&= \frac{\frac{1}{2} \sum_{j=1}^{k-1}(\pi_{j|R_1} + \pi_{j|R_2}) \left(\sum_{q=j+1}^{k} \pi_{q|T} \right) - \frac{1}{2} \sum_{j=1}^{k-1} \pi_{j|T} \left(\sum_{q=j+1}^{k}(\pi_{q|R_1} + \pi_{q|R_2}) \right)}{\sum_{j=1}^{k-1} \pi_{j|R_2} \left(\sum_{q=j+1}^{k} \pi_{q|R_1} \right) - \sum_{j=1}^{k-1} \pi_{j|R_1} \left(\sum_{q=j+1}^{k} \pi_{q|R_2} \right)},
\end{aligned}
\tag{3.4}
$$

where $\pi_{j|R_i}$ is cell probability of jth category in group taking one of reference drug from batch i, $i = 1, 2$ and $\pi_{j|T}$ is cell probability of jth category in group taking test drugs. Then, it can be seen that (see Appendix)

$$\sqrt{n_1} \left(\hat{\theta}_\Delta - \theta_\Delta \right) \xrightarrow{d} N \left(0, \sigma_\Delta^2 \right), \quad \sigma_\Delta^2 = B_1 \Sigma B_1', \tag{3.5}$$

where $B_1 = (dg_1/d\pi_{1|R_1}, \ldots, dg_1/d\pi_{k-1|R_1}, dg_1/d\pi_{1|R_2}, \ldots, dg_1/d\pi_{k-1|R_2}, dg_1/d\pi_{1|T}, \ldots, dg_1/d\pi_{k-1|R_T})$ and Σ is computed as follows.

$$\Sigma = \begin{bmatrix} \Sigma_1 & & \\ & \Sigma_2 & \\ & & \Sigma_3 \end{bmatrix}, \tag{3.6}$$

where

$$\Sigma_1 = \begin{bmatrix} \pi_{1|R_1}(1 - \pi_{1|R_1}) & -\pi_{1|R_1}\pi_{2|R_1} & \cdots & -\pi_{1|R_1}\pi_{k-1|R_1} \\ -\pi_{2|R_1}\pi_{1|R_1} & \pi_{2|R_1}(1 - \pi_{2|R_1}) & \cdots & -\pi_{2|R_1}\pi_{k-1|R_1} \\ \vdots & \vdots & \ddots & \vdots \\ -\pi_{k-1|R_1}\pi_{1|R_1} & -\pi_{k-1|R_1}\pi_{2|R_1} & \cdots & \pi_{k-1|R_1}(1 - \pi_{k-1|R_1}) \end{bmatrix}, \tag{3.7}$$

$$\Sigma_2 = \begin{bmatrix} \pi_{1|R_2}(1-\pi_{1|R_2}) & -\pi_{1|R_2}\pi_{2|R_2} & \cdots & -\pi_{1|R_2}\pi_{k-1|R_2} \\ -\pi_{2|R_2}\pi_{1|R_2} & \pi_{2|R_2}(1-\pi_{2|R_2}) & \cdots & -\pi_{2|R_2}\pi_{k-1|R_2} \\ \vdots & \vdots & \ddots & \vdots \\ -\pi_{k-1|R_2}\pi_{1|R_2} & -\pi_{k-1|R_2}\pi_{2|R_2} & \cdots & \pi_{k-1|R_2}(1-\pi_{k-1|R_2}) \end{bmatrix}, \tag{3.8}$$

$$\Sigma_3 = \begin{bmatrix} \pi_{1|T}(1-\pi_{1|T}) & -\pi_{1|T}\pi_{2|T} & \cdots & -\pi_{1|T}\pi_{k-1|T} \\ -\pi_{2|T}\pi_{1|T} & \pi_{2|T}(1-\pi_{2|T}) & \cdots & -\pi_{2|T}\pi_{k-1|T} \\ \vdots & \vdots & \ddots & \vdots \\ -\pi_{k-1|T}\pi_{1|T} & -\pi_{k-1|T}\pi_{2|T} & \cdots & \pi_{k-1|T}(1-\pi_{k-1|T}) \end{bmatrix}. \tag{3.9}$$

Using the asymptotic normal distribution of $\hat{\theta}_\Delta$, hypothesis tests in (3.1) and (3.2) could be conducted with $Z_{1\Delta}$ and $Z_{2\Delta}$, respectively. Therefore,

$$Z_{1\Delta} < -z_\alpha \quad \text{and} \quad Z_{2\Delta} > z_\alpha, \tag{3.10}$$

where z_α is the 100α upper percentile of the standard normal distribution. $Z_{1\Delta}$ and $Z_{2\Delta}$ is

$$Z_{1\Delta} = \frac{\hat{\theta}_\Delta - \delta_\Delta}{\hat{\sigma}_\Delta / \sqrt{n_1}} \quad \text{and} \quad Z_{2\Delta} = \frac{\hat{\theta}_\Delta + \delta_\Delta}{\hat{\sigma}_\Delta / \sqrt{n_1}}, \tag{3.11}$$

where $\hat{\sigma}_\Delta$ can be computed by substituting $\pi_{j|i}$ with $\hat{\pi}_{j|i}$. If each null hypothesis in (3.1) and (3.2) is rejected at a significance level α, then we can conclude that test drugs (biosimilar) is similar to reference drugs.

3.2. Statistical tests based on $\log \alpha$ with k-categories

Similarly, (2.13) can be decomposed into two one-sided tests as follows.

$$H_{01} : \theta_\alpha \geq \delta_\alpha \quad \text{versus} \quad H_{A1} : \theta_\alpha < \delta_\alpha, \tag{3.12}$$

and

$$H_{02} : \theta_\alpha \leq -\delta_\alpha \quad \text{versus} \quad H_{A2} : \theta_\alpha > -\delta_\alpha, \tag{3.13}$$

where $\delta_\alpha \,(> 0)$ is predefined margin. Note that θ_α can be expressed as g_2 function of $\boldsymbol{\pi_{R_1}}, \boldsymbol{\pi_{R_2}}$, and $\boldsymbol{\pi_T}$ such that

$$\begin{aligned} \theta_\alpha &= g_2\left(\boldsymbol{\pi_{R_1}}, \boldsymbol{\pi_{R_2}}, \boldsymbol{\pi_T}\right) \\ &= \frac{\log\left[\sum_{j=1}^{k-1} \frac{\pi_{j|R_1} + \pi_{j|R_2}}{2}\left(\sum_{q=j+1}^{k} \pi_{q|T}\right)\right] - \log\left[\sum_{j=1}^{k-1} \pi_{j|T}\left(\sum_{q=j+1}^{k} \frac{\pi_{q|R_1} + \pi_{q|R_2}}{2}\right)\right]}{\log\left[\sum_{j=1}^{k-1} \pi_{j|R_2}\left(\sum_{q=j+1}^{k} \pi_{q|R_1}\right)\right] - \log\left[\sum_{j=1}^{k-1} \pi_{j|R_1}\left(\sum_{q=j+1}^{k} \pi_{q|R_2}\right)\right]}. \end{aligned} \tag{3.14}$$

Then, it can also be shown that (see Appendix)

$$\sqrt{n_1}\left(\hat{\theta}_\alpha - \theta_\alpha\right) \xrightarrow{d} N\left(0, \sigma_\alpha^2\right), \quad \sigma_\alpha^2 = B_2 \Sigma B_2', \tag{3.15}$$

where $B_2 = (dg_2/d\pi_{1|R_1}, \ldots, dg_2/d\pi_{k-1|R_1}, dg_2/d\pi_{1|R_2}, \ldots, dg_2/d\pi_{k-1|R_2}, dg_2/d\pi_{1|T}, \ldots, dg_2/d\pi_{k-1|R_T})$ and Σ is as same as (3.6). Using asymptotic normality of $\hat{\theta}_\alpha$, hypotheses tests in (3.12) and (3.13) can be conducted. As before, we could claim that biosimilarity between test drugs (biosimilar) and

reference drugs are established if each null hypothesis in (3.12) and (3.13) is rejected at a significance level α. Therefore,

$$Z_{1\alpha} < -z_\alpha \quad \text{and} \quad Z_{2\alpha} > z_\alpha, \tag{3.16}$$

where $Z_{1\alpha} = (\hat{\theta}_\alpha - \delta_\alpha)/(\hat{\sigma}_\alpha/\sqrt{n_1})$, $Z_{2\alpha} = (\hat{\theta}_\alpha + \delta_\alpha)/(\hat{\sigma}_\alpha/\sqrt{n_1})$, and $z_\alpha = 100\alpha$ upper percentile of the standard normal distribution.

4. Type I error rate and power function

In this Section, type I error rate and power are investigated based on both measures Δ and $\log\alpha$. Shin and Kang (2016) derived power function for binary endpoints; however, we deal with ordinal endpoints. For predefined margin $\delta_\Delta > 0$, if $z_\alpha < \delta_\Delta/(\sigma_\Delta/\sqrt{n_1})$, then type I error rate at $\theta_\Delta = \delta_\Delta$ based on measure Δ is

$$P\left(\frac{\hat{\theta}_\Delta + \delta_\Delta}{\hat{\sigma}_\Delta/\sqrt{n_1}} > z_\alpha \quad \text{and} \quad \frac{\hat{\theta}_\Delta - \delta_\Delta}{\hat{\sigma}_\Delta/\sqrt{n_1}} < -z_\alpha | \theta_\Delta = \delta_\Delta\right)$$
$$\simeq P\left(Z_{1\Delta} > z_\alpha - \frac{2\delta_\Delta}{\sigma_\Delta/\sqrt{n_1}} \quad \text{and} \quad Z_{1\Delta} < -z_\alpha | \theta_\Delta = \delta_\Delta\right). \tag{4.1}$$

Similarly, type I error rate at $\theta_\Delta = -\delta_\Delta$ is

$$P\left(\frac{\hat{\theta}_\Delta + \delta_\Delta}{\hat{\sigma}_\Delta/\sqrt{n_1}} > z_\alpha \quad \text{and} \quad \frac{\hat{\theta}_\Delta - \delta_\Delta}{\hat{\sigma}_\Delta/\sqrt{n_1}} < -z_\alpha | \theta_\Delta = -\delta_\Delta\right) \tag{4.2}$$
$$= P\left(\frac{\hat{\theta}_\Delta - (-\delta_\Delta)}{\hat{\sigma}_\Delta/\sqrt{n_1}} > z_\alpha \quad \text{and} \quad \frac{\hat{\theta}_\Delta + \delta_\Delta - 2\delta_\Delta}{\hat{\sigma}_\Delta/\sqrt{n_1}} < -z_\alpha | \theta_\Delta = -\delta_\Delta\right)$$
$$\simeq P\left(Z_{2\Delta} > z_\alpha \quad \text{and} \quad Z_{2\Delta} < -z_\alpha + \frac{2\delta_\Delta}{\sigma_\Delta/\sqrt{n_1}} \bigg| \theta_\Delta = -\delta_\Delta\right),$$

where $Z_{1\Delta} = (\hat{\theta}_\Delta - \delta_\Delta)/(\hat{\sigma}_\Delta/\sqrt{n_1})$, $Z_{2\Delta} = (\hat{\theta}_\Delta + \delta_\Delta)/(\hat{\sigma}_\Delta/\sqrt{n_1})$. The power can be computed with the formulas (4.1) and (4.2) under the alternative hypothesis. In some cases, the condition $z_\alpha < \delta_\Delta/(\sigma_\Delta/\sqrt{n_1})$ is not satisfied so that type I error rate and power cannot be computed. Note that formula for both type I error rate and power based on $\log\alpha$ can be derived similarly with (4.1) and (4.2).

Under different settings, theoretical type I error rate and empirical type I error rate based on Δ and $\log\alpha$ are investigated. For the empirical type I error rate, random samples under null hypothesis are generated according to sample size. After 5,000 replication, empirical type I error rate is calculated as the proportion of rejecting null hypothesis. Note that the empirical power can be calculated similarly. In our simulation settings, number of categories, K, is fixed as 3, but similar results could be obtained for other cases. Tables 1–4 gives the results.

Tables 1 and 2 indicate that the theoretical type I error rate and empirical type I error rate vary according to parameter settings. For example, when $(\pi_{1|T}, \pi_{2|T}, \pi_{3|T}) = (0.6, 0.2, 0.2)$ the empirical type I error rate based on Δ is 0.078, whereas the one based on $\log\alpha$ is 0.065 for $n_1 = 1000$. When $(\pi_{1|T}, \pi_{2|T}, \pi_{3|T}) = (0.14, 0.41, 0.45)$, empirical type I error rate based on Δ is 0.056; however, the one based on measure $\log\alpha$ is 0.075 for $n_1 = 1000$. It could be found that empirical type I error rate of both proposed measures becomes closer to the theoretical type I error rate, in general, as n_1 becomes large. Tables 3 and 4 present the theoretical power and empirical power based on both measures under various settings. As in the case of type I error rate, theoretical power and empirical

Table 1: Comparison of theoretical and empirical type I error rate (%) based on Δ at significance level $\alpha = 5\%$

$\pi_{1\|R_1}, \ldots, \pi_{3\|R_1}$	$\pi_{1\|R_2}, \ldots, \pi_{3\|R_2}$	$\pi_{1\|T}, \ldots, \pi_{3\|T}$	θ_Δ	δ_Δ	$n_1 = 500$		$n_1 = 1000$	
					Theoretical	Empirical	Theoretical	Empirical
(0.3, 0.4, 0.3)	(0.2, 0.5, 0.3)	(0.1, 0.6, 0.3)	−1.50	1.50	3.3	5.7	4.9	5.5
(0.2, 0.55, 0.25)	(0.05, 0.7, 0.25)	(0.1, 0.6, 0.3)	−0.56	0.56	4.4	4.6	4.9	5.0
(0.25, 0.4, 0.35)	(0.15, 0.42, 0.43)	(0.14, 0.41, 0.45)	−0.67	0.67	4.8	6.0	5.0	5.6
(0.25, 0.4, 0.35)	(0.15, 0.42, 0.43)	(0.16, 0.41, 0.43)	−0.45	0.45	1.9	4.3	4.9	5.1
(0.3, 0.3, 0.4)	(0.1, 0.5, 0.4)	(0.2, 0.3, 0.5)	−0.67	0.67	4.8	5.6	5.0	5.6
(0.05, 0.65, 0.3)	(0.1, 0.5, 0.4)	(0.2, 0.5, 0.3)	2.12	2.12	2.2	8.5	4.9	7.9
(0.3, 0.4, 0.3)	(0.4, 0.4, 0.2)	(0.3, 0.38, 0.32)	0.59	0.59	7.9	7.6	5.0	6.3
(0.4, 0.4, 0.2)	(0.3, 0.4, 0.3)	(0.3, 0.4, 0.3)	−0.50	0.50	4.2	6.5	5.0	6.0
(0.4, 0.4, 0.2)	(0.5, 0.4, 0.1)	(0.3, 0.4, 0.3)	1.50	1.50	5.0	9.1	5.0	8.4
(0.4, 0.4, 0.2)	(0.1, 0.4, 0.5)	(0.3, 0.4, 0.3)	0.17	0.17	4.8	5.1	5.0	5.1
(0.3, 0.4, 0.3)	(0.4, 0.4, 0.2)	(0.4, 0.38, 0.22)	−0.40	0.40	2.0	5.8	4.9	5.6
(0.8, 0.1, 0.1)	(0.7, 0.2, 0.1)	(0.5, 0.3, 0.2)	−2.78	2.78	4.9	11.7	5.0	9.8
(0.7, 0.2, 0.1)	(0.4, 0.4, 0.2)	(0.5, 0.3, 0.2)	−0.21	0.21	4.7	5.3	5.0	5.1
(0.3, 0.4, 0.3)	(0.4, 0.4, 0.2)	(0.5, 0.38, 0.12)	−1.40	1.40	5.0	9.7	5.0	8.7
(0.3, 0.4, 0.3)	(0.4, 0.4, 0.2)	(0.5, 0.4, 0.1)	−1.50	1.50	5.0	9.8	5.0	8.7
(0.4, 0.4, 0.2)	(0.5, 0.4, 0.1)	(0.6, 0.2, 0.2)	−0.71	0.71	5.0	8.5	5.0	7.8
(0.5, 0.3, 0.2)	(0.6, 0.2, 0.2)	(0.7, 0.1, 0.2)	−1.50	1.50	4.4	12.2	5.0	11.3
(0.5, 0.1, 0.4)	(0.6, 0.2, 0.2)	(0.7, 0.1, 0.2)	−0.93	0.93	5.0	8.8	5.0	7.4

Table 2: Comparison of theoretical and empirical type I error rate (%) based on $\log \alpha$ at significance level $\alpha = 5\%$

$\pi_{1\|R_1}, \ldots, \pi_{3\|R_1}$	$\pi_{1\|R_2}, \ldots, \pi_{3\|R_2}$	$\pi_{1\|T}, \ldots, \pi_{3\|T}$	θ_α	δ_α	$n_1 = 500$		$n_1 = 1000$	
					Theoretical	Empirical	Theoretical	Empirical
(0.3, 0.4, 0.3)	(0.2, 0.5, 0.3)	(0.1, 0.6, 0.3)	−1.50	1.50	3.2	12.7	4.9	10.5
(0.2, 0.55, 0.25)	(0.05, 0.7, 0.25)	(0.1, 0.6, 0.3)	−0.55	0.55	4.4	6.4	5.0	6.3
(0.25, 0.4, 0.35)	(0.15, 0.42, 0.43)	(0.14, 0.41, 0.45)	−0.68	0.68	4.8	7.9	5.0	7.5
(0.25, 0.4, 0.35)	(0.15, 0.42, 0.43)	(0.16, 0.41, 0.43)	−0.45	0.45	1.6	6.0	4.9	6.1
(0.3, 0.3, 0.4)	(0.1, 0.5, 0.4)	(0.2, 0.3, 0.5)	−0.68	0.68	4.8	7.9	5.0	7.1
(0.05, 0.65, 0.3)	(0.1, 0.5, 0.4)	(0.2, 0.5, 0.3)	1.99	1.99	2.1	13.2	4.9	10.9
(0.3, 0.4, 0.3)	(0.4, 0.4, 0.2)	(0.3, 0.38, 0.32)	0.58	0.58	4.8	7.6	5.0	6.8
(0.4, 0.4, 0.2)	(0.3, 0.4, 0.3)	(0.3, 0.4, 0.3)	−0.49	0.49	4.2	7.2	5.0	6.4
(0.4, 0.4, 0.2)	(0.5, 0.4, 0.1)	(0.3, 0.4, 0.3)	1.43	1.43	5.0	9.8	5.0	7.6
(0.4, 0.4, 0.2)	(0.1, 0.4, 0.5)	(0.3, 0.4, 0.3)	0.15	0.15	4.8	5.3	5.0	5.1
(0.3, 0.4, 0.3)	(0.4, 0.4, 0.2)	(0.4, 0.38, 0.22)	−0.40	0.40	1.7	4.9	4.9	5.8
(0.8, 0.1, 0.1)	(0.7, 0.2, 0.1)	(0.5, 0.3, 0.2)	−2.15	2.15	5.0	10.7	5.0	9.2
(0.7, 0.2, 0.1)	(0.4, 0.4, 0.2)	(0.5, 0.3, 0.2)	−0.20	0.20	4.7	4.5	5.0	5.0
(0.3, 0.4, 0.3)	(0.4, 0.4, 0.2)	(0.5, 0.38, 0.12)	−1.46	−1.46	5.0	8.9	5.0	7.7
(0.3, 0.4, 0.3)	(0.4, 0.4, 0.2)	(0.5, 0.4, 0.1)	−1.58	1.58	5.0	8.7	5.0	7.6
(0.4, 0.4, 0.2)	(0.5, 0.4, 0.1)	(0.6, 0.2, 0.2)	−0.70	0.70	4.9	5.0	5.0	6.5
(0.5, 0.3, 0.2)	(0.6, 0.2, 0.2)	(0.7, 0.1, 0.2)	−1.65	1.65	4.3	11.2	5.0	9.8
(0.5, 0.1, 0.4)	(0.6, 0.2, 0.2)	(0.7, 0.1, 0.2)	−1.04	1.04	5.0	7.8	5.0	6.9

power depend on parameter settings. Both measure Δ and $\log \alpha$ yield a similar power. For example, when $(\pi_{1|T}, \pi_{2|T}, \pi_{3|T}) = (0.4, 0.38, 0.22)$, theoretical power based on measure Δ is 0.956; however, the one based on measure $\log \alpha$ is 0.961 for $n_1 = 500$. In addition, it also could be seen that both powers increases as sample size increases.

5. Numerical results

We assume that ordinal data with categories $K = 4$ are dichotomized into binary data; first and second ordinal categories are collapsed into one category and the other ordinal categories are collapsed

Table 3: Comparison of theoretical and empirical power (%) based on Δ at significance level $\alpha = 5\%$

$\pi_{1|R_1},\ldots,\pi_{3|R_1}$	$\pi_{1|R_2},\ldots,\pi_{3|R_2}$	$\pi_{1|T},\ldots,\pi_{3|T}$	θ_Δ	δ_Δ	$n_1 = 500$		$n_1 = 1000$	
					Theoretical	Empirical	Theoretical	Empirical
$(0.3, 0.4, 0.3)$	$(0.2, 0.5, 0.3)$	$(0.1, 0.6, 0.3)$	-1.50	3.5	81.0	62.3	97.3	76.1
$(0.2, 0.55, 0.25)$	$(0.05, 0.7, 0.25)$	$(0.1, 0.6, 0.3)$	-0.56	1.5	97.6	85.2	99.9	97.2
$(0.25, 0.4, 0.35)$	$(0.15, 0.42, 0.43)$	$(0.14, 0.41, 0.45)$	-0.67	1.7	97.6	93.6	99.9	99.6
$(0.25, 0.4, 0.35)$	$(0.15, 0.42, 0.43)$	$(0.16, 0.41, 0.43)$	-0.45	1.3	95.9	81.5	99.9	95.6
$(0.3, 0.3, 0.4)$	$(0.1, 0.5, 0.4)$	$(0.2, 0.3, 0.5)$	-0.67	1.4	84.3	71.2	98.2	87.6
$(0.05, 0.65, 0.3)$	$(0.1, 0.5, 0.4)$	$(0.2, 0.5, 0.3)$	2.12	5.0	78.0	60.3	96.2	74.2
$(0.3, 0.4, 0.3)$	$(0.4, 0.4, 0.2)$	$(0.3, 0.38, 0.32)$	0.59	1.3	85.2	73.1	98.4	91.3
$(0.4, 0.4, 0.2)$	$(0.3, 0.4, 0.3)$	$(0.3, 0.4, 0.3)$	-0.50	1.2	87.6	76.1	98.9	91.8
$(0.4, 0.4, 0.2)$	$(0.5, 0.4, 0.1)$	$(0.3, 0.4, 0.3)$	1.50	2.5	75.1	64.3	94.9	83.2
$(0.4, 0.4, 0.2)$	$(0.1, 0.4, 0.5)$	$(0.3, 0.4, 0.3)$	0.17	0.4	95.8	92.9	99.9	99.7
$(0.3, 0.4, 0.3)$	$(0.4, 0.4, 0.2)$	$(0.4, 0.38, 0.22)$	-0.40	1.2	95.6	83.9	99.9	96.4
$(0.8, 0.1, 0.1)$	$(0.7, 0.2, 0.1)$	$(0.5, 0.3, 0.2)$	-2.78	4.8	70.0	60.4	92.2	77.6
$(0.7, 0.2, 0.1)$	$(0.4, 0.4, 0.2)$	$(0.5, 0.3, 0.2)$	-0.21	0.5	88.0	88.4	99.0	98.7
$(0.3, 0.4, 0.3)$	$(0.4, 0.4, 0.2)$	$(0.5, 0.38, 0.12)$	-1.40	2.4	80.8	67.7	97.2	84.2
$(0.3, 0.4, 0.3)$	$(0.4, 0.4, 0.2)$	$(0.5, 0.4, 0.1)$	-1.50	2.5	78.2	65.6	96.2	82.1
$(0.4, 0.4, 0.2)$	$(0.5, 0.4, 0.1)$	$(0.6, 0.2, 0.2)$	-0.71	1.4	78.5	70.0	96.3	87.8
$(0.5, 0.3, 0.2)$	$(0.6, 0.2, 0.2)$	$(0.7, 0.1, 0.2)$	-1.50	3.0	65.5	58.2	89.3	71.5
$(0.5, 0.1, 0.4)$	$(0.6, 0.2, 0.2)$	$(0.7, 0.1, 0.2)$	-0.93	1.7	92.8	80.7	99.7	94.6

Table 4: Comparison of theoretical and empirical power (%) based on $\log \alpha$ at significance level $\alpha = 5\%$

$\pi_{1|R_1},\ldots,\pi_{3|R_1}$	$\pi_{1|R_2},\ldots,\pi_{3|R_2}$	$\pi_{1|T},\ldots,\pi_{3|T}$	θ_α	δ_α	$n_1 = 500$		$n_1 = 1000$	
					Theoretical	Empirical	Theoretical	Empirical
$(0.3, 0.4, 0.3)$	$(0.2, 0.5, 0.3)$	$(0.1, 0.6, 0.3)$	-1.50	3.5	72.4	59.2	93.6	72.4
$(0.2, 0.55, 0.25)$	$(0.05, 0.7, 0.25)$	$(0.1, 0.6, 0.3)$	-0.55	1.5	97.1	83.5	99.9	95.9
$(0.25, 0.4, 0.35)$	$(0.15, 0.42, 0.43)$	$(0.14, 0.41, 0.45)$	-0.68	1.7	96.0	93.3	99.9	99.4
$(0.25, 0.4, 0.35)$	$(0.15, 0.42, 0.43)$	$(0.16, 0.41, 0.43)$	-0.45	1.3	94.2	81.3	99.8	95.2
$(0.3, 0.3, 0.4)$	$(0.1, 0.5, 0.4)$	$(0.2, 0.3, 0.5)$	-0.68	1.4	76.7	66.7	95.6	85.0
$(0.05, 0.65, 0.3)$	$(0.1, 0.5, 0.4)$	$(0.2, 0.5, 0.3)$	1.99	5.0	84.6	61.6	98.3	74.8
$(0.3, 0.4, 0.3)$	$(0.4, 0.4, 0.2)$	$(0.3, 0.38, 0.32)$	0.58	1.3	87.6	75.1	98.9	91.0
$(0.4, 0.4, 0.2)$	$(0.3, 0.4, 0.3)$	$(0.3, 0.4, 0.3)$	-0.49	1.2	89.6	76.2	99.3	92.8
$(0.4, 0.4, 0.2)$	$(0.5, 0.4, 0.1)$	$(0.3, 0.4, 0.3)$	1.43	2.5	83.5	68.8	98.0	87.3
$(0.4, 0.4, 0.2)$	$(0.1, 0.4, 0.5)$	$(0.3, 0.4, 0.3)$	0.15	0.4	98.3	96.7	99.9	99.8
$(0.3, 0.4, 0.3)$	$(0.4, 0.4, 0.2)$	$(0.4, 0.38, 0.22)$	-0.40	1.2	96.1	84.1	99.9	97.0
$(0.8, 0.1, 0.1)$	$(0.7, 0.2, 0.1)$	$(0.5, 0.3, 0.2)$	-2.15	4.8	98.5	79.2	99.9	92.1
$(0.7, 0.2, 0.1)$	$(0.4, 0.4, 0.2)$	$(0.5, 0.3, 0.2)$	-0.20	0.5	96.7	94.2	99.9	99.6
$(0.3, 0.4, 0.3)$	$(0.4, 0.4, 0.2)$	$(0.5, 0.38, 0.12)$	-1.46	2.4	71.7	62.3	93.2	81.0
$(0.3, 0.4, 0.3)$	$(0.4, 0.4, 0.2)$	$(0.5, 0.4, 0.1)$	-1.58	2.5	66.5	57.9	90.0	76.3
$(0.4, 0.4, 0.2)$	$(0.5, 0.4, 0.1)$	$(0.6, 0.2, 0.2)$	-0.70	1.4	81.7	70.6	97.5	88.1
$(0.5, 0.3, 0.2)$	$(0.6, 0.2, 0.2)$	$(0.7, 0.1, 0.2)$	-1.65	3.0	51.2	50.8	76.6	64.6
$(0.5, 0.1, 0.4)$	$(0.6, 0.2, 0.2)$	$(0.7, 0.1, 0.2)$	-1.04	1.7	75.5	65.5	95.1	84.8

into the other category. To show the significance of ordinal endpoints in evaluating biosimilarity, we compare the following probabilities based on a simulation study.

$$P_o = \text{the probability of biosimilarity using ordinal data,}$$
$$P_b = \text{the probability of biosimilarity using dichotomized data.}$$

For P_b, we utilize method by Shin and Kang (2016) in which relative distances for binary endpoints are proposed as biosimilarity criterion. They take risk difference and log odds ratio as a component of relative distance. Under parameter settings and a predefined margin, $n_1 = 500$ random samples are generated. With 5000 replication, P_o and P_b are calculated as proportion that biosimilarity is

Table 5: Comparison of p_o and p_b based on Δ and risk difference at significance level $\alpha = 5\%$

$\pi_{1|R_1},\ldots,\pi_{4|R_1}$	$\pi_{1|R_2},\ldots,\pi_{4|R_2}$	$\pi_{1|T},\ldots,\pi_{4|T}$	$\pi_1^*|R_1,\pi_2^*|R_1$	$\pi_1^*|R_2,\pi_2^*|R_2$	$\pi_1^*|T,\pi_2^*|T$	δ_Δ	P_o	P_b
$(0.2,0.35,0.35,0.1)$	$(0.3,0.3,0.3,0.1)$	$(0.5,0.15,0.2,0.15)$	$(0.55,0.45)$	$(0.6,0.4)$	$(0.65,0.35)$	2.0	0.146	0.256
$(0.3,0.2,0.25,0.25)$	$(0.1,0.4,0.4,0.1)$	$(0.18,0.22,0.25,0.35)$	$(0.5,0.5)$	$(0.5,0.5)$	$(0.4,0.6)$	2.0	0.005	0.015
$(0.2,0.3,0.3,0.2)$	$(0.2,0.4,0.3,0.1)$	$(0.6,0.2,0.1,0.1)$	$(0.4,0.6)$	$(0.6,0.4)$	$(0.8,0.2)$	2.0	0.000	0.022
$(0.2,0.3,0.3,0.2)$	$(0.3,0.3,0.3,0.1)$	$(0.1,0.1,0.1,0.7)$	$(0.5,0.5)$	$(0.6,0.4)$	$(0.2,0.8)$	5.0	0.000	0.002
$(0.4,0.4,0.2,0.2)$	$(0.3,0.5,0.1,0.1)$	$(0.1,0.4,0.1,0.4)$	$(0.8,0.2)$	$(0.8,0.2)$	$(0.5,0.5)$	5.0	0.000	0.224
$(0.6,0.1,0.2,0.1)$	$(0.5,0.1,0.2,0.2)$	$(0.2,0.4,0.3,0.1)$	$(0.7,0.3)$	$(0.6,0.4)$	$(0.6,0.4)$	0.5	0.225	0.896
$(0.2,0.2,0.25,0.35)$	$(0.22,0.25,0.2,0.33)$	$(0.25,0.1,0.05,0.6)$	$(0.4,0.6)$	$(0.47,0.53)$	$(0.35,0.65)$	0.5	0.032	0.428
$(0.5,0.2,0.15,0.15)$	$(0.6,0.25,0.1,0.05)$	$(0.25,0.3,0.4,0.05)$	$(0.7,0.3)$	$(0.85,0.15)$	$(0.55,0.45)$	5.0	0.149	0.460
$(0.05,0.05,0.4,0.5)$	$(0.1,0.25,0.3,0.35)$	$(0.4,0.2,0.3,0.1)$	$(0.1,0.9)$	$(0.35,0.65)$	$(0.6,0.4)$	1.0	0.071	0.726

Table 6: Comparison of p_o and p_b based on $\log\alpha$ and log odds ratio at $\alpha = 5\%$

$\pi_{1|R_1},\ldots,\pi_{4|R_1}$	$\pi_{1|R_2},\ldots,\pi_{4|R_2}$	$\pi_{1|T},\ldots,\pi_{4|T}$	$\pi_1^*|R_1,\pi_2^*|R_1$	$\pi_1^*|R_2,\pi_2^*|R_2$	$\pi_1^*|T,\pi_2^*|T$	δ_α	P_o	P_b
$(0.2,0.35,0.35,0.1)$	$(0.3,0.3,0.3,0.1)$	$(0.5,0.15,0.2,0.15)$	$(0.55,0.45)$	$(0.6,0.4)$	$(0.65,0.35)$	2.0	0.159	0.259
$(0.3,0.2,0.25,0.25)$	$(0.1,0.4,0.4,0.1)$	$(0.18,0.22,0.25,0.35)$	$(0.5,0.5)$	$(0.5,0.5)$	$(0.4,0.6)$	2.0	0.004	0.014
$(0.2,0.3,0.3,0.2)$	$(0.2,0.4,0.3,0.1)$	$(0.6,0.2,0.1,0.1)$	$(0.4,0.6)$	$(0.6,0.4)$	$(0.8,0.2)$	2.0	0.000	0.005
$(0.2,0.3,0.3,0.2)$	$(0.3,0.3,0.3,0.1)$	$(0.1,0.1,0.1,0.7)$	$(0.5,0.5)$	$(0.6,0.4)$	$(0.2,0.8)$	5.0	0.000	0.000
$(0.4,0.4,0.2,0.2)$	$(0.3,0.5,0.1,0.1)$	$(0.1,0.4,0.1,0.4)$	$(0.8,0.2)$	$(0.8,0.2)$	$(0.5,0.5)$	5.0	0.000	0.459
$(0.6,0.1,0.2,0.1)$	$(0.5,0.1,0.2,0.2)$	$(0.2,0.4,0.3,0.1)$	$(0.7,0.3)$	$(0.6,0.4)$	$(0.6,0.4)$	0.5	0.442	0.911
$(0.2,0.2,0.25,0.35)$	$(0.22,0.25,0.2,0.33)$	$(0.25,0.1,0.05,0.6)$	$(0.4,0.6)$	$(0.47,0.53)$	$(0.35,0.65)$	0.5	0.015	0.392
$(0.5,0.2,0.15,0.15)$	$(0.6,0.25,0.1,0.05)$	$(0.25,0.3,0.4,0.05)$	$(0.7,0.3)$	$(0.85,0.15)$	$(0.55,0.45)$	5.0	0.244	0.814
$(0.05,0.05,0.4,0.5)$	$(0.1,0.25,0.3,0.35)$	$(0.4,0.2,0.3,0.1)$	$(0.1,0.9)$	$(0.35,0.65)$	$(0.6,0.4)$	1.0	0.088	0.996

concluded. Table 5 represents P_o and P_b for Δ and risk difference. P_o is less than P_b since noticeable difference of parameters in ordinal categories become unclear when categories are dichotomized, which can result in a greater probability of concluding biosimilarity. For example, under the following settings based on Δ and risk difference,

$$\pi_1|R_1,\ldots,\pi_4|R_1 = (0.6,0.1,0.2,0.1),$$
$$\pi_1|R_2,\ldots,\pi_4|R_2 = (0.5,0.1,0.2,0.2),$$
$$\pi_1|T,\ldots,\pi_4|T = (0.2,0.4,0.3,0.1)$$

parameters for reference drugs and test drugs (biosimilar) show significant difference, but this difference becomes obscure after being dichotomized as follows.

$$\pi_1^*|R_1,\pi_2^*|R_1 = (0.7,0.3),$$
$$\pi_1^*|R_2,\pi_2^*|R_2 = (0.6,0.4),$$
$$\pi_1^*|T,\pi_2^*|T = (0.6,0.4),$$

where $\pi_1^*|R_i$, $\pi_2^*|R_i$ are collapsed cell probabilities of R_i, $i = 1, 2$, and $\pi_1^*|T$, $\pi_2^*|T$ are collapsed cell probabilities of T. In this case, $P_o = 0.225$ means that chance of claiming biosimilarity is 22.5%. However, $P_b = 0.896$ means that there is a 89.6% chance that biosimilarity would be concluded, which is greater than P_o. Similar results can be seen in Table 6 when p_o and p_b are computed based on measures $\log\alpha$ and log odds ratio. In this sense, one may conclude that assessing biosimilarity using original ordinal data might result in different results from those using dichotomized data. It would be reasonable not to claim biosimilarity if ordinal parameters for simulation settings show significant difference between reference drugs and test drugs. In that sense, our numerical results show that using ordinal data as original might be more appropriate than dichotomized data.

6. Conclusion and future study

Biosimilar, which has a more affordable price than biological products, could provide people with alternative treatments for several types of disease. Existing statistical methods to assess similarity between reference drugs and test drugs cannot be used directly due to fundamental differences between biological products and chemical products. Accordingly, numerous research to show biosimilarity in a statistical aspect has been conducted. In particular, Shin and Kang (2016) considered three-arm parallel design for binary endpoints. However, determining biosimilarity with ordinal endpoints has not been studied. Unless there exists biosimilarity criterion for ordinal endpoints, ordinal data should be dichotomized, which might cause some loss of information in ordinal data.

We propose two biosimilarity criteria for ordinal endpoints by making an extension of three-arm parallel design, and define relative distances based on Δ and α. Type I error rate and power are investigated after theoretically and empirically deriving the asymptotic sampling distributions of each relative distance. The type I error rate are shown to depend on parameter settings. The theoretical and empirical type I error rate become close to a significance level as the sample size increases. Similar results could be found in the case for power.

We also compare the probability of biosimilarity using ordinal data and the one using dichotomized data to show the significance of assessing biosimilarity with ordinal endpoints. The probabilities of biosimilarity are shown to be different when a remarkable difference between parameters for reference drugs and those for test drugs becomes ambiguous after being dichotomized. This result indicates that careful consideration should precede dichotomizing data in assessing biosimilarity.

This work incorporates a prespecified margin in a three-arm parallel design. A prespecified margin should be statistically and clinically justified; however, they remain controversial. Determining a prespecified margin for evaluating biosimilarity is an important topic for future study. We believe that justified prespecified margin could result in a more accurate assessment of biosimilarity.

There can be more choices for nominal-ordinal association. For example, Piccarreta (2001) proposed new measure of nominal-ordinal association. Comparing the performance of three-arm parallel design using other choices represents an interesting topic for future study. Also, the discussion of our study is restricted to two different batches in calculating the distance between reference drugs. As a future study, our discussion can be extended to $(k + 1)$-arm parallel design for ordinal endpoints.

Appendix:

By the multivariate central limit theorem,

$$
\sqrt{n_1} \left[\begin{pmatrix} \hat{\pi}_{1|R_1} \\ \vdots \\ \hat{\pi}_{k-1|R_1} \\ \hat{\pi}_{1|R_2} \\ \vdots \\ \hat{\pi}_{k-1|R_2} \\ \hat{\pi}_{1|T} \\ \vdots \\ \hat{\pi}_{k-1|T} \end{pmatrix} - \begin{pmatrix} \pi_{1|R_1} \\ \vdots \\ \pi_{k-1|R_1} \\ \pi_{1|R_2} \\ \vdots \\ \pi_{k-1|R_2} \\ \pi_{1|T} \\ \vdots \\ \pi_{k-1|T} \end{pmatrix} \right] \xrightarrow{d} N_{3k-3} \left[\begin{pmatrix} 0 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ 0 \end{pmatrix}, \ \Sigma \right], \tag{A.1}
$$

where

$$\Sigma = \begin{bmatrix} \Sigma_1 & & \\ & \Sigma_2 & \\ & & \Sigma_3 \end{bmatrix}, \tag{A.2}$$

where

$$\Sigma_1 = \begin{bmatrix} \pi_{1|R_1}(1 - \pi_{1|R_1}) & -\pi_{1|R_1}\pi_{2|R_1} & \cdots & -\pi_{1|R_1}\pi_{k-1|R_1} \\ -\pi_{2|R_1}\pi_{1|R_1} & \pi_{2|R_1}(1 - \pi_{2|R_1}) & \cdots & -\pi_{2|R_1}\pi_{k-1|R_1} \\ \vdots & \vdots & \ddots & \vdots \\ -\pi_{k-1|R_1}\pi_{1|R_1} & -\pi_{k-1|R_1}\pi_{2|R_1} & \cdots & \pi_{k-1|R_1}(1 - \pi_{k-1|R_1}) \end{bmatrix},$$

$$\Sigma_2 = \begin{bmatrix} \pi_{1|R_2}(1 - \pi_{1|R_2}) & -\pi_{1|R_2}\pi_{2|R_2} & \cdots & -\pi_{1|R_2}\pi_{k-1|R_2} \\ -\pi_{2|R_2}\pi_{1|R_2} & \pi_{2|R_2}(1 - \pi_{2|R_2}) & \cdots & -\pi_{2|R_2}\pi_{k-1|R_2} \\ \vdots & \vdots & \ddots & \vdots \\ -\pi_{k-1|R_2}\pi_{1|R_2} & -\pi_{k-1|R_2}\pi_{2|R_2} & \cdots & \pi_{k-1|R_2}(1 - \pi_{k-1|R_2}) \end{bmatrix},$$

$$\Sigma_3 = \begin{bmatrix} \pi_{1|T}(1 - \pi_{1|T}) & -\pi_{1|T}\pi_{2|T} & \cdots & -\pi_{1|T}\pi_{k-1|T} \\ -\pi_{2|T}\pi_{1|T} & \pi_{2|T}(1 - \pi_{2|T}) & \cdots & -\pi_{2|T}\pi_{k-1|T} \\ \vdots & \vdots & \ddots & \vdots \\ -\pi_{k-1|T}\pi_{1|T} & -\pi_{k-1|T}\pi_{2|T} & \cdots & \pi_{k-1|T}(1 - \pi_{k-1|T}) \end{bmatrix},$$

and $\hat{\pi}_{j|R_1}, \hat{\pi}_{j|R_2}, \hat{\pi}_{j|T}$ denote sample proportions for reference groups from batch 1 and 2, and test group.

(1) measure Δ

Note that relative distance based on measure Δ can be seen as g_1 function of $\pi_{R_1}, \pi_{R_2}, \pi_T$ such that

$$\begin{aligned} \theta_\Delta &= g_1(\pi_{R_1}, \pi_{R_2}, \pi_T) \\ &= \frac{\frac{1}{2}\sum_{j=1}^{k-1}(\pi_{j|R_1} + \pi_{j|R_2})\left(\sum_{q=j+1}^{k}\pi_{q|T}\right) - \frac{1}{2}\sum_{j=1}^{k-1}\pi_{j|T}\left(\sum_{q=j+1}^{k}(\pi_{q|R_1} + \pi_{q|R_2})\right)}{\sum_{j=1}^{k-1}\pi_{j|R_2}\left(\sum_{q=j+1}^{k}\pi_{q|R_1}\right) - \sum_{j=1}^{k-1}\pi_{j|R_1}\left(\sum_{q=j+1}^{k}\pi_{q|R_2}\right)} \equiv \frac{x}{y}, \end{aligned} \tag{A.3}$$

where $\pi_{j|R_i}$ is cell probability of jth category in group taking reference drug from batch i, $i = 1, 2$, and $\pi_{j|T}$ is cell probability of jth category in group taking test drugs. Then, by multivariate delta method, it can be seen that

$$\sqrt{n_1}\left(\hat{\theta}_\Delta - \theta_\Delta\right) \xrightarrow{d} N\left(0, \sigma_\Delta^2\right), \quad \sigma_\Delta^2 = B_1 \Sigma B_1', \tag{A.4}$$

where Σ is the same as (6.2) and $B_1 = (dg_1/d\pi_{1|R_1}, \ldots, dg_1/d\pi_{k-1|R_1}, dg_1/d\pi_{1|R_2}, \ldots, dg_1/d\pi_{k-1|R_2}, dg_1/d\pi_{1|T}, \ldots, dg_1/d\pi_{k-1|R_T})$.

$$\frac{dg_1}{d\pi_{j|R_1}} = \frac{\left[\frac{1}{2}\sum_{q=j+1}^{k}\pi_{q|T} + \frac{1}{2}\sum_{q=j}^{k-1}\pi_{q|T}\right]y - x\left[(-1)\sum_{q=j}^{k-1}\pi_{q|R_2} - \sum_{q=j+1}^{k}\pi_{q|R_2}\right]}{y^2},$$

$$\frac{dg_1}{d\pi_{j|R_2}} = \frac{\left[\frac{1}{2}\sum_{q=j+1}^{k}\pi_{q|T} + \frac{1}{2}\sum_{q=j}^{k-1}\pi_{q|T}\right]y - x\left[\sum_{q=j+1}^{k}\pi_{q|R_1} + \sum_{q=j}^{k-1}\pi_{q|R_1}\right]}{y^2},$$

$$\frac{dg_1}{d\pi_{j|T}} = \frac{-\frac{1}{2}\sum_{q=j}^{k-1}\left(\pi_{q|R_1} + \pi_{q|R_2}\right) - \frac{1}{2}\sum_{q=j+1}^{k}\left(\pi_{q|R_1} + \pi_{q|R_2}\right)}{y}.$$

(2) measure $\log \alpha$

Note that θ_α can be expressed as g_2 function of π_T, π_{R_1}, and π_{R_2} such that

$$
\begin{aligned}
\theta_\alpha &= g_2\left(\pi_T, \pi_{R_1}, \pi_{R_2}\right) \\
&= \frac{\log\left[\sum_{j=1}^{k-1} \frac{\pi_{j|R_1}+\pi_{j|R_2}}{2}\left(\sum_{q=j+1}^{k} \pi_{q|T}\right)\right] - \log\left[\sum_{j=1}^{k-1}\pi_{j|T}\left(\sum_{q=j+1}^{k}\frac{\pi_{q|R_1}+\pi_{q|R_2}}{2}\right)\right]}{\log\left[\sum_{j=1}^{k-1}\pi_{j|R_2}\left(\sum_{q=j+1}^{k}\pi_{q|R_1}\right)\right] - \log\left[\sum_{j=1}^{k-1}\pi_{j|R_1}\left(\sum_{q=j+1}^{k}\pi_{q|R_2}\right)\right]} \equiv \frac{x}{y},
\end{aligned} \qquad (A.5)
$$

where $\pi_{j|R_1}, \pi_{j|R_2}$, and $\pi_{j|T}$ are defined as same as the Section 3.1. Then, it can also be seen by multivariate delta method that

$$
\sqrt{n_1}\left(\hat{\theta}_\alpha - \theta_\alpha\right) \xrightarrow{d} N\left(0, \sigma_\alpha^2\right), \quad \sigma_\alpha^2 = B_2 \Sigma B_2', \qquad (A.6)
$$

where $B_2 = (dg_2/d\pi_{1|R_1}, \ldots, dg_2/d\pi_{k-1|R_1}, dg_2/d\pi_{1|R_2}, \ldots, dg_2/d\pi_{k-1|R_2}, dg_2/d\pi_{1|T}, \ldots, dg_2/d\pi_{k-1|R_T})$ and Σ is the same as (6.2). Then, $dg_2/d\pi_{j|R_1} = (x'y - xy')/y^2$, where

$$
x' = \frac{\frac{1}{2}\sum_{q=j+1}^{k}\pi_{q|T}}{\left[\sum_{m=1}^{k-1}\frac{\pi_{m|R_1}+\pi_{m|R_2}}{2}\left(\sum_{p=m+1}^{k}\pi_{p|T}\right)\right]} - \frac{-\frac{1}{2}\sum_{q=j}^{k-1}\pi_{q|T}}{\left[\sum_{m=1}^{k-1}\pi_{m|T}\left(\sum_{p=m+1}^{k}\frac{\pi_{p|R_1}+\pi_{p|R_2}}{2}\right)\right]},
$$

$$
y' = \frac{(-1)\sum_{q=j}^{k-1}\pi_{q|R_2}}{\left[\sum_{m=1}^{k-1}\pi_{m|R_2}\left(\sum_{p=m+1}^{k}\pi_{p|R_1}\right)\right]} - \frac{\sum_{q=j+1}^{k}\pi_{q|R_2}}{\left[\sum_{m=1}^{k-1}\pi_{m|R_1}\left(\sum_{p=m+1}^{k}\pi_{p|R_2}\right)\right]}.
$$

Similarly, $dg_2/d\pi_{j|R_2} = (x'y - xy')/y^2$, where

$$
x' = \frac{\frac{1}{2}\sum_{q=j+1}^{k}\pi_{q|T}}{\left[\sum_{m=1}^{k-1}\frac{\pi_{m|R_1}+\pi_{m|R_2}}{2}\left(\sum_{p=m+1}^{k}\pi_{p|T}\right)\right]} - \frac{-\frac{1}{2}\sum_{q=j}^{k-1}\pi_{q|T}}{\left[\sum_{m=1}^{k-1}\pi_{m|T}\left(\sum_{p=m+1}^{k}\frac{\pi_{p|R_1}+\pi_{p|R_2}}{2}\right)\right]},
$$

$$
y' = \frac{\sum_{q=j+1}^{k}\pi_{q|R_1}}{\left[\sum_{m=1}^{k-1}\pi_{m|R_2}\left(\sum_{p=m+1}^{k}\pi_{p|R_1}\right)\right]} - \frac{(-1)\sum_{q=j}^{k-1}\pi_{q|R_1}}{\left[\sum_{m=1}^{k-1}\pi_{m|R_1}\left(\sum_{p=m+1}^{k}\pi_{p|R_2}\right)\right]}.
$$

Also, $dg_2/d\pi_{j|T} = x'/y$, where

$$
x' = \frac{-\frac{1}{2}\sum_{q=j}^{k-1}(\pi_{q|R_1}+\pi_{q|R_2})}{\left[\sum_{m=1}^{k-1}\frac{\pi_{m|R_1}+\pi_{m|R_2}}{2}\left(\sum_{p=m+1}^{k}\pi_{p|T}\right)\right]} - \frac{\frac{1}{2}\sum_{q=j+1}^{k}(\pi_{q|R_1}+\pi_{q|R_2})}{\left[\sum_{m=1}^{k-1}\pi_{m|T}\left(\sum_{p=m+1}^{k}\frac{\pi_{p|R_1}+\pi_{p|R_2}}{2}\right)\right]}.
$$

References

Agresti A (1981). Measures of nominal-ordinal association, *Journal of the American Statistical Association*, **76**, 524–529.

Chen C, Tsou H, Hsiao C, Lai Y, Chang W, and Liu J (2017). A tolerance interval approach to assessing the biosimilarity of follow-on biologics, *Statistics in Biopharmaceutical Research*, **9**, 286–292.

Doll R and Pygott F (1952). Factors influencing the rate of healing of gastric ulcers admission to hospital, phenobarbitone, and ascorbic acid, *Lancet*, **259**, 171–175.

Kang SH and Chow SC (2013). Statistical assessment of biosimilarity based on relative distance between follow-on biologics, *Statistics in Medicine*, **32**, 382–392.

Kang SH and Kim Y (2014). Sample size calculations for the development of biosimilar products, *Journal of Biopharmaceutical Statistics*, **24**, 1215–1224.

Kang SH, Jung J, and Baik S (2015). Sample size calculations for the development of biosimilar products based on binary endpoints, *Communications for Statistical Applications and Methods*, **22**, 389–399.

Kang SH and Shin W (2015). Statistical assesment of biosimilarity based on the relative distance between follow-on biologics in the $(k + 1)$-arm parallel design, *Communications for Statistical Applications and Methods*, **22**, 605–613.

Lawless JF (1982). *Statistical Models and Methods for Lifetime Data*, Wiley, New York.

Lu Y, Zhang Z, and Chow SC (2014). Frequency estimator for assessing of follow-on biologics, *Journal of Biopharmaceutical Statistics*, **24**, 1280–1297.

Piccarreta R (2001). A new measure of nominal-ordinal association, *Journal of Applied Statistics*, **28**, 107–120.

Roozenbeek B, Lingsma H, Perel P, Edwards P, Roberts I, Murray G, Maas A, and Steyerberg E (2011). The added value of ordinal analysis in clinical trials: an example in traumatic brain injury, *Critical Care*, **15**, R127.

Shin W and Kang SH (2016). Statistical assessment of biosimilarity based on the relative distance between follow-on biologics for binary endpoints, *Journal of Biopharmaceutical Statistics*, **26**, 227–239.

Sankey S and Weissfeld L (1998). A study of the effect of dichotomizing ordinal data upon modeling, *Communications in Statistics-Simulation and Computation*, **27**, 871–887.

US Food and Drug Administration (2015). Scientific consideration in demonstrating biosimilarity to a reference product: guidance for industry, Retrieved April, 2015, from: https://www.fda.gov/media /82647/download

Vulto A and Jaquez O (2017). The process defines the product: what really matters in biosimilar design and production?, *Rheumatology*, **56**, iv14–29.

Yang J, Zhang N, Chow SC, and Chi E (2012). An adapted F-test for homogeneity of variability in follow-on biological products, *Statistics in Medicine*, **32**, 415–423.

Zhang N, Yang J, Chow SC, and Chi E (2014). Nonparametric tests for evaluation of biosimilarity in variability of follow-on biologics, *Journal of Biopharmaceutical Statistics*, **24**, 1239–1253.