Computational Statistics and Data Analysis 56 (2012) 1301–1320

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Testing non-inferiority for clustered matched-pair binary data in diagnostic medicine[☆]

Zhao Yang[a,∗], Xuezheng Sun[b], James W. Hardin[c]

[a] Quintiles, Inc., 5927 S Miami Blvd, Morrisville, NC 27560, USA

[b] Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

[c] Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, Columbia, SC 29208, USA

ARTICLE INFO

Article history:
Available online 13 July 2011

Keywords:
Clustered matched-pair binary data
Diagnostic testing
Non-inferiority
Intra-cluster correlation coefficient

ABSTRACT

Testing non-inferiority in active-controlled clinical trials examines whether a new procedure is, to a pre-specified amount, no worse than an existing procedure. To assess non-inferiority between two procedures using clustered matched-pair binary data, two new statistical tests are systematically compared to existing tests. The calculation of corresponding confidence interval is also proposed. None of the tests considered requires structural within-cluster correlation or distributional assumptions. The results of an extensive Monte Carlo simulation study illustrate that the performance of the statistics depends on several factors including the number of clusters, cluster size, probability of success in the test procedure, the homogeneity of the probability of success across clusters, and the intra-cluster correlation coefficient (ICC). In evaluating non-inferiority for a clustered matched-pair study, one should consider all of these issues when choosing an appropriate test statistic. The ICC-adjusted test statistic is generally recommended to effectively control the nominal level when there is constant or small variability of cluster sizes. For a greater number of clusters, the other test statistics maintain the nominal level reasonably well and have higher power. Therefore, with the carefully designed clustered matched-pair study, a combination of the statistics investigated may serve best in data analysis. Finally, to illustrate the practical application of the recommendations, a real clustered matched-pair collection of data is used to illustrate testing non-inferiority.

1. Introduction

Inference from non-inferiority tests is often used to establish that a new procedure is, to a pre-specified amount, no worse than an existing procedure in effectiveness. Assessment of non-inferiority is an important issue in modern medical research and diagnostic medicine. Collected from subjects recruited in clusters, data from subjects in the same cluster are likely to be more similar than data from subjects in different clusters, requiring appropriate statistical analysis to address inter- and intra-cluster correlation. A relevant practical study by Nuñez et al. (2002) compares two tracers (fluorodeoxyglucose(FDG) and C11) in positron emission tomography (PET) for detecting prostate cancer metastatic to the bone. In practice, little is known about appropriate correlation structures. Hence, methods of analysis should adjust for multiple units within a cluster, yet avoid specific assumptions on the structure of inter- and intra-cluster correlation.

To assess non-inferiority in clustered matched-pair binary data, there are several available tests that maintain the nominal Type I error rate. Durkalski et al. (2003) proposed the first statistical test based on the method of moments, and

[☆] The complete extensive simulation results can be found in the supplementary data of the journal website (electronic version of the paper).

∗ Corresponding author. Tel.: +1 678 576 0784.
E-mail address: tonyyangsxz@gmail.com (Z. Yang).

0167-9473/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2011.06.019

Table 1
McNemar-type table for K clusters of matched-pair data.

Procedure 1 (New)	Procedure 2 (Standard)		Total
	Success (1)	Failure (0)	
Success(1)	$\sum_{k=1}^{K} a_k, (p_{11})$	$\sum_{k=1}^{K} b_k = b, (p_{10})$	$\sum_{k=1}^{K}(a_k + b_k) = \sum_{k=1}^{K} x_{1k}$
Failure(0)	$\sum_{k=1}^{K} c_k = c, (p_{01})$	$\sum_{k=1}^{K} d_k, (p_{00})$	
Total	$\sum_{k=1}^{K}(a_k + c_k) = \sum_{k=1}^{K} x_{2k}$		$\sum_{k=1}^{K} n_k = N$

Nam and Kwon (2009) investigated a test to evaluate non-inferiority based on earlier work by Obuchowski (1998). Based on the work of Eliasziw and Donner (1991), Nam and Kwon developed another test for non-inferiority using the idea of an inflation factor from the intra-cluster correlation coefficient (ICC). The inflation factor adjustment via the intra-cluster correlation coefficient is a well-known approach to account for underestimated variance resulting from clustering. In an extensive simulation study, Nam and Kwon (2009) compared these three methods and concluded that all of the tests are appropriate for general practice.

Recently, Yang and Sun (2010) simplified the test statistics proposed by Nam and Kwon (2009) and Obuchowski (1998). In addition, Yang et al. (2010) proposed a new test for clustered matched-pair binary data based on a modification on the covariance estimation in the test by Obuchowski (1998). The extensive simulation study by Yang et al. (2010) suggests that the new test has higher power. In addition, this new test can be straightforwardly extended to testing non-inferiority. A common feature for all tests considered herein is that each test adjusts for multiple sampling units within clusters without requiring assumptions for correlation structure and distribution.

Section 2 presents a general discussion of non-inferiority tests. An extensive Monte Carlo simulation study performed for the comparison of new and extant tests in terms of nominal level and power is presented in Section 3. Section 4 includes a detailed real data analysis example to illustrate the application and interpretation of the competing tests. Finally, further discussion and summary are provided in Section 5.

2. Non-inferiority tests

Suppose there is a random sample of K clusters with n_k units in the kth cluster, $k = 1, \ldots, K$. In the kth cluster, each unit is measured under the ith procedure, $i = 1, 2$, where the index values 1 and 2 correspond to the new and standard procedures, respectively. Let 1 (success) and 0 (failure) indicate the kth unit's response to the ith procedure. Thus, there are four pairs of outcomes which can be tabulated in a 2×2 table: (1, 1), (1, 0), (0, 1), and (0, 0), where the first element is the response to the new procedure, and the second element is the response to the standard procedure. Let a_k, b_k, c_k, and d_k be the observed frequencies of (1, 1), (1, 0), (0, 1), and (0, 0), respectively. Clearly, a_k and d_k are the concordant pairs, and b_k and c_k are the discordant pairs. Let x_{ik} denote the number of units in the kth cluster that respond to procedure i. That is, $x_{1k} = a_k + b_k$ and $x_{0k} = a_k + c_k$. The $n_k = a_k + b_k + c_k + d_k$ units in the kth cluster are assumed to be fixed for all K clusters, and denote $N = \sum_{k=1}^{K} n_k$ as the total number of units over all clusters. The 2×2 contingency in Table 1 presents the summary information for clustered matched-pair data summarized over all K clusters in which b and c are the frequencies of two types of discordant pairs, and p_{11}, p_{10}, p_{01}, and p_{00} are the respective response probabilities for the pairs (1, 1), (1, 0), (0, 1), and (0, 0).

To facilitate understanding of the data structure and notation, suppose there is a study with the objective to compare CT and PET scan images of patients with metastatic colorectal cancer. For each patient, there are 10 sites to be separately scanned using CT and PET. Hence, in this envisioned study, each patient is a cluster, each site of a patient is the unit, and the CT and PET scans are the diagnostic procedures.

Many active-controlled trials are designed to establish that the efficacy of an investigational (new) procedure is, to a pre-specified amount, no worse than an active comparator (standard procedure). Such trials are referred to as non-inferiority trials. Suppose that p_1 and p_2 are the true unknown success probabilities for the new and standard procedures. Let $\delta = p_1 - p_2$ be the true difference between the two procedures. We assume that a larger success probability denotes greater efficacy. Denoting the non-inferiority margin $\delta_0 < 0$, then for a non-inferiority trial, the alternative hypothesis is that the new procedure is not inferior to the standard by more than a pre-specified amount δ_0. This leads to the following hypothesis test for non-inferiority

$$\mathrm{H}_0 : p_1 - p_2 \le \delta_0; \quad \text{vs.} \quad \mathrm{H}_1 : p_1 - p_2 > \delta_0. \tag{1}$$

The test is based on the summary information of the clustered matched-pair binary data presented in Table 1.

2.1. Test by Durkalski et al. (2003)

Durkalski et al. (2003) proposed a non-inferiority test based on the robust method of moments as

$$Z_{\mathrm{D}} = \left[\sum_{k=1}^{K}\left(\frac{b_k - c_k}{n_k} - \delta_0\right)\right] \Big/ \left[\sum_{k=1}^{K}\left(\frac{b_k - c_k}{n_k} - \delta_0\right)^2\right]^{1/2}.$$