Computational Statistics and Data Analysis xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Non-inferiority test based on transformations for non-normal distributions☆

Santu Ghosh [a], Arpita Chatterjee [b], Samiran Ghosh [c,*]

[a] Department of Biostatistics and Epidemiology, Augusta University, United States
[b] Department of Mathematical Sciences, Georgia Southern University, United States
[c] Department of Family Medicine and Public Health Sciences, Wayne State University, United States

ARTICLE INFO

Article history:
Received 23 February 2016
Received in revised form 6 October 2016
Accepted 7 October 2016
Available online xxxx

Keywords:
Edgeworth expansion
Level error
Non-inferiority
Rank test
Three-arm trial

ABSTRACT

Non-inferiority trials are becoming very popular for comparative effectiveness research. These trials are required to show that the effect of an experimental treatment is not worse than that of a reference treatment by more than a specified margin. Hence non-inferiority trials are of great importance, when superiority cannot be claimed. A three-arm non-inferiority trial consists of a placebo, a reference treatment, and an experimental treatment is considered. However unlike the traditional choices, it is assumed that the distributions of the end points corresponding to these treatments are unknown and suggested test procedures for a three-arm non-inferiority trial based on monotone transformations in conjunction with a normal approximation. The resulting test procedures are flexible and robust. Theoretical properties of the proposed methods are also investigated. The performance of the suggested test procedures is compared to their counterparts using simulations. In terms of type I error and power, the proposed methods perform better than their counterparts in most cases. The usefulness of the proposed methods is further illustrated through an example.

1. Introduction

Non-inferiority trials comparing an experimental treatment with a reference treatment are becoming very popular as more and more effective treatments become available and fewer discoveries of new treatments are made. When clear superiority of an experimental treatment is not evident, the objective may be to demonstrate the non-inferiority of an experimental treatment compared to a reference treatment. A slightly less efficacious experimental treatment might be preferred to an established reference treatment due to its other benefits, such as: less toxicity, less costly, less debilitating, and easy to administer etc. In this kind of trial non-inferiority is established by showing that the efficacy of an experimental treatment is not less than that of a reference treatment by a specified small margin, also known as non-inferiority margin. Traditionally, these types of trials do not include a placebo group due to ethical reasons and are often termed as "two-arm" trials that include the reference and experimental groups. But due to the absence of placebo arm they are unable to establish direct proof of efficacy of the reference treatment over the placebo, and require external validation which is often questionable. Hence, two-arm non-inferiority trials without placebo lack the support of internal assay sensitivity.

☆ The R-code for our proposed methods is available in supplement A_R-code (see Appendix B).

* Corresponding author.
E-mail address: sghos@med.wayne.edu (S. Ghosh).

http://dx.doi.org/10.1016/j.csda.2016.10.004
0167-9473/

For detailed description of the problem see Hung et al. (2003) and D'Agostino et al. (2003). Assay sensitivity refers to the ability of a trial to distinguish between effective and ineffective treatments. This can be established by assuming the constancy condition, i.e. patient population in the current active control trial and in the past placebo control trial remains unchanged (ICH, 2000). However, in practice it is difficult to validate this constancy assumption. As a consequence, it is often suggested to include placebo group whenever it is feasible and ethically justifiable as addressed in several regulatory guidelines (ICHE10, 2000; EMEA, 2005).

Pigeot et al. (2003) and Koch and Rohmel (2004) considered three-arm non-inferiority trials with the inclusion of the placebo group. These three-arm non-inferiority trials are useful as they are free from some of the difficulties described above. When placebo is included, one approach to establish the non-inferiority of experimental treatment is to show that the ratio, $(\mu_E - \mu_P)/(\mu_R - \mu_P)$ is greater than $\theta \in (0, 1)$, where μ_E, μ_R, and μ_P are the mean effects corresponding to the experimental (E), reference (R), and placebo (P) groups respectively and θ is determined through clinical reasoning that considers knowledge about the diseases. For further details about the ratio method see Pigeot et al. (2003) and the references therein. A crucial assumption of some exiting methods (see Pigeot et al., 2003; Koti, 2007; Hasler et al., 2008) for testing non-inferiority hypothesis is that the endpoints are normally distributed. This assumption can lead us to unreliable conclusions when the normality of the endpoints is questionable. In this article, we consider a general situation without assuming any distributions corresponding to the endpoints. When the distributions of the endpoints are unknown, usually the non-inferiority testing can be performed based on a normal approximation using large sample theory. However, the true probability of incorrectly rejecting the null hypothesis (type I error) associated with an approximate test is never equal to the chosen nominal level for finite samples. Therefore the true type I error of an approximate test is either smaller or greater than the nominal level. If level error that is the difference between true type I error and nominal level is negative (or positive) then an approximate test is conservative (or anticonservative). A conservative test would lack power and an anticonservative test gives inflated type I error. Our analysis in Section 2 shows that magnitude of the level error associated the normal approximation test depends on the degree of skewness of underlying distributions and sample sizes associated with the three arms. Moreover, how large the sample size has to be also depends on the skewness of the distribution (Boos and Hughes-Oliver, 2000) and the hypothesized effect size.

Hence, the primary objective of this article is to reduce the level error associated to the non-inferiority test based on the normal approximation by removing the effects of skewness of underlying distributions. One way to achieve this goal is to convert the test statistic using a monotone transformation so that the resulting distribution of transformed test statistic is nearly symmetric and finally we can construct the critical point by inverting back the transformation. For one sample case, Hall (1992) investigated the effects of such monotone transformations to construct confidence intervals for the mean parameter, albeit not for the non-inferiority test setup. In a sense this article extends Hall's (1992) method to a three-arm trial in the non-inferiority context. The proposed tests are alternatives to the nonparametric non-inferiority test suggested by Munzell (2009) for three arm non-inferiority trials based on ranks. Munzell's method tackles the breakdown of the normality assumption via rank based test method that defines non-inferiority hypothesis in terms of relative treatment effects rather than usual mean effects for continuous data. Relative treatment effects are defined in terms of expectation of asymptotic rank transformation (page 3646, Munzell, 2009). Apart from lack of straight forward interpretation of relative treatment effects, our simulation studies show that Munzell's method tends to be conservative when the effect size $(\mu_R - \mu_p)$ is moderate to large (see Cohen, 1988).

The rest of the paper is organized as follows. Section 2 reviews three-arm non-inferiority trials. Section 3 develops our proposed test procedures for a three-arm non-inferiority trial based on transformations. In Section 4, we discuss a three-arm inferiority test based on ranks proposed by Munzell (2009). Simulation results are reported in Section 5. The analysis approach is illustrated in Section 6 using a data from a bone health study. A discussion follows in Section 7. For brevity, derivations of the theoretical results are provided in Appendix A.

2. Three-arm non-inferiority trials

To facilitate the discussion of a three-arm non-inferiority trial, let $X_{E,i}$, $X_{R,j}$, and $X_{P,k}(i = 1, \ldots, n_E, j = 1, \ldots, n_R, k = 1, \ldots, n_P)$ denote the observations corresponding to the treatment response in the experimental (E), reference (R), and placebo (P) groups, respectively. We assume that

$$X_{E,i} \overset{\text{i.i.d.}}{\sim} F_E(\mu_E, \sigma_E^2), \qquad X_{R,j} \overset{\text{i.i.d.}}{\sim} F_R(\mu_R, \sigma_R^2), \quad \text{and} \quad X_{P,k} \overset{\text{i.i.d.}}{\sim} F_P(\mu_P, \sigma_P^2),$$

where $\mu_l = E(X_l)$, $\sigma_l^2 = V(X_l)$, and F_l are absolute continuous as X_l are continuous random variables, $l \in \{E, R, P\}$. Without loss of generality, we assume that the large values of the endpoints represent large treatment effects. Conventionally, the hypothesis for a two-arm non-inferiority trial can be formulated in terms of difference between means

$$H_0 : \mu_E - \mu_R \leq -\Delta \quad \text{vs.} \quad H_1 : \mu_E - \mu_R > -\Delta, \tag{2.1}$$

where $\Delta > 0$ denotes the prespecified non-inferiority margin. Essentially, the rejection of H_0 supports the non-inferiority of the experimental treatment.

However, as we mentioned before even when $(\mu_E - \mu_R) > -\Delta$ for a fixed Δ, it is difficult to establish assay sensitivity.

Please cite this article in press as: Ghosh, S., et al., Non-inferiority test based on transformations for non-normal distributions. Computational Statistics and Data Analysis (2016), http://dx.doi.org/10.1016/j.csda.2016.10.004