Computational Statistics and Data Analysis 124 (2018) 168–180

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Response-adaptive treatment allocation for non-inferiority trials with heterogeneous variances

Wenfu Xu [a], Jingya Gao [a], Feifang Hu [b], Siu Hung Cheung [c,d,*]

[a] School of Statistics, Renmin University of China, Beijing, China
[b] Department of Statistics, George Washington University, WA, DC, USA
[c] Department of Statistics, The Chinese University of Hong Kong, Hong Kong, China
[d] Department of Statistics, National Cheng Kung University, Tainan, Taiwan

ARTICLE INFO

Article history:
Received 16 July 2017
Received in revised form 15 November 2017
Accepted 8 March 2018
Available online 20 March 2018

Keywords:
Non-inferiority trials
Adaptive design
Doubly adaptive biased coin design
Heterogeneous variances
Familywise error rate

ABSTRACT

In clinical studies, patients usually accrue sequentially. The response-adaptive design has been shown to be a valuable treatment allocation apparatus that skews the treatment allocation probabilities to achieve certain objectives such as reducing the number of patients who receive inferior treatments. The doubly adaptive biased coin design was successfully derived for the three-arm non-inferiority (NI) trial. For an NI study, an experimental treatment can be considered a possible substitute for the standard treatment if the loss of clinically tolerable efficacy is compensated by benefits such as the alleviation of side effects. Previous applications of the doubly adaptive biased coin design in NI trials were developed only for homogeneous treatment variances. However, it is worth to examine the more complicated, but nevertheless popular, scenarios in which the treatment variances are heterogeneous. The proposed treatment allocation scheme is superior when the treatment variances differ and remains very competitive when they are homogeneous. A clinical example is given for demonstrative purposes.

1. Introduction

If there are several treatments, the balanced design is a popular choice for treatment allocation. However, patients usually do not arrive all at once and a response-adaptive design could provide an ethical treatment allocation mechanism that skews allocation probabilities to favor better treatments. A detailed account of the development of response-adaptive designs can be found in Hu and Rosenberger (2006). The statistical information related to treatment effectiveness that accumulates at each point of the trial provides valuable information for assigning a treatment to the next incoming patient. As indicated by Montgomery (2017), in recent years, the awareness of the usefulness of response-adaptive designs has been growing in mainstream drug development. For example, Rugo et al. (2016) used a response-adaptive design for a breast cancer clinical study. In fact, adaptive designs are believed to be valuable tools for saving human lives in the field of global health. The outbreak of the West African Ebola epidemic in 2014 provides an example for which response-adaptive designs are being advocated (Montgomery, 2017; Berry et al., 2016; Lang, 2011).

The theme of this paper is related to the application of response-adaptive designs in three-arm non-inferiority (NI) trials. The growing popularity of NI trials in the past decade represents the drug development industry's increasing need to search for appropriate substitutes for standard (reference) treatments (Mauri and D'Agostino, 2017). Unlike superiority studies,

* Corresponding author at: Department of Statistics, The Chinese University of Hong Kong, Hong Kong, China.
E-mail address: shcheung@cuhk.edu.hk (S.H. Cheung).

https://doi.org/10.1016/j.csda.2018.03.005
0167-9473/

an NI trial is designed to declare that an experimental treatment can be considered a potential substitute for the reference treatment when the former is not inferior to the latter by a clinically significant margin (NI margin). The loss of efficacy, even though clinically insignificant and smaller than the NI margin, must be justified by compelling reasons, such as an easing of side effects, a reduction of the cost of treatment, or a less complicated treatment regimen (Beck et al., 2011; Fleming and Powers, 2008; Burger et al., 2011). Influential guidelines put forth by the U.S. Food and Drug Administrations (FDA) (FDA, 2016) and the European Medicines Agency (EMA) (EMA, 2006) provide templates for the design of NI trials and the proper selection of the NI margin.

A three-arm NI trial comprises the reference treatment R, the placebo P, and G experimental treatments $E_1, \ldots, E_G$. The existence of multiple experimental treatments, $G > 1$, is popular when the reference treatment is compared with different doses of a new drug (e.g., Valenzuela et al., 2016) or different combinations of several new drugs (e.g., Sundar et al., 2011).

In an NI trial, it is crucial to ensure assay sensitivity, which refers to the ability of the reference treatment to maintain an expected effect size (treatment effect of R as compared with that of P) comparable with that reported in previous placebo-controlled studies; if not, the validity of the trial will be questionable (FDA, 2016; EMA, 2006). The effect size of R is a vital quantity from which the NI margin is derived. For a two-arm NI trial without the placebo, confirmation of assay sensitivity may be a difficult task because the effect size of the reference treatment must be estimated from previous studies (Burger et al., 2011; Julious, 2011). However, if the placebo is included in an NI trial, one can directly verify the assay sensitivity of the NI trial. Hence, as advocated by the EMA (2006), a three-arm NI trial is preferred unless it has very serious adverse consequences for the placebo-treated patients. While the three-arm NI design is the gold standard, the inclusion of the placebo may not be feasible for ethical reasons. In such cases, the effect size of the active treatment over the placebo may have to be estimated using historical data and it has been recognized to be an intricate task due to possible bias induced by between-trial variability (Huang et al., 2009; Ellenberg and Temple, 2000).

For NI studies, a partially balanced design is frequently adopted in which the sample sizes of the experimental treatments and reference treatments are the same, whereas that of the placebo is halved (see for example Bossche and Vanderstraeten, 2015; Nauck et al., 2014). Let $n_{E_1}, \ldots, n_{E_G}, n_P, n_R$ be the sample sizes of $E_1, \ldots, E_G, P, R$, respectively. A partially balanced design then implies that $n_{E_1} = \cdots = n_{E_G} = 2n_P = n_R$. In other words, in terms of ratio, $(n_{E_1} : \cdots : n_{E_G} : n_P : n_R) = (2 : \cdots : 2 : 1 : 2)$. Based on the testing procedure of Kwong et al. (2012), an adaptive allocation scheme was proposed by Xu et al. (2017). Under their proposed adaptive treatment allocation scheme, more patients will receive the better treatments. However, their procedure assumes that the responses exhibit homogeneous variance across different types of treatments. As reported by Huang et al. (2015), when the variances are heterogeneous, the testing procedure of Kwong et al. (2012) is not appropriate because the familywise type I error rate (i.e., the probability of committing at least one type I error) may be highly inflated.

The objective of this paper is to modify the test procedure of Huang et al. (2015) and derive an adaptive treatment allocation algorithm that can be used for NI studies with heterogeneous outcomes. Our proposed treatment allocation scheme has two advantages. More patients undergo the better treatment than with the partially balanced design, and the test power has been demonstrated to remain at a very comparable level.

This paper is organized as follows. In Section 2, a brief discussion of the response-adaptive design is given, focusing on the doubly adaptive biased coin design, which has been shown to be very powerful among different classes of adaptive designs. In Section 3, NI test procedures for heterogeneous treatment variances are examined. Then, in Section 4, different allocation rules are defined. To compare the performance of different allocation rules, simulation results are presented in Section 5. A clinical example is provided in Section 6, and the last section contains several concluding remarks.

2. Response-adaptive randomization procedures

2.1. Doubly adaptive biased coin design

Sequential accumulation of patients in most clinical studies permits the use of a response-adaptive treatment allocation scheme that skews treatment assignment probabilities for incoming patients to achieve certain objectives. One major ethics-related purpose of using adaptive designs is to reduce the proportion of patients who undergo the less effective treatments. Another crucial objective is to maintain adequate testing power for the comparison of treatment efficacies.

Under a response-adaptive treatment allocation scheme, an incoming patient is assigned to a specific treatment with a probability computed based on the responses obtained thus far. Across various classes of adaptive designs, the target-driven response-adaptive design is a framework in which optimal criteria can be incorporated (Hu and Rosenberger, 2006; Tymofyeyev et al., 2007), which coincides with our intention to integrate the aforementioned objectives in the treatment allocation mechanism. For target-driven responses-adaptive designs, the doubly adaptive biased coin design (DBCD), first proposed by Eisele (1994) and Eisele and Woodroofe (1995) to compare two treatments, is a superior treatment allocation procedure. Important asymptotic properties that enable a more comprehensive understanding of the distributional probabilities of allocation proportions are derived in Hu and Zhang (2004). The essential feature of the DBCD is that its allocation mechanism is a function of (a) the proportion of subjects that is currently being assigned to each treatment and (b) the current estimate of the desired allocation proportion. The attractive property of DBCD is its ability to target desired allocation proportions. In addition, with the same limiting allocation proportions, the DBCD is shown to be preferable to other target-driven responsive-adaptive designs in terms of variance and test power (Hu and Rosenberger, 2003). For the Bayesian-type response-adaptive designs which are valuable for Phase II clinical trials, one could refer to the work of Berry and Eick (1995) and Cheng and Berry (2007).