Contemporary Clinical Trials Communications 8 (2017) 39–48

Contents lists available at ScienceDirect

Contemporary Clinical Trials Communications

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

Sensitivity of adaptive enrichment trial designs to accrual rates, time to outcome measurement, and prognostic variables

Tianchen Qian*, Elizabeth Colantuoni, Aaron Fisher, Michael Rosenblum, for the Alzheimer's Disease Neuroimaging Initiative[1]

Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA

ARTICLE INFO

Keywords:
Adaptive enrichment
Clinical trial design
Group sequential stopping
Multiple testing procedure
Treatment effect heterogeneity

ABSTRACT

Adaptive enrichment designs involve rules for restricting enrollment to a subset of the population during the course of an ongoing trial. This can be used to target those who benefit from the experimental treatment. Trial characteristics such as the accrual rate and the prognostic value of baseline variables are typically unknown when a trial is being planned; these values are typically assumed based on information available before the trial starts. Because of the added complexity in adaptive enrichment designs compared to standard designs, it may be of special concern how sensitive the trial performance is to deviations from assumptions. Through simulation studies, we evaluate the sensitivity of Type I error, power, expected sample size, and trial duration to different design characteristics. Our simulation distributions mimic features of data from the Alzheimer's Disease Neuroimaging Initiative cohort study, and involve two subpopulations based on a genetic marker. We investigate the impact of the following design characteristics: the accrual rate, the time from enrollment to measurement of a short-term outcome and the primary outcome, and the prognostic value of baseline variables and short-term outcomes. To leverage prognostic information in baseline variables and short-term outcomes, we use a semiparametric, locally efficient estimator, and investigate its strengths and limitations compared to standard estimators. We apply information-based monitoring, and evaluate how accurately information can be estimated in an ongoing trial.

1. Introduction

Adaptive enrichment designs involve pre-planned rules for restricting enrollment based on accrued data in an ongoing trial [1]. If, for example, a subpopulation shows evidence of no benefit of treatment, its enrollment could be stopped while the complementary subpopulation continues to be enrolled [2]. give an overview of statistical methods for adaptive enrichment designs, including the p-value combination approach [3–6]; the conditional error function approach [7]; and approaches using group sequential computations [8,9]. We use an adaptive enrichment design from the general class of [10]; which is based on the group sequential computation approach.

We consider trials where the primary outcome is observed a fixed amount of time after enrollment (called the delay); we refer to such outcomes as delayed responses. To illustrate, we use data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort study. We set the primary outcome to be a measure of change in severity of dementia symptoms from baseline to 2 year of follow-up described below; this is similar to the primary outcome in an ongoing, Phase 3 clinical trial of a drug to slow cognitive and functional decline from early Alzheimer's Disease [11]. Also recorded are baseline variables and the short-term outcome of change in severity of dementia symptoms measured at 1 year of follow-up.

To leverage prognostic information in baseline variables and the short-term outcome, we use a semiparametric, locally efficient estimator (called the adjusted estimator, for conciseness) from Ref. [12]. The adjusted estimator in a randomized trial is consistent under mild regularity conditions without requiring any parametric model assumptions. It has potential to improve precision, power, expected sample size, and trial duration when variables are sufficiently prognostic for the outcome. In trials with delayed responses, the adjusted estimator uses information from pipeline participants, i.e., enrollees whose primary outcome has not yet been observed.

We evaluate the sensitivity of Type I error, power, expected sample

* Corresponding author.
E-mail address: tqian2@jhu.edu (T. Qian).

[1] Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

http://dx.doi.org/10.1016/j.conctc.2017.08.003
Received 28 July 2016; Received in revised form 19 April 2017; Accepted 11 August 2017
Available online 16 August 2017

size, and trial duration to different design characteristics through simulation studies. Our simulation distributions mimic features of data from the Alzheimer's Disease Neuroimaging Initiative, and involve two subpopulations of interest based on a genetic marker. We investigate the impact of the following design characteristics: the accrual rate, the delay time of the short-term outcome and the primary outcome, and the prognostic value of baseline variables and short-term outcomes. The simulated trials involve multiple stages, and information-based monitoring is used to determine the time of interim analyses.

We focus on adaptive enrichment designs since their added complexity (compared to standard designs) may raise special concern about how sensitive their performance is to deviations from initial assumptions. Since statistics from multiple populations are used in the stopping rule and multiple testing procedure, changes to assumptions (which affect the joint distribution of these statistics) could have impacts that are not easy to predict a priori. This was observed, for example, when we varied the ratio of information accrual rates in the two subpopulations; in these cases the covariance structure of the test statistics is affected. This sometimes resulted in higher than 80% power for certain hypothesis tests, despite the fact that we used information-based monitoring (which in a single population trial design would maintain constant power at a given alternative). These results are described in Section 5.

In Section 2 we describe the ADNI study. In Section 3 we present notation. The simulation setup is given in Section 4. Section 5 presents simulation results, including the impact of prognostic baseline variables and a short-term outcome (Section 5.1), the impact of varying delay time (Section 5.2), and the impact of varying the accrual rates (Section 5.3) on the performance of the adaptive design. In Section 6 we discuss information accrual rates and how accurately these can be estimated in an ongoing trial. Section 7 concludes with discussions and future research directions.

2. Data example

Our simulations are based on distributions that mimic features of the data from the Alzheimer's Disease Neuroimaging Initiative (ADNI), an observational longitudinal study of cognitive impairment and progression to Alzheimer's disease. The ADNI was initiated in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of the study has been to test whether serial magnetic resonance imaging, positron emission tomography, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment and early Alzheimer's disease.[2] The Clinical Dementia Rating (CDR) scale is used to assess the severity of dementia symptoms and provides both a numeric global score ranging from 0 to 3, and a sum of boxes (SOB) score ranging from 0 to 18.

Our data come from 286 patients who entered the ADNI study with mild cognitive impairment (CDR 0.5 with a SOB score 2.5 or less) and who remained in the study for the full 12 months of follow-up. For conciseness, we refer to the CDR sum of boxes score as the CDR score. We define the primary outcome Y as the difference between the CDR score at baseline and at 2 years. We define the short-term outcome L as the difference between the CDR score at baseline and at 1 year. Let W denote the following five prognostic baseline variables: CDR score at baseline; age; $A\beta_{42}$ (a type of amyloid plaque involved in Alzheimer's disease progression); Alzheimer's Disease Association (ADA, 13 items) scale; and the Mini Mental State Examination (MMSE) score. We consider two distinct subpopulations defined by apolipoprotein E (APOE) $\varepsilon 4$ carrier status. Subpopulation 1 consists of those with no $\varepsilon 4$ alleles, and subpopulation 2 consists of those with at least one $\varepsilon 4$ allele. Among the 286 patients, 47% carry no APOE $\varepsilon 4$ alleles. We consider a hypothetical treatment whose goal is to delay the progression of disease.

[2] For up-to-date information, see www.adni-info.org.

3. Notation

When followed up completely, each participant i in the trial has full data vector $\mathbf{D}_i = (S_i, W_i, A_i, L_i, Y_i)$. We use the vector $\mathbf{D} = (S, W, A, L, Y)$ when referring to a generic participant. The variable $S_i \in \{1,2\}$ denotes the subpopulation that participant i belongs to; W_i denotes a vector of baseline variables; A_i denotes the treatment assignment indicator; L_i denotes the short-term outcome; and Y_i denotes the primary outcome. We assume that (S_i,W_i,A_i) are observed when participant i is enrolled, and that L_i and Y_i are observed at time d_L and d_Y, respectively, from the time of enrollment. Assume $d_L \le d_Y$. Each vector $\mathbf{D}$ is assumed to be an independent, identically distributed draw from an unknown distribution Q, with the only restriction being that A is randomized by design with equal probability of being 0 or 1, independent of S,W. The short-term outcome L can be any predefined measurement made after randomization. No assumptions on its relationship to Y are needed in order that our estimators (adjusted and unadjusted) are consistent and asymptotically normal [13].

For a given population, the average treatment effect is defined to be the difference between the population mean of the primary outcome under treatment ($A = 1$) versus under control ($A = 0$). Denote the average treatment effect in subpopulation 1, subpopulation 2, and the combined population by Δ_1, Δ_2, and Δ_0, respectively, where $\Delta_0 = E(Y|A = 1) - E(Y|A = 0)$ and for each subpopulation $s \in \{1,2\}$, $\Delta_s = E(Y|A = 1,S = s) - E(Y|A = 0,S = s)$. Let p_s denote the proportion of subpopulation s in the combined population, and we have $\Delta_0 = p_1\Delta_1 + p_2\Delta_2$. Define the null hypotheses

$$H_{01}: \Delta_1 \le 0; \quad H_{02}: \Delta_2 \le 0; \quad H_{00}: \Delta_0 \le 0,$$

which represent no average treatment benefit in subpopulation 1, subpopulation 2, and the combined population, respectively.

We quantify the prognostic value of W and L for explaining variance in the primary outcome Y for the combined population. Define the R-squared of W and R-squared of L as

$$R_W^2 = \frac{\mathrm{Var}\{E(Y|W)\}}{\mathrm{Var}(Y)}, \quad R_L^2 = \frac{\mathrm{Var}\{E(Y|L)\}}{\mathrm{Var}(Y)}. \tag{1}$$

R_W^2 represents the fraction of variance in Y explained by W. R_L^2 represents the fraction of variance in Y explained by L.

Using the ADNI study data, we approximated (1) to roughly determine how much of the variance of the outcome Y is explained by W or L. The empirical R_W^2 is computed as in (1), with $E(Y|W)$ estimated by a linear model with intercept and main terms W_3,W_4, and the variances are estimated by the empirical variance. (We use only W_3,W_4 in the working model for constructing the adjusted estimator; see Section 4.2.) A similar computation was done to obtain the empirical R_L^2. The resulting values are 0.20 and 0.48 for R_W^2 and R_L^2, respectively, for the combined population.

We also estimated R_W^2 and R_L^2 within each subpopulation, and found the prognostic values differ by subpopulation. The corresponding empirical R_W^2 is 0.30 for subpopulation 1 and 0.14 for subpopulation 2; the empirical R_L^2 is 0.44 for subpopulation 1 and 0.50 for subpopulation 2. This differential prognostic value by subpopulation impacts information accrual and power for the adjusted estimator as described in Section 5. In what follows, R_W^2 and R_L^2 refer to (1) for the combined population.

4. Simulation setup

4.1. Overview

Our goal is to evaluate the performance of an adaptive enrichment design with a delayed response when we vary the prognostic values in