Contemporary Clinical Trials 38 (2014) 157–162

Contents lists available at ScienceDirect

Contemporary Clinical Trials

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

Exploring the statistical and clinical impact of two interim analyses on the Phase II design with option for direct assignment ☆

Ming-Wen An [a,*], Sumithra J. Mandrekar [b], Martin J. Edelman [c], Daniel J. Sargent [b]

[a] Department of Mathematics, Vassar College, Poughkeepsie, NY, United States
[b] Department of Health Sciences Research, Mayo Clinic, Rochester, MN, United States
[c] Division of Hematology and Oncology, University of New Mexico Cancer Center, Albuquerque, NM, United States

ARTICLE INFO

Article history:
Received 21 January 2014
Received in revised form 10 April 2014
Accepted 12 April 2014
Available online 24 April 2014

Keywords:
Phase II
Direct assignment
Two interim analyses
Biomarker

ABSTRACT

Purpose: The primary goal of Phase II clinical trials is to understand better a treatment's safety and efficacy to inform a Phase III go/no-go decision. Many Phase II designs have been proposed, incorporating randomization, interim analyses, adaptation, and patient selection. The Phase II design with an option for direct assignment (i.e. stop randomization and assign all patients to the experimental arm based on a single interim analysis (IA) at 50% accrual) was recently proposed [An et al., 2012]. We discuss this design in the context of existing designs, and extend it from a single-IA to a two-IA design.
Methods: We compared the statistical properties and clinical relevance of the direct assignment design with two IA (DAD-2) versus a balanced randomized design with two IA (BRD-2) and a direct assignment design with one IA (DAD-1), over a range of response rate ratios (2.0–3.0).
Results: The DAD-2 has minimal loss in power (<2.2%) and minimal increase in T1ER (<1.6%) compared to a BRD-2. As many as 80% more patients were treated with experimental vs. control in the DAD-2 than with the BRD-2 (experimental vs. control ratio: 1.8 vs. 1.0), and as many as 64% more in the DAD-2 than with the DAD-1 (1.8 vs. 1.1). We illustrate the DAD-2 using a case study in lung cancer.
Conclusion: In the spectrum of Phase II designs, the direct assignment design, especially with two IA, provides a middle ground with desirable statistical properties and likely appeal to both clinicians and patients.

1. Introduction

There are typically three phases of clinical trials in the development of new treatments in oncology. In Phase I, the primary goal is to demonstrate the treatment's safety and identify a maximally tolerated dose in a small group of patients. In Phase II, the primary goal is to gain a better understanding of the treatment's safety and efficacy, to inform a go/no-go decision to Phase III testing through a single arm or a small randomized trial in a homogenous group of patients, with the trial size varying from <100 patients to as many as 300 patients. Finally, in Phase III, the primary goal is to compare the new treatment with the standard treatment in a large group of patients to demonstrate a clinical benefit. In this paper, we focus on designs for Phase II trials. Historically, Phase II trials in oncology have been single-arm trials where all patients receive the new

☆ Financial Support: Mandrekar and Sargent: National Cancer Institute grants no. CA-15083 (Mayo Clinic Cancer Center) and no. CA-25224 (North Central Cancer Treatment Group).

* Corresponding author at: Department of Mathematics, Vassar College, 124 Raymond Avenue, Poughkeepsie, NY 12604, United States. Tel.: +1 845 437 5981; fax: +1 845 437 7544.
E-mail address: mian@vassar.edu (M.-W. An).

http://dx.doi.org/10.1016/j.cct.2014.04.007
1551-7144/

treatment. More recently, randomization to either the new treatment or the standard treatment has become standard for Phase II trials [1]. In addition to being either single-arm or randomized trials, Phase II trials may also include planned interim analyses for safety and/or futility and/or efficacy. Bayesian methods have also been incorporated into Phase II design and analysis, allowing for flexible adaptive designs while requiring more resources than non-Bayesian designs.

Recently, a "hybrid" Phase II design was proposed that blends attractive features of existing designs, including randomization, adaptive accrual and simple implementation. The proposed design (An et al., [2]) is initiated as a randomized trial with equal allocation between two treatment arms. At the time of a single interim analysis (IA), after one-half of planned accrual, the design includes an option, based on pre-specified rules, of terminating accrual entirely (for efficacy or futility), continuing as a randomized trial, or continuing as a single-arm trial ("direct assignment"). An et al. ([2]) demonstrate that this design maintains comparable statistical properties in terms of power and type I error rate as those of a balanced randomized design with IA. Sensitivity analyses suggested minimal effects on statistical properties from the *timing* of single IA and *population shifts* in the case where the direct assignment option is adopted. This design with direct assignment option has ready applications to general settings of cytotoxic therapies.

In the setting of Phase II trials, many have argued that a single interim analysis may be inadequate, and that multiple looks improve both the statistical and ethical properties of the design. Further, a trial with interim analysis may terminate early, potentially resulting in cost savings and earlier delivery of effective treatments to patients (e.g. [3–6]). Additionally, the design of [2] with a single IA has mostly been studied so far for its statistical properties. In this paper, therefore, we study the design of [2] by incorporating two IA, focusing more on the clinical relevance of this design as it relates to number of patients treated on the experimental regimen, and illustrate with an example using a real trial. Specifically, we first review the statistical properties of a design with the option for direct assignment and with two IA (DAD-2) but then shift our focus to the impact on sample size and proportion of patients receiving experimental vs. control treatment associated with a DAD-2 relative to both a balanced randomized design with two IA (BRD-2) and a design with option for direct assignment and one IA after 1/2 accrual (DAD-1). We then illustrate the DAD-2 with an example from a non-small cell lung cancer trial in which a retrospective analysis identified a treatment benefit in a subgroup of patients with elevated Cox2 enzyme levels [7].

2. Methods

2.1. Design framework

We consider a binary outcome. We specify two interim analyses (IA) after 1/3 and 2/3 of planned accrual. At the first IA (i.e. IA-1), there are *4 options*: stop for efficacy, continue with direct assignment, continue with randomization, or stop for futility. If direct assignment is adopted at IA-1, then the trial continues with direct assignment to the end without a second IA, enrolling the planned accrual to active treatment for the remainder of the trial (i.e. $1/2 \times (1/3 + 1/3) =$ one-third of the total planned accrual). Otherwise if randomization continues at IA-1, then at the second IA (i.e. IA-2), there are again *4 options*: stop for efficacy, continue with direct assignment, continue with randomization, or stop for futility.

Extending the framework of [1], the IA decisions are based on the p-values from a test comparing the experimental to control treatment using cumulative data. In particular, the first IA (IA-1) uses data from Stage I, the second IA (IA-2) uses data from Stages I and II, and the final analysis uses data from all available stages. We specify the overall type I error rate (α) and power (1-β); and the expected response rates in control ($p^{control}$) and in treated (p^{treat}) patients, with an associated treatment effect (response rate ratio, $RRR = p^{treat}/p^{control}$). The maximum sample size (N) is calculated based on α, β, and the expected treatment effect size, using a one-sided two sample test of proportions assuming 1:1 randomization and O'Brien-Fleming (OF) stopping rules for efficacy and futility. At any given IA, the cut-off boundary for deciding between direct assignment and randomization is taken to be the cut-off boundary for efficacy in the subsequent IA (or final analysis, in the case of the last IA). An advantage of using the known framework of OF stopping rules is that this design can be readily implemented using existing software.

As an example of the cut-off boundaries, consider the case of two IA, $\alpha = 0.10$, and $\beta = 0.8$. We specify the cut-off for deciding between direct and randomized assignments at IA-1 to be 0.043, which corresponds to the cut-off for efficacy at IA-2, and the cut-off for deciding between direct and randomized assignments at IA-2 to be 0.087, corresponding to the cut-off for efficacy at the final analysis (Tables 1A, 1B).

2.2. Simulation settings

We conducted simulation studies to evaluate the operating characteristics of the designs. In particular, we first recorded the probabilities of IA decisions (estimated by the proportion of trials adopting the decisions) for the DAD-2, to understand basic properties of the DAD-2. Then we compared the power and type I error rates for a DAD-2 with those for a BRD-2. Finally we compared the following accrual outcomes of a DAD-2 with those of a BRD-2 and a DAD-1, under the null and alternative hypotheses: the expected total accrual and ratio of the number of patients treated with experimental versus control treatment if the experimental treatment is truly effective. Typically in Phase II studies, minimizing expected accrual is of primary interest, yet it might be simultaneously of interest to maximize the expected ratio of patients treated with experimental versus control treatment. To quantify this trade-off, we additionally considered the following composite measure: *Expected Accrual/Expected Ratio of Number Treated with Experimental* vs. *Control*, where lower values reflect lower expected accrual and proportionally more patients treated with experimental (vs. control), and are therefore desirable under the alternative hypothesis. We generated 6000 trials and specified $\beta = 0.80$, and $\alpha = 0.10$ and 0.20. Sample size calculations were based on a control response rate of $p^{control} = 0.20$ and experimental treatment response rates of $p^{treat} = 0.40$, with an associated response rate ratio (RRR) of 2.0 (Tables 1A, 1B). We additionally considered other treatment effects by varying the experimental treatment response rates according to RRR = 2.1 to 3.0. All tests of equal proportions were based on a normal-approximation z-test without correction for continuity.