A 2-Stage Phase II Design with Direct Assignment Option in Stage II for Initial Marker Validation

Ming-Wen An¹, Sumithra J. Mandrekar², and Daniel J. Sargent²

Abstract

Biomarkers are critical to targeted therapies, as they may identify patients more likely to benefit from a treatment. Several prospective designs for biomarker-directed therapy have been previously proposed, differing primarily in the study population, randomization scheme, or both. Recognizing the need for randomization, yet acknowledging the possibility of promising but inconclusive results after a stage I cohort of randomized patients, we propose a 2-stage phase II design on marker-positive patients that allows for direct assignment in a stage II cohort. In stage I, marker-positive patients are equally randomized to receive experimental treatment or control. Stage II has the option to adopt "direct assignment" whereby all patients receive experimental treatment. Through simulation, we studied the power and type I error rate of our design compared with a balanced randomized two-stage design, and conducted sensitivity analyses to study the effect of timing of stage I analysis, population shift effects, and unbalanced randomization. Our proposed design has minimal loss in power (<1.8%) and increased type I error rate (<2.1%) compared with a balanced randomized design. The maximum increase in type I error rate in the presence of a population shift was between 3.1% and 5%, and the loss in power across possible timings of stage I analysis was less than 1.2%. Our proposed design has desirable statistical properties with potential appeal in practice. The direct assignment option, if adopted, provides for an "extended confirmation phase" as an alternative to stopping the trial early for evidence of efficacy in stage I. Clin Cancer Res; 18(16); 1–9. ©2012 AACR.

Introduction

Targeted therapies play an increasingly important role in the medical treatment of oncology patients. Biomarkers are a critical component of targeted therapies, as they can be used to identify patients who are more likely to benefit from a particular treatment. A biomarker is a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention (1). A biomarker can be useful for treatment either through the estimation of disease-related trajectories (i.e., prognostic signatures) and/or prediction of patient-specific benefit to a particular treatment (i.e., predictive tool). Clinical trials to validate biomarkers represent a crucial step in translating basic science research into improved clinical practice.

Several phase III prospective clinical trial designs for biomarker-directed therapy have been previously proposed; we briefly describe 3 such design classes. All-comers designs enroll and randomize all patients, thus allowing for the testing of an interaction between treatment and marker, thereby facilitating validation of the marker’s predictive ability. All-comers designs can be broadly classified into 2 types that differ by randomization schema and analytical strategies: (i) the marker-stratified design (2) randomizes all patients, stratified by marker status, followed either by a test for interaction or separate tests for efficacy in each marker group, and (ii) the sequential testing strategy design (3) randomizes all patients to treatment but allows for testing a treatment effect in both the overall population and the marker-positive group. The second set of designs, adaptive designs, refers to a class of designs that adapt design parameters during the course of a trial based on accumulating data. These adaptations can take the form of dropping a particular marker-defined subgroup based on initial futility (4) or changing randomization probabilities within marker subgroups (5). The enrichment design (6) aims to understand the safety and clinical benefit of a treatment in a subgroup of patients defined by a specific marker status. All patients are screened for the presence of a marker, and only those patients with certain molecular features are included in the trial and randomized to treatment or control. These designs are particularly pertinent when there is compelling preliminary evidence that treatment benefit, if any, will be restricted to a subgroup of patients who express a specific molecular feature.

Randomization remains the gold standard for testing the predictive ability of a marker and the only way to ensure an unbiased estimation of a treatment effect. As such, all of the...
designs mentioned above appropriately randomize patients to receive treatment or not throughout the entire duration of the trial. However, are there certain situations in which it may not be necessary to randomize until the end of a trial? For example, consider the setting of a binary biomarker and a targeted therapy believed to have a beneficial effect in the marker-positive group and an unknown effect in the marker-negative group. One example of such a situation is sorafenib in combination with ganetespib in patients with non–small cell lung cancer (NSCLC), whose tumors harbor K-ras mutations. Preliminary evidence suggests that patients with K-ras mutations treated with single-agent sorafenib achieve objective response, contrary to what is observed in the general NSCLC patient population (7, 8). Preclinical data also suggest activity of ganetespib in K-ras–mutant NSCLC cell lines (9). Suppose there is evidence for cytotoxic synergy between sorafenib and ganetespib. It may, therefore, be of interest to study the benefit of combination therapy versus single-agent therapy in K-ras–mutant NSCLCs with the hypothesis that there will be enhanced activity associated with the combination therapy. To investigate this hypothesis, we could adopt one of the above designs depending on strength of preliminary evidence, marker prevalence, and assay properties. However, suppose in the course of the trial, promising, although not definitive, evidence of efficacy associated with the combination therapy becomes apparent. In that case, a design that allowed for a switch to direct assignment (i.e., all patients receive the combination therapy in our example) after an interim analysis could be considered. Such a design would likely have clinical and patient appeal, while still providing information on treatment benefit. Colton (10; 11) first proposed several designs that involved directly assigning patients to 1 of 2 treatment arms. He considered a cost function approach to clinical trial design for comparison of 2 treatments, whereby the choice of design parameters was driven by minimization of the cost associated with treating patients. In his class of designs, the second stage always was a direct assignment.

In this article, we propose a phase II design that includes an option for direct assignment to the experimental treatment, when there is promising, but not definitive, evidence of a treatment benefit at the end of an initial randomized stage of the trial. Specifically, we propose a 2-stage enrichment design (i.e., screen all patients for marker status but only enroll and randomize a particular marker subgroup, e.g., marker-positive or marker-negative for a binary marker) that may stop early for futility or efficacy. While we focus on an enrichment design to illustrate our proposal, the design can easily be extended to include the other marker group(s). Furthermore, although we motivate this proposed design in the context of a targeted therapy, we note that this design could in fact be used for a cytotoxic agent, as no decisions are made with respect to the targeted versus an overall hypothesis. If the trial does not stop early for futility after stage I, then in stage II, the trial can continue in 1 of 2 ways: (i) continue with randomization as in stage I or (ii) switch to “direct assignment,” in which all patients are given the experimental treatment. The decision for direct assignment is based on observing promising, but not definitive, results indicating treatment benefit in stage I. Through simulation, we study the empirical power and type I error rate of our design compared with a balanced randomized design for different treatment effect sizes.

Methods

We consider a binary outcome, for example, objective response and progression-free status (PFS) at a predefined time point, or percentage of change in an expression level dichotomized as high or low. Figure 1 gives a schematic of our proposed design, which we describe in this section.

Design framework

Patients who meet the trial eligibility criteria with a valid test result for the marker (M) and who belong to the specific marker group (say, marker-positive) are randomized (1:1) to receive either an experimental treatment or control. We assume that a planned interim analysis occurs after half of the patients are accrued. The interim analysis decisions are based on the P values from a test comparing the experimental treatment to the control using stage I data. A decision is made to stop early for futility, continue with 1:1 randomization, continue with direct assignment, or stop early for efficacy. If accrual continues to stage II, then at the end of stage II, a decision is made about treatment efficacy. This decision is based on the P value from a test comparing the experimental treatment versus control using accumulated data from both stages. Of note, if the trial continues into stage II with direct assignment, then the final comparison is between all the patients treated in the experimental treatment (from stages I and II) and the control (from stage I) patients.

Design parameters and assumptions

We specify the overall type I error rate (α) and power (1 − β) and the expected response rates (RR) in control (p0) and in treated (p1) patients, with associated treatment effect [response rate ratio (RRR) = p1/p0]. As we are assuming the preliminary evidence suggests strong benefit of the new treatment in the marker-positive group, we specify α to be 1-sided, as is common for phase II trials.

The maximum sample size (N) is calculated on the basis of α, β, and the expected effect size using a 2-sample test of proportions for a 2-stage design, with a 1:1 randomization and O’Brien–Fleming (OF) stopping rules, allowing for early stopping for efficacy (c1) or futility (c2) after stage I. In addition, the decision to adopt direct assignment (vs. continuing with balanced randomization) in stage II uses the same boundary as for concluding efficacy at the end of the trial. That is, if d is the OF boundary for concluding efficacy at the end of the trial, then we set the boundary for deciding between adopting direct assignment (vs. continuing with balanced randomization) for stage II to be d. If the trial continues to stage II with direct assignment, then we direct assign (i.e., enroll) half of the planned stage II sample.
size. Thus, the effective trial accrual depends on the interim analysis decisions: If the trial stops early for either efficacy or futility, then the total accrual is $N/2$; if the trial continues with randomization, then total accrual is $N$; and if the trial continues with the direct assignment option, then total accrual is $3N/4$.

Simulation study: parameter values

We conducted a simulation study to evaluate the performance of the design in terms of power and type I error rate. We generated 6,000 trials and specified $\alpha = 0.10$ and $\beta = 0.20$ and $\alpha = 0.20$ and $\beta = 0.80$ with associated RRRs of 2.00, 2.25, 2.50, and 3.00, respectively. These values were chosen to be consistent with what is commonly targeted in phase II oncology trials. For each of the 2 $\alpha$ values, we used a fixed sample size that was calculated on the basis of a RRR of 2.00 and $\beta = 0.80$, assuming a 2-stage balanced randomized design based on a 2-sample test of proportions using OF stopping rules. In particular, for $\alpha = 0.10$, $\beta = 0.80$, and RRR $= 2.00$, the maximum total sample size is $N = 101$; and for $\alpha = 0.20$, $\beta = 0.80$, and RRR $= 2.00$, the maximum total sample size is $N = 65$ (Table 1). Then, using the same sample sizes and for each of $\alpha = 0.10$ and 0.20, we calculated power and type I error rate for the different RRRs as follows. Under the alternative hypothesis $H_A: \rho_{\text{control}} \neq \rho_{\text{treat}}$, we calculated empirical power as the proportion of the 6,000 simulated trials in which the null hypothesis was (correctly) rejected using significance levels $\alpha = 0.10$ and 0.20. Under the null hypothesis $H_0: \rho_{\text{control}} = \rho_{\text{treat}}$, we calculated the empirical type I error rate as the proportion of the 6,000 simulated trials in which the null hypothesis was (incorrectly) rejected using significance levels $\alpha = 0.10$ and 0.20. All tests of equal proportions were based on the normal-approximation $z$ test without correction for continuity. We compared our design against a balanced randomized design.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>Early efficacy ($c_1$)</th>
<th>Early futility ($c_2$)</th>
<th>Overall efficacy ($d$)</th>
<th>Max $N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>0.0200</td>
<td>0.4566</td>
<td>0.0940</td>
<td>101</td>
</tr>
<tr>
<td>0.20</td>
<td>0.0699</td>
<td>0.5765</td>
<td>0.1803</td>
<td>65</td>
</tr>
</tbody>
</table>

Figure 1. Design schematic (among marker-positive patients only). Square brackets indicate number of patients enrolled at the given stage. $R$ = randomize; $N$ = total number of patients allocated at start of trial; $p_1 = P$-value based on stage I patient data; $c_1$, $c_2$, and $d$ are OF stopping boundaries ($d$ is the boundary for concluding overall trial efficacy, and is also used to decide between adopting the direct assignment option versus continuing with randomization in stage II).
Sensitivity analyses

In addition to the parameter values specified above, we conducted sensitivity analyses to explore the effects on type I and II error of (i) population shift, (ii) timing of stage I analysis, and (iii) unbalanced randomization.

Population shift is a potential concern of our design as it is with other designs such as outcome-adaptive randomization or single-arm phase II trials that use historical controls. Specifically, with the direct assignment option, patients who enroll during stage II know they will receive the experimental treatment and so may differ fundamentally with respect to outcomes from patients in stage I, with whom comparisons are being made in the final analysis. To investigate this concern and the potential effect on the type I error rate, we conducted a sensitivity analysis in which we hypothesized the stage II RR for the experimental treatment group is shifted by an amount (δ) from the stage I RR, if direct assignment is used for stage II. We considered shift values of δ = ±0.025, 0.05, 0.10, 0.20, and +0.30 for α = 0.10. The shift values of δ = +0.30, ±0.20, and ±0.10 were included to examine what may happen in the unlikely case of an extreme population shift. We compared the empirical type I error rate and power under these possible population shifts with the corresponding rates under no population shift (δ = 0). In particular, we obtained the type I error rate by assuming the null hypothesis in stage I (i.e., RR = 1.0 (Null)) and then in stage II if the direct assignment option was adopted, we specify RR = 0.20. To obtain the power, we assumed the alternative hypothesis in stage I (i.e., RR = 2.0, p\textsuperscript{control} = 0.20, p\textsuperscript{treat} = 0.40), and in stage II if the direct assignment option was adopted, we specify p\textsuperscript{treat} = 0.40 + δ.

A second sensitivity analysis addresses the timing of stage I analysis. In the proposed design, we consider a single analysis at half (f = 0.50) accrual. As a sensitivity analysis, we considered analyses at one third (f = 0.33) and two thirds (f = 0.67) of accrual.

Finally, a variation of our proposed design would be to allow for an unbalanced (e.g., 4:1 or 5:1 randomization for experimental treatment vs. control) design in stage II, when stage I results are promising but not definitive to stop trial for evidence of efficacy. We refer to this as the "unbalanced randomized design." This would be an intermediate option between our proposed design and an adaptive design. We explored this possibility with the option for 4:1 randomization, instead of direct assignment, in stage II, specifying an overall alpha level of α = 0.10, RR = 2.00, and with interim at half accrual.

Results

The results from the simulation study are summarized in Tables 2 and 3. Table 2 summarizes the distribution of interim analysis decisions across the 6,000 trials: the proportion of trials that stopped early for either efficacy or futility, or continued with direct assignment or balanced randomization, under both the null (RR = 1.0) and the alternative (RR = 2.0) hypotheses. Table 3 summarizes the operating characteristics: power and type I error rate. In general, the operating characteristics of our proposed design with the option for direct assignment are comparable with a balanced randomized design. For α = 0.10 and the range of RRR values considered, the power varied from 79.3% to 99.3% for a design with the direct assignment option, versus from 80.6% to 99.7% for a balanced randomized design. For α = 0.20, the power varied from 78.0% to 98.7% and from 79.3% to 99.0% for a design with the direct assignment option and a balanced randomized design, respectively. The maximum loss in power associated with implementing a design with the direct assignment option versus a balanced randomized design is 1.7% (88.0% vs. 86.3%) at α = 0.20 and RR = 2.25. The effect of the direct assignment option on power decreases with larger treatment effects. For instance, at α = 0.10, the decrease in power for detecting a RRR of 2.0 is 1.3% (80.6% vs. 79.3%) for a balanced randomized design and design with the direct assignment option, respectively, compared with 0.4% (99.7% vs. 99.3%) for detecting a RRR of 3.0. The type I error rate increased slightly for a design with the direct assignment option (11.5% vs. 10.4%) for a balanced randomized design at the nominal alpha level of α = 0.10, and 21.8% vs. 19.7% for a balanced randomized design at α = 0.20.

Sensitivity analyses

Population shift. Type I error rate as a function of population shift size is presented in Table 2. Because we only
assumed a population shift in the case when the direct assignment option is adopted, we compared properties of a design with the direct assignment option across different population shift values; the balanced randomized design would not be affected by a population shift and thus is not included in comparisons. For $\alpha = 0.10$ (Fig. 2, blue line and circles), the empirical type I error rate of a design with the direct assignment option, assuming no population shift (i.e., $\delta = 0$) was 11.5%. For small, but plausible, positive shifts (i.e., $\delta = +0.025$ and $+0.05$), there was minimal effect on empirical type I error rate (type I error rates: 11.9% and 12.9%). As expected, the type I error rate decreased with negative population shifts (reaching a minimum of 6.4% for a shift of $\delta = -0.20$) and increased

<table>
<thead>
<tr>
<th>Shift in experimental treatment group response rates in stage II ($\delta$)</th>
<th>Type I error rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.2</td>
<td>10.4</td>
</tr>
<tr>
<td>-0.1</td>
<td>14.0</td>
</tr>
<tr>
<td>0.0</td>
<td>17.1</td>
</tr>
<tr>
<td>0.1</td>
<td>20.2</td>
</tr>
<tr>
<td>0.2</td>
<td>23.3</td>
</tr>
<tr>
<td>0.3</td>
<td>26.4</td>
</tr>
</tbody>
</table>

NOTE: Trial simulations ($N = 6,000$); $1 - \beta = 0.80$; $p_{\text{control}} = 0.20$, stage I analysis at half of accrual ($f = 0.50$); total sample size planned $N = 101$ and 65 for $\alpha = 0.10$ and 0.20, respectively.

Table 3. Comparison of operating characteristics for a design with direct assignment option versus a balanced randomized design

<table>
<thead>
<tr>
<th></th>
<th>$\alpha = 0.10$</th>
<th>$\alpha = 0.20$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Balanced randomized design</td>
<td>Design with direct assignment option</td>
</tr>
<tr>
<td>Power (%)</td>
<td>2.00</td>
<td>80.6</td>
</tr>
<tr>
<td></td>
<td>2.25</td>
<td>90.5</td>
</tr>
<tr>
<td></td>
<td>2.50</td>
<td>96.5</td>
</tr>
<tr>
<td></td>
<td>3.00</td>
<td>99.7</td>
</tr>
<tr>
<td>Type I error rate (%)</td>
<td>-</td>
<td>10.4</td>
</tr>
</tbody>
</table>

Figure 2. Type I error rate as a function of shift in RR for experimental treatment group, only if direct assignment is adopted (vertical dotted line at $\delta = 0$, i.e., no population shift). Nominal $\alpha$-level = 0.10 (blue line and circles) and 0.20 (red line and triangles).
with positive population shifts (reaching a maximum of 14.6% for the unlikely shift of $\delta = +0.30$). The increase in type I error rate at a nominal $\alpha$-level of 0.10 was at most 3.1% (from 11.5% to 14.6%, associated with a shift of $\delta = +0.30$). For a nominal $\alpha$-level of 0.20 (Fig. 2, red line and triangles), the empirical type I error rate of a design with the direct assignment option, assuming no population shift (i.e., $\delta = 0$) was 21.5%. The maximum increase in type I error rate was 5% (associated with a shift of $\delta = +0.30$). Although the increase in type I error rate is, as expected, greater for higher nominal $\alpha$-levels, it is still relatively minor.

We also considered the effect of population shift on power. For $\alpha = 0.10$, the empirical power of a design with the direct assignment option assuming no population shift was 79.3%. The power substantially decreased with extreme negative population shifts (63.0% for a shift of $\delta = -0.20$) and minimally increased with positive population shifts (79.2%, 80.6%, and 82.3% for shifts of $\delta = +0.025$, +0.05, and +0.30, respectively).

**Timing of stage I analysis.** Figure 3 displays power as a function of timing of stage I analysis (Fig. 3, red line and triangles; overall $\alpha$-level ($\alpha = 0.10$ and 0.20; by rows); and RRR (RRR = 2.0, 2.25, 2.5, and 3.0; on x axis).
Two-Stage Phase II Design with Direct Assignment Option

Assignment option compared with a balanced randomized design. For interim after one-third accrual \( (f = 0.33) \), the loss in power (over RRR = 2.0, 2.25, 2.5, 3.0; and \( \alpha = 0.10, 0.20 \)) ranged from 1.0% to 2.8%; and for interim after two-thirds accrual \( (f = 0.67) \), the loss in power ranged from 0.08% to 1.2%. We also considered type I error rate. In terms of type I error rates, for interim after one-third accrual, the increase in type I error rate was 3.0% and 5.0% for \( \alpha = 0.10 \) and 0.20, respectively; and for interim after two-thirds accrual, the increase was 0.6% and 0.4% for \( \alpha = 0.10 \) and 0.20, respectively.

Unbalanced randomization. The loss in power (to detect a RRR of 2.0 at \( \alpha = 0.10 \)) associated with an unbalanced randomized (4:1) design compared with a balanced randomized design is 0.4% (80.2% vs. 80.6%). This is marginally less than the loss in power associated with a design with the direct assignment option compared with a balanced randomized design (1.3%; 79.3% vs. 80.6%). This result is expected, as unbalanced randomized designs are less powerful than balanced randomized designs, and the design with a direct assignment option is the most extreme case of an unbalanced randomized design. There is nearly no effect on type I error rate at \( \alpha = 0.10 \) associated with the unbalanced randomized design (4:1) compared with a balanced randomized design (both 10.4%).

Discussion

Testing of novel targeted agents with predictive biomarkers in clinical trials requires new strategies. In the case of a targeted agent with promising efficacy, it is desirable to allow as many patients as possible to the new treatment. Several studies have been conducted to investigate factors associated with participation in cancer clinical trials. Among patients with breast cancer, one of the principle drawbacks of participating in clinical trials was the potential for receiving less effective treatment (12). The direct assignment phase of our proposed design removes the uncertainty about treatment assignment (associated with randomized treatment assignment) when an initial promising benefit has been observed. Such certainty about receiving treatment received in a clinical trial is only possible with a direct assignment. Even outcome adaptive designs with reweighted randomization probabilities based on accumulating outcome data do not provide such a level of certainty.

Colton (10, 11) previously proposed several designs that involved directly assigning patients to 1 of 2 treatment arms, which differs from our proposed design in 2 key ways. First, the parameters in his design were driven by minimization of the cost associated with treating patients, which is in contrast to our objectives, which are to control for power and type I error rates. Another key distinction is that in the Colton designs, the second stage always goes to direct assignment. In our proposed design, a data-derived decision is made at interim based on stage I data regarding whether or not to direct assign in stage II. We emphasize that our design does not always switch to direct assignment in stage II. Only when there is convincing, though not definitive evidence, from stage I does the trial design switch to direct assignment. In the absence of such evidence, the trial continues with randomization in stage II. As such, the direct assignment option provides an "extended confirmation phase" as an alternative to stopping the trial early for efficacy, which may help to avoid possibly prematurely launching into a phase III trial. This latter benefit of our design is especially important, given the high failure rates of phase III trials (13).

Our simulation studies show that the addition of a direct assignment option, in the absence of extreme population shift, results in similar statistical operating characteristics relative to a balanced randomized design. This relatively minimal loss in power (Table 3) is in part explained by the probability that a trial adopts the direct assignment option, as it is only with direct assignment where power is lost relative to balanced randomization. Table 2 presents the proportion of the 6,000 simulated trials that adopted direct assignment under both the null (RRR = 1.0) and the alternative hypotheses (RRR = 2.0; Table 2). For \( \alpha = 0.10 \) and 0.20, the probability that a trial adopts direct assignment is 0.30 and 0.22, respectively, under the alternative, which is meaningfully high to make this design with a direct assignment option attractive, yet, as we have found in our simulation, not too high so as to lead to a substantial loss in power.

On the basis of our sensitivity analyses, we believe that the risk associated with false trial conclusions associated with a plausible population shift in our proposed design is minor. In addition, the timing of the stage I analysis seems to have little effect on the loss in power. A design that allows for extreme unbalanced randomization (e.g., 4:1) in stage II is slightly more powerful than a design with a direct assignment option in stage II, and could be considered as an alternative to our design. Nevertheless, the direct assignment design has marginally little loss of power and provides the advantage of certainty of treatment received.

One could imagine various extensions to our proposed design. For example, a design could plan for 2 interim analyses, for example, at one-third and two-thirds accrual, but with the option for direct assignment only available after the second interim analysis (i.e., the first interim analysis is for futility only). Extensions to our design to incorporate continuous or time-to-event endpoints are also a possibility. Furthermore, whereas we used an enrichment design in the marker-positive group as a framework to our design, it may be of interest to extend to a situation involving multiple marker group(s). A final possibility, if ethical, would be to use a placebo control throughout the trial, in which both the patient and the treating physician are blinded to the treatment assignment. In this case, if in stage II the direct assignment option is adopted, then while the randomization would continue at the patient level (so the patient and the treating physician are unaware that the design has switched to a direct assignment), the placebo arm would be discontinued and all patients would be assigned to the experimental arm.

Biomarker-based clinical trials are a rapidly developing field. In the spectrum of designs proposed to date, with
adaptive designs comprising one end and fixed balanced randomized designs the other, our proposed design with a direct assignment option in stage II provides a possible middle ground. It provides (relative) logistic simplicity, whereas still allowing adaptation based on accumulated data, and maintaining desirable statistical properties. We believe that this design may have a better appeal to clinicians and patients, and could in fact be used for a cytotoxic agent, as no decisions are made with respect to the targeted versus an overall hypothesis.

Appendix

Interim (stage I) decision based on number of responses

The interim decision is based on the cutoff points of the stage I P values. However, as we use a normal-based z test, we can translate these decision cutoff points into an approximate number of excess responses needed in the experimental arm relative to the control arm. General expressions are derived later, but first consider an example with n = 50 per arm in stage I, overall type I error rate $\alpha = 0.10$, the null hypothesis that $p^{\text{control}} = p^{\text{treat}} = 0.2$, and the OF stopping boundaries given in Table 1. In this case, if the number of excess responses in the experimental arm relative to the control arm is at least 8, then the trial stops early for efficacy; if the excess is either 6 or 7, then the trial continues into stage II with direct assignment; if the excess is between 1 and 5, then the trial continues to stage II with randomization; otherwise the trial stops early for futility. In other words, to stop for futility, the number of responses on the experimental arm exceeds those on the control arm by at most 1.

Here, we derive general expressions for translating the stage I decision stopping rules into number of responses. At stage I analysis, we test the null hypothesis that $p^{\text{control}} = p^{\text{treat}}$ using data collected in stage I. In particular, let $n$ denote the stage I sample sizes for each group and $P$ denote the null hypothesized common RR, then the test statistic calculated under the null hypothesis is

$$
\text{I.S.} = \frac{\hat{p}_1 - \hat{p}_2}{\text{SE}(\hat{p}_1 - \hat{p}_2)} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{2p(1-p)/n}} = \frac{X_1 - X_2}{\sqrt{2np(1-p)}}
$$

where $\hat{p}_1, \hat{p}_2$ denotes the observed RRs in the treatment and control groups, respectively; and $X_1, X_2$ denotes the observed number of responses in the treatment and control groups, respectively. At the end of stage I, we “stop the trial early for efficacy” when the stage I P value is less than c1, which corresponds to

$$
X_1 - X_2 > z_{1-c1} \sqrt{2np(1-p)},
$$

where $z_c$ denotes the $(1 - p) \times 100$th percentile of a standard normal distribution for $0 \leq p \leq 1$. Similarly, we “continue to stage II with direct assignment” when the stage I P value is between c1 and d, which corresponds to

$$
z_{1-d} \sqrt{2np(1-p)} < X_1 - X_2 < z_{1-c1} \sqrt{2np(1-p)};
$$

“continue to stage II with randomization” when the stage I P value is between d and c2, which corresponds to

$$
z_{1-c2} \sqrt{2np(1-p)} < X_1 - X_2 < z_{1-d} \sqrt{2np(1-p)};
$$

and “stop the trial early for futility” when the stage I P value is less than c2, which corresponds to

$$
X_1 - X_2 < z_{1-c2} \sqrt{2np(1-p)}.
$$

Disclosure of Potential Conflicts of Interest

No potential conflicts of interests were disclosed.

Authors’ Contributions

Conception and design: M.-W. An, S.J. Mandrekar, D.J. Sargent

Development of methodology: M.-W. An, S.J. Mandrekar, D.J. Sargent

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): S.J. Mandrekar

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): M.-W. An, S.J. Mandrekar, D.J. Sargent

Writing, review, and/or revision of the manuscript: M.-W. An, S.J. Mandrekar, D.J. Sargent

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): M.-W. An, S.J. Mandrekar, D.J. Sargent

Study supervision: M.-W. An, S.J. Mandrekar

Grant Support

This work was supported by the National Cancer Institute grants CA-15083 (Mayo Clinic Cancer Center) and CA-25224 (North Central Cancer Treatment Group) to S.J. Mandrekar and D.J. Sargent.

Received March 2, 2012; revised May 7, 2012; accepted May 24, 2012; published OnlineFirst June 14, 2012.

References


Clinical Cancer Research

A 2-Stage Phase II Design with Direct Assignment Option in Stage II for Initial Marker Validation

Ming-Wen An, Sumithra J. Mandrekar and Daniel J. Sargent

Clin Cancer Res  Published OnlineFirst June 14, 2012.

Updated version  Access the most recent version of this article at:
doi:10.1158/1078-0432.CCR-12-0686

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.
Permissions  To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.