Published in final edited form as:

Stat Med. 2017 January 15; 36(1): 54–66. doi:10.1002/sim.7128.

Optimal Sequential Enrichment Designs for Phase II Clinical Trials

Yong Zang[1,*,†] and **Ying Yuan**[2,*,‡]

[1]Department of Biostatistics, School of Medicine, Indiana University, Indianapolis, Indiana 46202, U.S.A

[2]Department of Biostatistics, The University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030, U.S.A

Abstract

In the early phase development of molecularly targeted agents (MTAs), a commonly encountered situation is that the MTA is expected to be more effective for a certain biomarker subgroup, say marker-positive patients, but there is no adequate evidence to show that the MTA does not work for the other subgroup, i.e., marker-negative patients. After establishing that marker-positive patients benefit from the treatment, it is often of great clinical interest to determine whether the treatment benefit extends to marker-negative patients. The authors propose optimal sequential enrichment (OSE) designs to address this practical issue in the context of phase II clinical trials. The OSE designs evaluate the treatment effect first in marker-positive patients and then in marker-negative patients if needed. The designs are optimal in the sense that they minimize the expected sample size or the maximum sample size under the null hypothesis that the MTA is futile. An efficient, accurate optimization algorithm is proposed to find the optimal design parameters. One important advantage of the OSE design is that the go/no-go interim decision rules are specified prior to the trial conduct, which makes the design particularly easy to use in practice. A simulation study shows that the OSE designs perform well and are ethically more desirable than the commonly used marker-stratified design. The OSE design is applied to an endometrial carcinoma trial.

Keywords

subgroups; personalized medicine; molecularly targeted agents; optimal design; phase II trials

1. Introduction

Personalized or precision medicine is revolutionizing medical research. The premise of precision medicine is that patients are heterogeneous and respond differently to treatment agents known as molecularly targeted agents (MTAs). The development of MTAs relies on

*Correspondence to: Yong Zang, Department of Biostatistics, School of Medicine, Indiana University and Ying Yuan, Department of Biostatistics, The University of Texas M. D. Anderson Cancer Center.

†zangy@iu.edu

‡Equal Contribution

biomarkers that identify a sensitive subgroup (of patients) who will favorably respond to the treatment. It is assumed that one or a set of biomarkers has been selected to use in classifying patients into two subgroups, generically referred to as a marker-positive group and a marker-negative group, and that the marker-positive group presumably responds more favorably to the MTA (e.g., the marker-positive patients bear the genomic aberration that the MTA targets). If there are established evidences from the biological and clinical studies assuring that the biomarkers are truly predictive, the study population should be limited to the marker-positive patients to evaluate the treatment effect of the MTA using the enrichment trial design. However, at the early phase of the clinical development of MTAs, such as phase II clinical trials, a common issue is that there is a substantial amount of uncertainty about the ability of the biomarkers to identify the true sensitive subgroup. Therefore, after establishing that marker-positive patients benefit from the treatment, it is often of great clinical interest to determine whether the treatment benefit extends to marker-negative patients; otherwise, a large subset of the patient population will be deprived of beneficial treatment if the treatment actually is effective for the marker-negative patients. The objective of this article is to develop a optimal sequential enrichment (OSE) design to tackle the above practical issue in single-arm phase II clinical trials with binary outcomes.

Simon's optimal two-stage design is arguably the most commonly used phase II clinical trial design to test the treatment effect [1]. That design uses an interim look to enhance the individual ethics of the trial: if the number of responses at the interim look is less than a certain cutoff, the trial should be terminated early to avoid treating more patients with an ineffective agent. Simon's design focuses on the treatment effect for the overall population and thus is not suitable for testing the efficacy of MTAs in the presence of subgroups.

To handle MTAs, a straightforward extension of Simon's optimal design is to stratify the patients by their biomarker status, and then apply two parallel, independent Simon's two-stage designs to each of the strata (i.e., the marker-positive group and the marker-negative group). This design is known as the marker-stratified design (MSD) [2, 3], and has been frequently used to investigate a variety of MTAs [4, 5]. Although the MSD allows for the evaluation of the treatment effect for both marker-positive and marker-negative groups, it ignores the fact that in many applications, clinicians are interested in testing the treatment effect in marker-negative patients (i.e., whether the treatment benefit can extend to marker-negative patients) only after they have established that the MTA is effective for marker-positive patients. The rationale is that as the MTA targets genomic aberrations that are enriched in marker-positive patients, the MTA is expected to be more effective in those patients than in patients who do not show that particular genomic enrichment. If the MTA does not work for marker-positive patients, it is unlikely to benefit marker-negative patients and thus there is no need to test it in that group of patients from the viewpoint of both ethics and economics.

The OSE design is developed to handle the case in which the MTA is expected to be more effective in marker-positive patients, but there is no definite evidence to show that the treatment benefit of the MTA potentially extends to marker-negative patients, which is particularly common in early phase drug development. As illustrated in Figure 1, marker-positive and marker-negative groups are treated sequentially under the OSE design. The

proposed design starts from the marker-positive group, using one interim analysis and one final analysis for marker-positive patients to evaluate treatment efficacy. If the trial fails to achieve certain efficacy requirements at any of the analyses, the trial is terminated and no marker-negative patients will be treated. If the trial passes these two analyses, the trial proceeds to test the marker-negative patients, during which an interim check and a final analysis will be performed. In the OSE design, the interim decision rules are chosen to optimize a certain utility function (e.g., the expected sample size under the null hypothesis that the MTA is not effective), while controlling the subgroup-specific type I and II error rates. As shown later, similar to Simon's optimal design, one important advantage of the OSE design is that the go/no-go decision rules can be enumerated prior to the trial conduct, which makes the design particularly easy to use in practice.

The OSE design is motivated by a phase II clinical trial at MD Anderson Cancer Center for patients diagnosed with endometrial carcinoma. The MTA used in that trial is a phosphatidylinositol 3-kinase (PI3K) inhibitor. Patients are classified into marker-positive or marker-negative groups on the basis of their PI3K pathway activation scores [6]. Because the marker-positive patients have higher degrees of PI3K pathway activation, the MTA is expected to perform better in the marker-positive patients than in the marker-negative patients. The objective of this phase II trial is to investigate the efficacy of the MTA for marker-positive patients, and possibly for marker-negative patients if the MTA is shown to be effective for the marker-positive patients.

The rest of this article is organized as follows. The OSE design is proposed and a computationally efficient algorithm is developed to optimize the design parameters in Section 2. In Section 3, comprehensive simulation studies are carried out to investigate the operating characteristics of the proposed design and compare them to those of the MSD. In Section 4, the OSE design is applied to the motivating trial. A brief discussion and concluding remarks is provided in Section 5.

2. Optimal Sequential Enrichment Design

2.1. Design

Consider a phase II trial with a binary efficacy endpoint. Let p^+ denote the response rate for the marker-positive patients and p^- denote the response rate for the marker-negative patients with the assumption $p^+ \geq p^-$ to reflect that the MTA is expected to be more effective for the marker-positive patients. Let p_0 denote the highest unacceptable response rate such that the MTA is deemed futile, and let p_1 and p_2 denote the lowest acceptable response rates such that the MTA is promising for further development for marker-positive and marker-negative patients, respectively. In practice, p_1 is often equal to p_2. The interest of the OSE is to testing the following two pairs of subgroup-specific hypotheses:

$$H_0^+ : p^+ \leq p_0 \quad \text{versus} \quad H_1^+ : p^+ \geq p_1, \quad (1)$$

$$H_0^-: p^- \le p_0 \quad \text{versus} \quad H_1^-: p^- \ge p_2, \quad (2)$$

where the first pair of hypotheses tests whether the MTA is promising for marker-positive patients, and the second pair of hypotheses tests whether the MTA is promising for marker-negative patients. Under the assumption $p^+ \ge p^-$, if H_0^+ is failed to reject, H_0^- is failed to reject automatically; however, if H_0^+ is rejected, it may not be able to reject H_1^-.

The OSE design evaluates the treatment effect of the MTA first in marker-positive patients and then in marker-negative patients if needed. As illustrated in Figure 1 and enumerated below, the OSE design has 4 stages. The first two stages concern marker-positive patients and the last two stages concern marker-negative patients:

1. Enroll n_1^+ marker-positive patients. If the number of responses $X_1^+ > r_1^+$, move to the next stage. Otherwise, terminate the trial and conclude that the MTA is not effective for the overall population, including both marker-positive and marker-negative patients (i.e., fail to reject H_0^+ and H_0^-).
2. Enroll an additional n_2^+ marker-positive patients. Out of the total of $n^+ = n_1^+ + n_2^+$ enrolled marker-positive patients, if the number of responses $X^+ > r^+$, H_0^+ is rejected, claiming that the drug is effective for marker-positive patients, and initiate the next stage with marker-negative patients. Otherwise, the trial is terminated and the conclusion is that the MTA is not effective for the overall population.
3. Enroll n_1^- marker-negative patients. If the number of responses $X_1^- > r_1^-$, move to the next stage. Otherwise, terminate the trial and conclude that the MTA is not effective for marker-negative patients (i.e., fail to reject H_0^-), but effective for marker-positive patients (i.e., reject H_0^+).
4. Enroll an additional n_2^- marker-negative patients. Out of the total of $n^- = n_1^- + n_2^-$ enrolled marker-negative patients, if the number of response $X^- > r^-$, H_0^- is rejected and the drug is effective for the overall population (i.e., reject both H_0^- and H_0^+); otherwise, the conclusion is that the MTA is not effective for marker-negative patients, but is still effective for marker-positive patients.

2.2. Methods

The operating characteristics of the OSE design depend on four pairs of design parameters, (r_1^+, n_1^+), (r^+, n^+), (r_1^-, n_1^-), and (r^-, n^-). Let $\boldsymbol{r} = (r_1^+, r^+, r_1^-, r^-)$ and $\boldsymbol{n} = (n_1^+, n^+, n_1^-, n^-)$. The values of ($\boldsymbol{r}$,$\boldsymbol{n}$) are chose to minimize the expected total number of patients under the null hypothesis $p^+ = p^- = p_0$, while controlling the subgroup-specific type I and II error rates at certain prespecified levels. It is worthy noting that the null hypothesis $p^+ = p^- = p_0$ is only used for the purpose of minimizing the expected total number of patients, not for defining

type I error. The type I error that the design aims to control is the subgroup-specific type I errors, defined under subgroup-specific null hypotheses (1) and (2). Thus, the closed testing procedure [7, 8] is not applicable here.

Let α^+ and α^- denote the desirable subgroup-specific type I error rates for marker-positive and marker-negative patients, respectively; and β^+ and β^- denote the desirable subgroup-specific type II error rates for marker-positive and marker-negative patients. Because marker-positive patients and marker-negative patients are tested squentially, if we fail to reject H_0^+, we automatically fail to reject H_0^-. As a result, $\beta^- > \beta^+$ by design. This seems restrictive, but actually no different from the standard clinical trial paradigm, where the drug moves to phase III for testing only when the test in phase II is positive. In other words, if we view phase II and III trials as a single drug testing process, the type II error of phase III must be greater than that of phase II because if we fail to reject null in phase II, we automatically fail to reject the null in phase III. Let X_1^+ and X^+ be the numbers of responses at the first and second interim analyses among marker-positive patients, and let X_1^- and X^- be the number of responses at the third interim analysis and the final analysis among marker-negative patients. The design parameters ($\boldsymbol{r}$,$\boldsymbol{n}$) must satisfy the following type I and type II error constraints:

$$Pr(X_1^+ > r_1^+ \cap X^+ > r^+ | p^+ = p_0) \leq \alpha^+, \quad (3)$$

$$1 - Pr(X_1^+ > r_1^+ \cap X^+ > r^+ | p^+ = p_1) \leq \beta^+, \quad (4)$$

$$Pr(X_1^+ > r_1^+ \cap X^+ > r^+ \cap X_1^- > r_1^- \cap X^- > r^- | p^- = p_0) \leq \alpha^-, \quad (5)$$

$$1 - Pr(X_1^+ > r_1^+ \cap X^+ > r^+ \cap X_1^- > r_1^- \cap X^- > r^- | p^- = p_2) \leq \beta^-, \quad (6)$$

where the first two conditions define the type I and II error requirements for the test of the treatment effect for marker-positive patients, and the last two conditions define the type I and II error requirements for the test of the treatment effect for marker-negative patients. Noting that $X_1^+, X_1^-, (X^+ - X_1^+)$ and $X^- - X_1^-$ follow independent binomial distributions and defining $B(x; n, p) = \sum_{j=x+1}^{n} \binom{n}{j} p^j (1-p)^{n-j}$, it follows that the above constraints can be expressed as

$$
\begin{aligned}
B(r_1^+;n_1^+,p_0)B(r^+-r_1^+;n^+-n_1^+,p_0) &\le \alpha^+\\
1-B(r_1^+;n_1^+,p_1)B(r^+-r_1^+;n^+-n_1^+,p_1) &\le \beta^+\\
B(r_1^+;n_1^+,p_0)B(r^+-r_1^+;n^+-n_1^+,p_0)B(r_1^-;n_1^-,p_0)B(r^--r_1^-;n^--n_1^-,p_0) &\le \alpha^-\\
1-B(r_1^+;n_1^+,p_2)B(r^+-r_1^+;n^+-n_1^+,p_2)B(r_1^-;n_1^-,p_2)B(r^--r_1^-;n^--n_1^-,p_2) &\le \beta^-.
\end{aligned}
$$

Let $S(\boldsymbol{r},\boldsymbol{n} \mid p_0, p_1, p_2, \alpha^+, \alpha^-, \beta^+, \beta^-)$ denote the set of all possible values of $(\boldsymbol{r},\boldsymbol{n})$ that satisfy the type I and II error constraints. The goal of the OSE design is to find $(\boldsymbol{r}^*,\boldsymbol{n}^*) \in S(\boldsymbol{r},\boldsymbol{n} \mid p_0, p_1, p_2, \alpha^+, \alpha^-, \beta^+, \beta^-)$ that minimizes the expected total sample size $\mathrm{E}(n|p_0)$ under the null hypothesis $p^+ = p^- = p_0$ (i.e., the MTA is not effective for both marker-positive and marker-negative patients),

$$
(\boldsymbol{r}^*,\boldsymbol{n}^*)=\mathrm{argmin}_{\mathrm{E}(n|p_0)} S(\boldsymbol{r},\boldsymbol{n}|p_0,p_1,p_2,\alpha^+,\alpha^-,\beta^+,\beta^-). \quad (7)
$$

Letting $\mathrm{PET}_i(p)$ denote the probability of early termination after the ith stage when the response rate is p, $\mathrm{E}(n|p_0)$ is given by

$$
\mathrm{E}(n|p_0)=n_1^+ +\{1-\mathrm{PET}_1(p_0)\}n_2^+ +\{1-\mathrm{PET}_2(p_0)\}n_1^- +\{1-\mathrm{PET}_3(p_0)\}n_2^- \quad (8)
$$

where

$$
\begin{aligned}
\mathrm{PET}_1(p_0) &= 1-B(r_1^+;n_1^+,p_0)\\
\mathrm{PET}_2(p_0) &= 1-B(r_1^+;n_1^+,p_0)B(r^+-r_1^+;n^+-n_1^+,p_0)\\
\mathrm{PET}_3(p_0) &= 1-B(r_1^+;n_1^+,p_0)B(r^+-r_1^+;n^+-n_1^+,p_0)B(r_1^-;n_1^-,p_0).
\end{aligned}
$$

2.3. Optimization algorithm

In principle, an exhaustive numerical search can be conducted to optimize the objective function (7) and find the optimal design parameters $(\boldsymbol{r}^*,\boldsymbol{n}^*)$, which is the approach adopted by Simon's optimal two-stage design. This bruteforce approach, however, is not suitable in the OSE case because of the larger search space: simultaneous grid searching for 8 design parameters is extremely computationally expensive. To speed the optimization process, the authors propose a divide-and-conquer algorithm, which divides the optimization problem into two lower dimension optimization problems, and then they are optimized one by one. The divide-and-conquer algorithm is based on the following decomposition of the expected total sample size:

$$
\begin{aligned}
\mathrm{E}(n|p_0) &= n_1^+ +\{1-\mathrm{PET}_1(p_0)\}n_2^+ +\{1-\mathrm{PET}_2(p_0)\}n_1^- +\{1-\mathrm{PET}_3(p_0)\}n_2^-\\
&= n_1^+ +\{1-\mathrm{PET}_1(p_0)\}n_2^+ +\{1-\mathrm{PET}_2(p_0)\}\left\{n_1^- +\frac{1-\mathrm{PET}_3(p_0)}{1-\mathrm{PET}_2(p_0)}n_2^-\right\}\\
&= \mathrm{E}(n^+|p_0)+\{1-\mathrm{PET}_2(p_0)\}\mathrm{E}(n^-|p_0)
\end{aligned}
$$

where $E(n^+|p_0)$ and $E(n^-|p_0)$ are the expected sample sizes under $p^+ = p_0$ and $p^- = p_0$ for marker-positive and marker-negative patients, respectively. In other words, the overall expected sample size can be decomposed into the expected sample sizes for marker-positive and marker-negative patients from two separate trials.

Thus, the authors propose to minimize $E(n|p_0)$ by first minimizing $E(n^+|p_0)$, and then, conditional on the optimal solution of $E(n^+|p_0)$, minimizing $E(n^-|p_0)$. Strictly speaking, this two-step optimization procedure does not guarantee that its solution is globally optimal because $E(n^+|p_0)$ and $E(n^-|p_0)$ are correlated through $PET_2(p_0)$. However, in practice, the solution yielded by the above procedure almost always matches the global optimal solution. The key observation is that $\{1 - PET_2(p_0)\}$ is the type I error rate for the marker-positive patients (i.e., α^+), which is typically small, e.g., 0.1 or 0.05. As a result, $E(n|p_0)$ is predominantly determined by the value of $E(n^+|p_0)$, and thus $E(n|p_0)$ is approximately optimized when $E(n^+|p_0)$ is optimized. Among all possible solutions that minimize $E(n^+|p_0)$, the solution provided by the proposed algorithm is optimal in minimizing $E(n|p_0)$. This is verified by the numerical studies described later, which find that almost all the solutions from the proposed optimization algorithm match the global optimization solution. However, the proposed divide-and-conquer algorithm reduces the computation time by over 20-fold.

The optimization of $E(n^+|p_0)$ is straightforward because it is exactly the same as Simon's optimal two-stage design, noting that both the type I and II requirements (3) and (4) and objection function $E(n^+|p_0)$ do not depend on the marker-negative patients. Let $(r^+_{1,\text{opt}}, n^+_{1,\text{opt}})$ and $(r^+_{\text{opt}}, n^+_{\text{opt}})$ denote the solution that minimizes $E(n^+|p_0)$. Then, given $(r^+_{1,\text{opt}}, n^+_{1,\text{opt}})$ and $(r^+_{\text{opt}}, n^+_{\text{opt}})$, the optimization of $E(n^-|p_0)$ is determined by choosing appropriate values of (r^-_1, n^-_1) and (r^-, n^-) that satisfy the type I and type II requirements (5) and (6).

Optimizing $E(n^-|p_0)$ is less straightforward because the corresponding type I and II error requirements (5) and (6) depend on the design parameters of the marker-positive patients. That is, the part of the two-stage design that evaluates the treatment effect for marker-negative patients is not an independent, stand-alone, two-stage design, but rather depends on the part of the design that evaluates the treatment effect for marker-positive patients. The strategy is to reformulate the problem such that the part of the two-stage design that evaluates the treatment effect for marker-negative patients can be viewed as an independent, stand-alone, two-stage design. This is based on the results that follow (the proof of Lemma 1 is provided in the Appendix).

Lemma 1—Controlling the type I error rate and the type II error rate of the marker-negative group so that they are no greater than α^- and β^-, respectively, is equivalent to controlling

$$Pr(X_1^- > r_1^- \cap X^- > r^- | p^- = p_0) \leq \frac{\alpha^-}{Pr(X_1^+ > r_1^+ \cap X^+ > r^+)}, \quad (9)$$

$$1-\Pr(X_1^- > r_1^- \cap X^- > r^- | p^- = p_2) \leq \frac{\Pr(X_1^+ > r_1^+ \cap X^+ > r^+ | p^+ = p_2) + \beta^- - 1}{\Pr(X_1^+ > r_1^+ \cap X^+ > r^+ | p^+ = p_2)}. \quad (10)$$

Noting that the left sides of equations (9) and (10) are the definitions of type I and II error rates for an independent two-stage design for marker-negative patients, the above problem can be converted into a standard two-stage design optimization problem: minimize E($n^-|p_0$) for an independent two-stage design (for marker-negative patients) while controlling the type I and II error rates at the respective levels of

$$\alpha^* = \frac{\alpha^-}{Pr(X_1^+ > r_{1,\text{opt}}^+ \cap X^+ > r_{\text{opt}}^+)}$$
$$\beta^* = \frac{\Pr(X_1^+ > r_{1,\text{opt}}^+ \cap X^+ > r_{\text{opt}}^+ | p^+ = p_2) + \beta^- - 1}{\Pr(X_1^+ > r_{1,\text{opt}}^+ \cap X^+ > r_{\text{opt}}^+ | p^+ = p_2)},$$

where α^* and β^* can be viewed as the adjusted type I and type II error rates, accounting for the sequential testing procedure (i.e., marker-negative patients are tested after marker-positive patients). Given the values of ($r_{1,\text{opt}}^+, n_{1,\text{opt}}^+$) and ($r_{\text{opt}}^+, n_{\text{opt}}^+$), the value of β^* can be easily calculated, noting that X^+ and X_1^+ follow binomial distributions.

The difficulty is that α^* depends on the response rate of the marker-positive patients (i.e., p^+), which is typically unknown at the design stage. To circumvent that issue, the authors propose to replace p^+ with its upper bound u, and calculate the adjusted type I error rate as

$$\alpha^* = \frac{\alpha^-}{B(r_{1,\text{opt}}^+; X_1^+, u) B(r_{\text{opt}}^+ - r_{1,\text{opt}}^+; X^+ - X_1^+, u)}. \quad (11)$$

In practice, the value of u can be elicited from physicians. For example, for a certain treatment, physicians may expect that the response rate is unlikely to be higher than u = 60%. Typically, it is required that $u \geq p_1$. The validity of the above approach is given by Theorem 1.

Theorem 1—Given $p^+ \leq u$, if the design parameters (r_1^-, n_1^-) and (r^-, n^-) are chose based on the adjusted type I error rate α^* as given by (11), the type I error rate for the marker-negative patients is maintained under level α^-.

The proof is provided in the Appendix. One may be concerned about the overly specification of u; however, the sensitivity analysis (described later) shows that the OSE design is rather robust to the specification of u.

The proposed divide-and-conquer algorithm thus converts the original optimization problem into the optimization of two independent, 2-stage designs. The optimal values of (r_1^+, n_1^+) and (r^+, n^+) are obtained by minimizing E($n^+|p_0$) while controlling the type I and type II

error rates at α^+ and β^+. Then, given the optimal values of (r_1^+, n_1^+) and (r^+, n^+), the optimal values of (r_1^-, n_1^-) and (r^-, n^-) can be obtained by minimizing E($n^-|p_0$) while controlling the type I and type II error rates at α^* and β^*. For convenience, this design is referred to as OSE-O.

Thus far, the focus is on minimizing the expected total sample size E($n|p_0$). Alternatively, the minimax design can be developed with the aim of minimizing the maximal total sample size. That is, instead of minimizing the expected total sample size, the purpose of the minimax design is to find ($\boldsymbol{r}^\dagger$, $\boldsymbol{n}^\dagger$) $\in S(\boldsymbol{r}, \boldsymbol{n} \mid p_0, p_1, p_2, \alpha^+, \alpha^-, \beta^+, \beta^-)$ that minimizes the maximal total sample size Max($n|p_0$) under the null hypothesis $p^+ = p^- = p_0$,

$$(\boldsymbol{r}^\dagger, \boldsymbol{n}^\dagger) = \operatorname{argmin}_{\mathrm{Max(n|p_0)}} S(\boldsymbol{r}, \boldsymbol{n} | p_0, p_1, p_2, \alpha^+, \alpha^-, \beta^+, \beta^-). \quad (12)$$

The developed divide-and-conquer algorithm can be applied to find the optimal parameter ($\boldsymbol{r}^\dagger$, $\boldsymbol{n}^\dagger$) alone the same line. The only modification is to use the maximum sample size to replace the expected sample size during the optimization procedure. The resulting minimax design is denoted as the OSE-M.

3. Numerical studies

3.1. Operating characteristics of the OSE design

Throughout the numerical studies, the authors fix $p_1 = p_2$, as is typically the case in practice, i.e., the lowest acceptable response rate for the MTA is the same for marker-positive and marker-negative patients. Table 1 provides the operating characteristics of the OSE-O design under different values of (p_0, p_1), including optimal design parameters (r_1^+, n_1^+), (r^+, n^+), (r_1^-, n_1^-) and (r^-, n^-), the minimum expected sample size E($n|p_0$), and the probabilities of early termination PET_1, PET_2, and PET_3. The results are based on analytic calculations with $u = p_1 + 0.3$. Under each set of values of (p_0, p_1), the first, second and third rows correspond to different type I and type II requirements (α^+, β^+, α^-, β^-) = (0.05, 0.2, 0.05, 0.3), (0.05, 0.1, 0.05, 0.15) and (0.05, 0.1, 0.05, 0.2), respectively. We are aware that β^- should be greater than β^+ as the error is accumulating from the marker positive group to the marker negative group. In general, the OSE-O design terminates the trial early with high probabilities when the MTA is futile (i.e., $p^+ = p^- = p_0$), leading to small sample sizes. For example, assuming $p_0 = 0.1$, $p_1 = p_2 = 0.3$, and (α^+, β^+, α^-, β^-) = (0.05, 0.2, 0.05, 0.3), the design parameters for the OSE-O design are $(r_1^+, n_1^+) = (1, 10)$, (r^+, n^+) = (5, 29), $(r_1^-, n_1^-) = (1, 12)$ and (r^-, n^-) = (6, 35). Although the maximum sample size for this design is 64 (i.e., 29 marker-positive patients + 35 marker-negative patients), on average, only about 16 patients are needed to enroll when the treatment is futile for marker-positive and marker-negative patients. In this case, the probabilities of early termination up to the first, second and third interim analyses are 0.74, 0.95 and 0.98, respectively. To examine the accuracy of the proposed divide-and-conquer optimization algorithm, an exhaustive numerical search is also used to find the optimal design parameters (results not shown in Table 1 and Table 2). In all the scenarios considered in Table 1, the solutions obtained from the two optimization

methods all matched, but the computation time of the proposed algorithm was less than 1/20 of that of the exhaustive numerical search.

Table 2 summarizes the results for the OSE-M (i.e., minimax) design. Due to the different optimization criteria, compared to the OSE-O design, the OSE-M design has a smaller maximum sample size but a larger expected sample size under p_0. For example, comparing the first rows between Table 1 and Table 2, the maximum sample size under the OSE-M design is 7 patients fewer than that under the OSE-O design. However, on average, the OSE-M design still recruits about 4.3 more patients than the OSE-O design. Nevertheless, there are circumstances for which the OSE-M design may be preferable. For example, when p_0 = 0.2, p_1 = 0.4 and (α^+, β^+, α^-, β^-) = (0.05, 0.1, 0.05, 0.15), the OSE-M and OSE-O designs have almost the same E($n|p_0$) (33.47 versus 32.27), but the OSE-M design has a smaller maximum sample size than the OSE-O design (98 versus 120). More numerical studies with $p_1 - p_0 = 0.15$ can be found in the supplementary materials with Table S1 representing OSE-O and Table S2 representing OSE-M.

3.2. Comparison to the MSD

Comprehensive simulation studies are carried out to compare the proposed OSE-O and OSE-M designs to the MSD, which stratifies patients into marker-positive and marker-negative subgroups and then independently applies Simon's optimal two-stage design to each of the two subgroups. Two different versions of MSD are considered: the first one minimizes the overall sample size (denoted as MSD-O), and the second one is based on the minimax criterion (denoted as MSD-M). Let (α^+, β^+) = (0.05, 0.1), (α^-, β^-) = (0.05, 0.2), p_0 = 0.2, $p_2 = p_1$ = 0.4 and u = 0.7. Different configurations of the response rates p^+ and p^- are studied. Under each of the response rate configurations, 10,000 simulated trials were conducted for each design.

Table 3 summarizes the simulation results, including the power/type I error rate and average sample size. Scenarios 1 and 2 consider the cases in which the MTA does not work for both marker-positive and marker-negative patients. In these cases, the OSE designs show substantial advantage in individual patient ethics by exposing substantially fewer marker-negative patients to the ineffective treatment. For example, when $p^+ = p^- = 0.2$, the OSE designs enroll only 1.5 marker-negative patients, on average; whereas the MSD designs enroll more than 20 patients, on average. For the marker-positive patients, both the OSE designs and MSD designs yield type I error rates close to 5%. Note that for the marker-negative patients, the OSE designs yielded type I error rates substantially lower than 5%. This is because the upper bound u is used (to replace the true value of p^+) to calculate the adjusted type I error rate α^*. If the true value of p^+ is used, the type I error rate is about 5% (results not shown). This should not be regarded as a drawback because a low type I error rate is a desirable property if it does not notably affect the power. In scenarios 3 and 4, the MTA is effective for marker-positive patients, but not for marker-negative patients. The OSE designs and MSDs have similar performances in terms of the power to detect the treatment effect for the marker-positive patients and in controlling the type I error rate for the marker-negative patients. Scenarios 5 and 6 represent the situation in which the treatment is effective for both the marker-positive and marker-negative patients. Compared to the MSDs, the OSE

designs enroll approximately 5 more marker-negative patients on average when $p^+ = p^- = 0.4$ (i.e., scenario 5) and 9 more marker-negative patients on average when $p^+ = 0.5$ and $p^- = 0.4$ (i.e., scenario 6). When the treatment is effective for marker-negative patients, the OSE designs remain ethical because more patients benefit from the treatment within the trial. In addition, although the OSE designs recruit more patients, they are about 8% more powerful than the MSDs in detecting the treatment effect for marker-negative patients in scenario 6. In addition, more simulation results with different parameters' configuration can be found in the supplementary materials. Specifically, Table S3 focuses on different lowest acceptable response rate of $p_1 = p_2 = 0.35$ and Table S4 focuses on different type I and type II error requirements of $(\alpha^+, \beta^+) = (0.1, 0.15)$ and $(\alpha^-, \beta^-) = (0.1, 0.2)$.

In summary, the OSE designs outperform the MSDs with smaller expected sample sizes when no treatment effect is expected in both marker-positive and marker-negative subgroups. In the case that the treatment is effective for marker-positive patients, the OSE designs offer no clear advantage over the MSDs—the OSE designs may result in slightly larger sample size and require screening more patients with potentially longer trial time. Nevertheless, considering the fact that the success rate of phase II clinical trials is not high [9], the OSE designs are still attractive for some applications, for example, when the biomarker is cheap to measure and there is considerable uncertainty on efficacy of the experimental agents.

3.3. Sensitivity analysis

In the above simulations, $u = 0.7$ is used as the upper bound of p^+ (i.e., the response rate for marker-positive patients) to derive the optimal design parameters for the OSE designs. To assess the sensitivity to this upper bound, the operating characteristics of the OSE designs under different u are also examined, with the value increasing from 0.8 to 1.0, while using the same values for the other simulation configurations as those shown in Table 3. The results of the sensitivity analysis are shown in Table 4. It is easy to see that the results are rather stable across different choices of u and are very close to the results given in Table 3. Hence, if there is no empirical data to speculate a value of u a prior, a practical resolution is to set u directly at 1.

Table 4 only reports the results where u is correctly specified (greater than p^+). Table S5 in the supplementary materials investigates the performances of the OSE designs when u is mis-specified (less than p^+). According to the simulation results, the mis-specified u does not affect the power; however; it dose inflate the type I errors. The raise of the type I errors are marginal when u is close to p^+ ($u = 0.4$) and are substantial when u is far away from p^+ ($u = 0.2$).

Both the OSE designs and MSDs rely on a key prerequisite that an precise biomarker classifier exists at the beginning of the trial which can correctly classify every patient into either the marker-positive or marker-negative subgroups. However, in practice, such precise classifier may not be available considering the exploratory native of the phase II trial. Hence, it is important to study the performances of the OSE designs and MSDs in the presence of imperfect biomarker classifier. Table 5 summarizes the results of a sensitivity study with a non-informative biomarker classifier, which always classifies patients as marker-positive (or

marker-negative) with a probability of 50%, regardless the true biomarker status of the patients. According to the simulation results, if the treatment is promising or unpromising for both biomarker subgroups (Scenarios 1 and 2), the non-informative biomarker classifier has little impact on evaluating the treatment. However, if the treatment effect is limited to the marker-positive subgroup only (Scenarios 3 and 4), both OSE designs and MSDs fail to control type I and type II errors at their nominal levels. Hence, to implement the biomarker-based clinical designs, such as OSE designs and MSDs, it is important that the biomarker classifier is precise and validated. The authors of this manuscript have investigated this issue and published a series of papers that handle the biomarker classifier with misclassification errors [10, 11, 12, 13].

4. Application

The proposed OSE-O design is applied to the phase II trial to evaluate a PI3K inhibitor in endometrial carcinoma. The unacceptable response rate in the trial is $p_0 = 0.1$ and the lowest acceptable response rate is $p_1 = p_2 = 0.3$. The type I and type II error rates are controlled at $(\alpha^+, \beta^+) = (0.05, 0.2)$ for marker-positive patients and $(\alpha^-, \beta^-) = (0.05, 0.4)$ for marker-negative patients. Under the OSE-O design, the trial first enroll 10 marker-positive patients. If 1 or no patients respond to the PI3K inhibitor, the PIs terminate the trial and conclude that the treatment is unpromising; otherwise, 19 more marker-positive patients will be enrolled. Among the total 29 marker-positive patients, if 5 or fewer patients respond to the PI3K inhibitor, the PIs terminate the trial and conclude that the treatment is unpromising; otherwise, the PIs continue the trial to test marker-negative patients, initially enrolling 10 such patients. If 1 or no patients respond, the PIs terminate the trial and conclude that the PI3K inhibitor is effective only for marker-positive patients; otherwise, 12 additional marker-negative patients are enrolled. Among the total of 22 marker-negative patients, if more than 4 respond to the treatment, the PIs conclude that the PI3K inhibitor is effective for both marker-positive and marker-negative patients; otherwise the PI3K inhibitor is effective only for marker-positive patients. The maximal sample size under the OSE-O design is 51. Under the null hypothesis that the PI3K inhibitor is not effective for both marker-positive and marker-negative patients, the trial has 73.6%, 95.3% and 98.8% chance of being terminated early at the first, second and third interim checks. Therefore, according to equation (8), the expected sample size under the null hypotheses is $10 + 0.264 \times 19 + 0.047 \times 10 + 0.012 \times 12 = 15.6$.

For comparison, the authors also considered the MSD-O design, in which the following Simon's optimal two-stage design is used independently for the marker-positive and marker-negative patient treatment arms. The marker-positive arm for the MSD-O design is idential as that for the OSE-O design. For the marker-negative arm, at the first stage, 3 marker-negative patients are enrolled. If none of the patients responds, the trial is terminated, otherwise the PIs enroll 20 additional marker-negative patients. Out of the total of 23 marker-negative patients, if more than 4 patients respond, the PIs conclude that the treatment is promising; otherwise, the PIs conclude that the treatment is not effective for the marker-negative patients. The MSD-O design has a maximal sample size of 52, which is comparable with the OSE-O design. However, under the null hypothesis that the PI3K inhibitor is not effective for both marker-positive and marker-negative patients, the probability of early

termination for the marker-positive and marker-negative arms is 73.6% and 72.9% respectively. Therefore, the average sample size for the MSD-O design is $(10 \times 0.74 + 29 \times 0.26) + (3 \times 0.73 + 23 \times 0.27) = 23.4$. Therefore, by adopting the proposed OSE-O design, the trial could potentially treat 8 fewer (marker-negative) patients with a futile treatment.

5. Discussion

We have proposed the OSE designs to test the treatment effect of an MTA when it is expected that the treatment is more effective for marker-positive patients, but there is no adequate evidence to show that the MTA does not work for marker-negative patients. The OSE designs test the treatment effect for the marker-positive group first and then test it for the marker-negative patients if the MTA is shown to be effective for the marker-positive patients. Multiple interim analyses are incorporated into the design to terminate the trial early if the MTA is futile. The OSE designs minimize the expected sample size or the maximum value of the total sample size, while controlling the type I and II error rates. Simulation studies and an application to a endometrial carcinoma trial show that the OSE designs possess desirable operating characteristics and enhance the individual ethics of the trial. Because the go/no-go interim decision rules of the OSE designs can be enumerated prior to the trial conduct, it is particularly easy to use the design in practice.

Although the OSE designs yield desirable operating characteristics and are easy to implement, they have some limitations. Compared to the MSD that treats all comers, the OSE designs evaluate subgroups sequentially, and thus require to screen more subjects and often prolong the trial duration when the treatment is effective for marker-positive patients. Therefore, the OSE designs are more suitable for the case that there is substantial uncertainty on the efficacy of the experimental agent, or the case that patients can be easily screened and limited resource does not allow to treat many patients (all comers) simultaneously. Given the fact that most of experimental agents tested in phase II trials failed to move to phase III [9], the OSE designs provide an attractive design option for testing some MTAs. In addition, the OSE designs are built on the assumption that marker-positive patients are more likely to benefit from the treatment than marker-negative patients. In other words, the biomarker is predictive. The OSE designs are not a good choice when little information is available on the predictive ability of the biomarker. They are more appropriate for the case that there are strong biological rationale and evidence that the biomarker is predictive to a certain degree such that if the MTA does not work for marker-positive patients, it is less like to work for marker-negative patients.

Within each marker group, the two-stage design is adopted, which can be easily extended to three or more stages. However, as the decrease in the sample size from a two-stage design to one with three or more stages is limited [14], further extension may not be very useful. A more interesting extension is to consider multiple marker groups (larger than 2), with the ordinal relationship existing for only part of the groups. Further research in this direction is warranted. The proposed designs are appropriate for binary response outcomes. It is also of interest to extend the proposed designs to handle other response outcomes, such as time-to-event outcomes. The sequential designs take advantage of the ordinal relationship between two marker groups. If the order is unknown, adopting the sequential design may decrease the

power of the test for the marker-negative patients. If this is the case, a parallel test such as the stratified design may be more appropriate. Otherwise, a treatment effect comparison stage is needed at the beginning of the sequential design to identify the order [15]. However, controlling the additional type I and type II errors induced by the comparison stage will remain a challenge.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Yuan's research was partially supported by Award Number R01CA154591, 5P50CA098258 and P30CA016672 from the National Cancer Institute. The authors thank two referees and the associate editor for their valuable comments and LeeAnn Chastain for her editorial assistance.

References

1. Simon R. Optimal two-stage designs for phase II clinical trials. Controlled Clinical Trials. 1989; 10:1–10. [PubMed: 2702835]
2. Sargent DJ, Conley BA, Allegra C, Collette L. Clinical trial designs for predictive marker validation in cancer treatment trials. Journal of Clinical Oncology. 2005; 23:2020–2027. [PubMed: 15774793]
3. Mandrekar SJ, Sargent DJ. Clinical trial designs for predictive biomarker validation: theoretical considerations and practical challenges. Journal of Clinical Oncology. 2009; 27:4027–4034. [PubMed: 19597023]
4. Barker AD, et al. I-SPY 2: an adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. Clinical Pharmacology & Therapeutics. 2009; 86:97–100. [PubMed: 19440188]
5. Berry DA, Herbst RS, Rubin EH. Reports from the 2010 clinical and translational cancer research think tank meeting: design strategies for personalized therapy trials. Clinical Cancer Research. 2012; 18:638–644. [PubMed: 22298897]
6. Saal LH, et al. Poor prognosis in carcinoma is associated with a gene expression signature of aberrant PTEN tumor suppressor pathway activity. Proceedings of the National Academy of Sciences of the United States of America. 2007; 104:7564–7569. [PubMed: 17452630]
7. Marcus R, Peritz E, Gabriel KR. On closed testing procedures with special reference to ordered analysis of variance. Biometrika. 1976; 63:655–660.
8. Bretz F, Maurer W, Brannath W, Posch M. A graphical approach to sequentially rejective multiple test procedures. Statistics in Medicine. 2009; 28(4):586–604. [PubMed: 19051220]
9. Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J. Clinical development success rates for investigational drugs. Nature Biotechnology. 2014; 30:40–51.
10. Zang Y, Liu S, Yuan Y. Optimal marker-adaptive designs for targeted therapy based on imperfectly measured biomarkers. Journal of the Royal Statistical Society: Series C. 2015; 64:635–650.
11. Zang Y, Guo B. Optimal two-stage enrichment design correcting for biomarker misclassification. Statistical Methods in Medical Research. 2016 In Press.
12. Zang Y, Lee JJ, Yuan Y. Two stage marker-stratified clinical trial design in the presence of biomarker misclassification. Journal of the Royal Statistical Society: Series C. 2016; 65:585–601.
13. Zang Y, Guo B. Two-stage methods to implement and analyze the biomarker-guided clinical trial designs in the presence of biomarker misclassification. Precision Medicine. 2016; 2:1–5.
14. Chen TT. Optimal three-stage designs for phase II cancer clinical trials. Statistics in Medicine. 1997; 16:2701–2711. [PubMed: 9421870]
15. Thall PF, Simon R, Ellenberg SS. Two-stage selection and testing designs for comparative clinical trials. Biometrika. 1988; 75:303–310.

APPENDIX

Proof of Lemma 1

As (X_1^+, X^+) are independent of (X_1^-, X^-), the left part of the inequality (5) for the type I error is maximized at

$$Pr(X_1^+ > r_1^+ \cap X^+ > r^+) * Pr(X_1^- > r_1^- \cap X^- > r^- | p^- = p_0).$$

Thus, to control this probability so that it is no greater than α^-, the following condition is needed

$$Pr(X_1^- > r_1^- \cap X^- > r^- | p^- = p_0) \leq \frac{\alpha^-}{Pr(X_1^+ > r_1^+ \cap X^+ > r^+)}.$$

Let us turn to the type II error. As $p^+ \geq p^- \geq p_2$, the left part of inequality (6) is maximized at

$$1 - \Pr(X_1^+ > r_1^+ \cap X^+ > r^+ | p^+ = p_2) * \Pr(X_1^- > r_1^- \cap X^- > r^- | p^- = p_2).$$

Hence, to control the type II error so that it remains under β^-, it needs

$$1 - \Pr(X_1^- > r_1^- \cap X^- > r^- | p^- = p_+) \leq \frac{\Pr(X_1^+ > r_1^+ \cap X^+ > r^+ | p^+ = p_2) + \beta^- - 1}{\Pr(X_1^+ > r_1^+ \cap X^+ > r^+ | p^+ = p_2)}.$$

Proof of Theorem 1

$B(r, n, p)$ can be represented in terms of the regularized incomplete beta function as

$$B(r, n, p) = 1 - (n - r) \begin{pmatrix} n \\ r \end{pmatrix} \int_0^{1-p} t^{n-r-1} (1-t)^r dt.$$

Then, by taking the derivative of $B(r, n, p)$ with respect to p and get

$$\frac{\partial}{\partial p} B(r, n, p) = (n - r) \begin{pmatrix} n \\ r \end{pmatrix} (1-p)^{n-r-1} p^r > 0.$$

Hence, $B(r, n, p)$ is monotonically increasing with p, and

$$\frac{\alpha^-}{Pr(X_1^+ > r_1^+ \cap X^+ > r^+)} = \frac{\alpha^-}{B(r_1^+, n_1^+, p^+)B(r^+ - r_1^+, n^+ - n_1^+, p^+)} \geq \frac{\alpha^-}{B(r_1^+, n_1^+, u)B(r^+ - r_1^+, n^+ - n_1^+, u)} = \alpha^*$$

given $p^+ \leq u$. Therefore, according to Lemma 1, once the adjusted type I error rate is controlled at α^*, the type I error rate for the marker-negative patients is maintained under α^-.

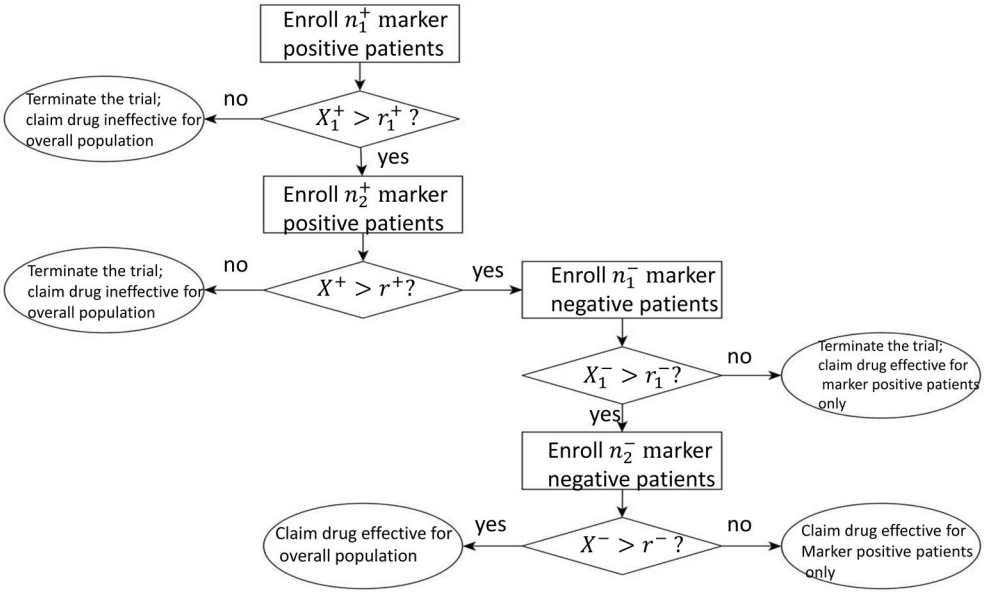

Figure 1.
Diagram of the OSE design.

Table 1

Operating characteristics of the OSE-O design for $p_1 - p_0 = 0.2$.

		Design parameters					Probability of early termination		
p_0	p_1	r_1^+/n_1^+	r^+/n^+	r_1^-/n_1^-	r^-/n^-	E(n|p_0)	PET$_1$(p_0)	PET$_2$(p_0)	PET$_3$(p_0)
0.1	0.3	1/10	5/29	1/12	6/35	15.95	0.74	0.95	0.98
		2/18	6/35	2/21	7/42	23.87	0.73	0.95	0.98
		2/18	6/35	1/13	6/35	23.54	0.73	0.95	0.98
0.2	0.4	3/13	12/43	4/18	14/50	21.92	0.75	0.95	0.99
		4/19	15/54	6/27	18/66	32.27	0.67	0.95	0.99
		4/19	15/54	5/22	14/49	31.84	0.67	0.95	0.99
0.3	0.5	5/15	18/46	6/19	21/54	25.16	0.72	0.95	0.98
		8/24	24/63	9/29	26/68	36.87	0.73	0.95	0.98
		8/24	24/63	6/19	23/60	36.35	0.73	0.95	0.98
0.4	0.6	7/16	23/46	8/19	29/59	26.09	0.72	0.95	0.99
		11/25	32/66	13/31	35/72	38.17	0.73	0.95	0.98
		11/25	32/66	11/25	29/59	37.64	0.73	0.95	0.99
0.5	0.7	8/15	26/43	10/19	34/57	25.06	0.70	0.95	0.98
		13/24	36/61	17/32	40/68	36.09	0.73	0.95	0.99
		13/24	36/61	10/19	36/61	35.60	0.73	0.95	0.98

The first, second and third rows of each value of (p_0, p_1) correspond to (α^+, β^+, α^-, β^-) = (0.05, 0.2, 0.05, 0.3), (0.05, 0.1, 0.05, 0.15) and (0.05, 0.1, 0.05, 0.2) respectively.

Table 2

Operating characteristics of the OSE-M design for $p_1 - p_0 - 0.2$.

		Design parameters					Probability of early termination		
p_0	p_1	r_1^+/n_1^+	r^+/n^+	r_1^-/n_1^-	r^-/n^-	E($n|p_0$)	PET$_1$(p_0)	PET$_2$(p_0)	PET$_3$(p_0)
0.1	0.3	1/15	5/25	1/16	6/32	20.29	0.55	0.97	0.99
		2/22	6/33	3/32	7/40	27.62	0.62	0.96	0.98
		2/22	6/33	1/15	6/33	27.12	0.62	0.96	0.98
0.2	0.4	4/18	10/33	4/21	12/41	23.60	0.72	0.95	0.98
		5/24	13/45	11/45	15/53	33.47	0.66	0.95	0.99
		5/24	13/45	4/22	13/44	32.77	0.66	0.95	0.98
0.3	0.5	6/19	16/39	8/28	19/47	27.37	0.67	0.95	0.98
		7/24	21/53	9/32	25/64	38.87	0.56	0.95	0.98
		7/24	21/53	6/22	20/50	38.31	0.56	0.95	0.98
0.4	0.6	17/34	20/39	21/43	25/50	36.57	0.91	0.95	0.99
		12/29	27/54	15/37	32/65	40.43	0.64	0.95	0.98
		12/29	27/54	17/37	26/52	40.01	0.64	0.95	0.99
0.5	0.7	12/23	23/37	24/41	28/46	29.75	0.66	0.95	0.99
		14/27	32/53	14/30	37/62	38.34	0.65	0.95	0.97
		14/27	32/53	29/48	30/50	38.33	0.65	0.95	0.99

The first, second and third rows of each value of (p_0, p_1) correspond to (α^+, β^+, α^-, β^-) = (0.05, 0.2, 0.05, 0.3), (0.05, 0.1, 0.05, 0.15) and (0.05, 0.1, 0.05, 0.2) respectively.

Table 3

Simulation results for the OSE and MSD deigns, with $p_0 = 0.2$, $p_1 = p_2 = 0.4$, $(\alpha^+, \beta^+) = (0.05, 0.1)$ and $(\alpha^-, \beta^-) = (0.05, 0.2)$.

				Power or type I error (%)		Sample size		
Scenario	p^+	p^-	**Design**	**Positive**	**Negative**	**Positive**	**Negative**	**Total**
1	0.2	0.2	OSE-O	5.0	0.2	30.4	1.5	31.9
			MSD-O	5.4	5.0	30.5	20.7	51.2
			OSE-M	4.5	0.2	31.2	1.4	32.6
			MSD-M	4.6	4.6	31.1	22.3	53.4
2	0.2	0.1	OSE-O	5.0	0	30.4	1.1	31.5
			MSD-O	5.1	0	30.6	14.0	44.6
			OSE-M	4.6	0	31.3	1.1	32.4
			MSD-M	4.9	0	31.4	18.4	49.8
3	0.4	0.1	OSE-O	90.8	0	51.6	20.5	72.1
			MSD-O	90.5	0	51.5	16.0	67.5
			OSE-M	89.7	0	44.1	20.9	65.0
			MSD-M	89.8	0	44.2	20.4	64.6
4	0.4	0.2	OSE-O	91.0	4.2	51.7	26.5	78.3
			MSD-O	91.1	5.2	51.7	22.5	74.2
			OSE-M	90.0	3.4	44.1	28.8	72.9
			MSD-M	89.9	4.9	44.1	24.3	68.4
5	0.4	0.4	OSE-O	90.3	80.8	51.5	42.5	94.0
			MSD-O	90.5	80.1	51.5	38.0	89.5
			OSE-M	89.8	80.0	44.2	39.1	83.2
			MSD-M	90.5	80.3	44.2	31.6	75.8
6	0.5	0.4	OSE-O	99.0	88.1	53.7	46.6	100.2
			MSD-O	99.2	80.1	53.7	38.0	91.7
			OSE-M	99.5	88.3	44.9	43.2	88.1
			MSD-M	99.4	80.5	44.9	31.6	76.5

The differences between the OSE and MSD designs in the marker-positive subgroup are due to the random variations from the Monte-Carlo simulation.

Stat Med. Author manuscript; available in PMC 2018 January 15.

Table 4

Sensitivity analysis of u with $p_0 = 0.2$, $p_1 = p_2 = 0.4$, $(\alpha^+, \beta^+) = (0.05, 0.1)$ and $(\alpha^-, \beta^-) = (0.05, 0.2)$.

				Power or type I error (%)		Sample size		
p^+	p^-	Design	u	Positive	Negative	Positive	Negative	Total
0.2	0.2	OSE-O	0.8	5.0	0.2	30.4	1.5	31.9
			1.0	4.9	0.2	30.3	1.6	31.9
		OSE-M	0.8	4.4	0.2	31.0	1.3	32.3
			1.0	4.5	0.3	31.1	1.5	32.6
0.2	0.1	OSE-O	0.8	4.9	0	30.3	1.2	31.5
			1.0	4.7	0	30.0	1.1	31.1
		OSE-M	0.8	4.8	0	31.4	1.0	32.4
			1.0	4.6	0	31.2	0.8	32.0
0.4	0.1	OSE-O	0.8	90.4	0	51.3	20.5	71.8
			1.0	90.2	0	51.2	20.4	71.6
		OSE-M	0.8	89.9	0	44.0	21.0	65.0
			1.0	89.7	0	44.3	20.9	65.2
0.4	0.2	OSE-O	0.8	91.0	4.0	51.7	26.4	78.1
			1.0	91.1	3.9	51.7	26.0	77.7
		OSE-M	0.8	90.0	3.3	44.2	28.5	72.7
			1.0	90.2	3.1	44.0	28.3	72.3
0.4	0.4	OSE-O	0.8	90.5	80.6	51.4	42.2	93.6
			1.0	90.4	80.3	51.4	41.9	93.3
		OSE-M	0.8	89.9	79.7	44.4	38.7	83.1
			1.0	90.1	79.3	44.1	38.5	82.6
0.5	0.4	OSE-O	0.8	99.1	87.8	53.8	46.2	100.0
			1.0	99.1	87.4	53.9	45.8	99.7
		OSE-M	0.8	99.5	88.0	44.9	42.8	87.7
			1.0	99.4	88.1	44.9	42.7	87.6

Table 5

Sensitivity analysis of non-informative biomarker classifier for the OSE and MSD designs, with $p_0 = 0.2$, $p_1 = p_2 = 0.4$, $(\alpha^+, \beta^+) = (0.05, 0.1)$ and $(\alpha^-, \beta^-) = (0.05, 0.2)$.

				Power or type I error (%)		Sample size		
Scenario	p^+	p^-	**Design**	**Positive**	**Negative**	**Positive**	**Negative**	**Total**
1	0.2	0.2	OSE-O	5.0	0.2	30.4	1.5	31.9
			MSD-O	5.4	5.0	30.5	20.7	51.2
			OSE-M	5.1	0.2	31.3	1.6	32.9
			MSD-M	5.3	4.8	31.3	22.3	53.6
2	0.6	0.4	OSE-O	99.0	98.1	53.7	48.3	102.0
			MSD-O	99.0	95.5	53.7	41.7	95.4
			OSE-M	99.4	98.8	44.9	43.7	88.6
			MSD-M	99.4	97.1	44.9	32.8	77.7
3	0.4	0.2	OSE-O	49.8	23.1	44.2	20.3	64.5
			MSD-O	50.0	41.2	44.2	30.5	74.7
			OSE-M	47.3	21.2	40.2	19.1	59.3
			MSD-M	47.2	38.1	40.2	28.1	68.3
4	0.5	0.2	OSE-O	75.7	55.0	48.8	33.8	82.6
			MSD-O	75.6	63.1	48.8	34.6	83.4
			OSE-M	73.9	52.5	42.9	31.3	74.2
			MSD-M	74.0	60.3	42.8	30.1	72.9