



Published in final edited form as:

Clin Trials. 2014 February ; 11(1): 49–59. doi:10.1177/1740774513500081.

SEARS: A Seamless Dose Escalation/Expansion with Adaptive Randomization Scheme

Haitao Pan^{1,2}, Fang Xie³, Ping Liu⁴, Jielai Xia¹, and Yuan Ji⁵

Yuan Ji: jiyuan@uchicago.edu

¹Department of Health Statistics, Fourth Military Medical University, Xi'an 710032, China

²School of Statistics, Xi'an University of Finance and Economics, Xi'an 710061, China

³Teva Pharmaceutical, USA

⁴Abbott Laboratory, USA

⁵Center for Clinical and Research Informatics, NorthShore University HealthSystem, Evanston IL 60201, USA. Prytzker School of Medicine, The University of Chicago, Chicago IL 60637, USA

Abstract

Background—Standard drug development conducts phase I dose finding and phase II dose expansion sequentially and separately. Information between the two phases is rarely shared. Administratively, such a sequential process is time consuming and burdensome.

Purpose—We propose SEARS, a seamless design that combines phase I dose escalation based on toxicity with phase II dose expansion and dose comparison based on efficacy. SEARS allows extension from phase I to phase II under one design with no gap in between, and employs a dynamic and parallel procedure involving simultaneous dose escalation, dose graduation, and adaptive randomization.

Methods—SEARS integrates three components into a seamless scheme. Specifically, in phase I, SEARS applies the mTPI method to monitor dose escalation based on toxicity outcome. Doses that show promising efficacy and safety are immediately graduated from phase I and placed to a phase II stage in which patients are adaptively randomized based on efficacy outcome. Phase I dose escalation, dose graduation, and phase II adaptive randomization proceed simultaneously throughout the entire trial.

Results—Examples are given comparing SEARS with two other designs, in which superior performance of SEARS is demonstrated. An important and promising finding is that SEARS reduces sample sizes without losing power. R program and demo slides of SEARS can be obtained at <http://www.northshore.org/research/investigators/yuan-ji-phd/>

Limitation—We assume that the binary efficacy and toxicity response can be measured in the same time frame. This is often achievable with surrogate efficacy markers in practice.

Keywords

Bayesian adaptive designs; Graduation rule; Phase I; Phase II; Seamless

1 Introduction

Phase I and II studies are early stages of drug development aiming to identify tolerable and efficacious doses of a regimen to be recommended for phase III confirmatory studies. Traditionally, phase I and phase II trials are conducted sequentially and separately, and trial data across different phases are rarely shared in statistical and medical decision making. For example, although patient efficacy response data might be recorded in phase I studies, they are rarely formally used to inform decisions in phase II trials based on the same drug and/or dose. Another drawback of the traditional strategy of conducting phase I and phase II studies separately is that patient populations could be time-dependent and the gap between the two phases might cause biased inference due to changes in patient characteristics. Lastly, investigators often need to prepare separate protocols and go through multiple review processes for administrative purposes, which increases the duration and cost for the entire drug development process. Thus, there is a need to seamlessly combine phase I and phase II trials.

Recently, there has been increasing research in the development of dose-finding methodologies based on both toxicity and efficacy outcomes, as opposed to toxicity outcome alone. Representative work includes [1], [2], [3], [4], [5], [6], among others. [7] and [8] proposed two-stage designs, in which, after completion of phase I studies, phase II studies subsequently borrow information from the previous phase to improve decision making. More recently Xie, Ji, and Tremmel (2012) [9] proposed to combine the toxicity dose finding in phase I with dose expansion in phase II. Specifically doses passing safety criteria in phase I are compared to placebo by randomizing patients between them. In addition, the initiation of phase II is not contingent on the completion of phase I. In other words, phase I dose-finding and phase II randomization are conducted in parallel. Disappointingly, they did not consider a formal dose-finding design for phase I.

Motivated by this work, we propose a seamless design, SEARS, aiming to combine phase I dose finding with phase II dose comparison in one trial. SEARS is characterized by three main features:

1. SEARS allows doses to transition from phase I evaluation to phase II with no gap in between. When a dose demonstrates initial safety and efficacy, it will graduate from phase I and directly join existing doses in phase II for head-to-head comparison under a randomization scheme.
2. Under SEARS, phase I and phase II are conducted in a parallel and dynamic fashion. Specifically, the two phases proceed simultaneously throughout the trial with promising doses leaving phase I and joining phase II whenever a graduation rule is satisfied.

3. Posterior inferences are used to bridge the two phases and allow for efficient information sharing between phases.

In summary, SEARS possesses two Bayesian adaptive methods and a graduation rule to realize these features. In phase I, the modified toxicity probability interval (mTPI) method (Ji et al., 2010) guides the toxicity monitoring and dose escalation. A graduation rule is continuously applied to send safe and promising doses to phase II for immediate evaluation. Owing to the flexibility of mTPI, when doses are taken out of phase I and graduate to phase II, dose finding in phase I proceeds without the need to modify the design or protocol. In phase II, an outcome-adaptive randomization scheme realizes direct comparison between graduated doses. To efficiently use all the information, response data from phase I are included in the adaptive randomization calculation. Since the randomization probabilities are based on coherent posterior inference, different levels of variabilities in the observed dose response data are automatically accounted for.

A schema of SEARS is presented in Figure 1. As a hypothetical case, the schema shows that doses 2 and 5 graduated from phase I to phase II, and dose 2 was eventually selected for further confirmatory studies. The dynamics of the seamless extension from phase I to phase II can be easily seen.

The rest of the paper is organized as follows. In Section 2, we describe the proposed SEARS design, including the specific methods for phase I and phase II, and the rule for graduating doses from phase I to II. In Section 3, we provide examples including a simulation study that compares SEARS with two other designs; we also provide a demo with a set of slides online for a hypothetical trial. We end with a brief discussion in Section 4.

2 SEARS Design

We present SEARS by sequentially introducing the design for phase I, the graduation rule, and the design for phase II.

2.1 Phase I Dose Finding

In phase I studies, the goal is to identify the maximum tolerated dose (MTD), a high dose with a tolerable toxicity rate less than a target probability, p_T (e.g., $p_T = 0.3$). Let $\mathbf{p} = (p_1, \dots, p_d)$ denote the toxicity probabilities for doses $i = 1, \dots, d$, where d is the total number of candidate doses in the trial. The observed data include n_i patients treated at dose i , and x_i the number of patients among n_i that experienced toxicity. The likelihood function for data $\{(x_i, n_i), i = 1, \dots, d\}$ is a product of binomial densities, $l(\mathbf{p}) \propto \prod_{i=1}^d p_i^{x_i} (1-p_i)^{n_i-x_i}$. Statistical inference is sequentially applied to estimate p_i and decide future doses that are close to the true MTD.

Anticipating that doses might graduate to phase II during the course of phase I dose finding, we apply the mTPI design (Ji et al., 2010) to monitor toxicity and conduct dose escalation. The mTPI design is an extension of the toxicity probability interval method (Ji et al., 2007) and employs a simple beta-binomial hierarchical model. Decision rules are based on calculating the unit probability mass (UPM) of three intervals corresponding to *under-*,

proper-, and *over*-dosing in terms of toxicity. Here, under-, proper-, or overdosing refers to whether a dose is lower, around, or higher than the MTD, respectively. The under-dosing interval is defined as $(0, p_T - \varepsilon_1)$, the over-dosing interval $(p_T + \varepsilon_2, 1)$, and the proper-dosing interval $(p_T - \varepsilon_1, p_T + \varepsilon_2)$, where ε_1 and ε_2 are small fractions, such as 0.05. The three intervals are associated with three different dose-escalation decisions. The under-dosing interval corresponds to a dose escalation, over-dosing corresponds to a dose de-escalation, and proper-dosing corresponds to staying at the current dose. Given an interval and a probability distribution, define the unit probability mass (UPM) of that interval as the probability of the interval divided by the length of the interval. The mTPI design calculates the UPM for the three dosing intervals, and the one with the largest UPM implies the corresponding dose-finding decision. Specifically, assume dose i is currently used to treat patients. Denote the three dose-finding decisions as escalation (E), de-escalation (D), and stay (S). To apply mTPI, we simply calculate the three UPMs for under-, proper-, and over-dosing intervals.

A dose-assignment rule \mathbf{B}_i based on these three UPMs chooses the decision with the largest UPM, that is,

$$\mathbf{B}_i = \arg \max_{m \in \{D, S, E\}} \text{UPM}(m, i). \quad (1)$$

Ji et al. (2010) showed that the decision \mathbf{B}_i is optimal in that it minimizes a posterior expected loss, in which the loss function is determined to achieve equal prior expected loss for the three decisions, D, S, and E.

The mTPI design assumes independent and vague priors $p_i \sim \text{beta}(1, 1)$, with $\text{beta}(a, b)$ denoting the beta density proportional to $x^{a-1}(1-x)^{b-1}$. It is often possible to borrow information between doses using a dependent prior based on a dose-response curve (e.g., a logistic regression). However, in many phase I studies, little information is known about the toxicity of the candidate doses and their dependence. In addition, phase I studies have small sample sizes. Therefore the dependence prior should not be informative so that posterior inference will not be severely biased toward the prior model. Combined with the likelihood $l(\mathbf{p})$, the posterior of p_i follows independent $\text{beta}(1 + x_i, 1 + n_i - x_i)$, for $i = 1, \dots, d$. When strong prior information on the toxicity of the candidate doses is available, informative beta priors can replace the vague priors. We recommend that the readers refer to [10, 11] for more details on mTPI.

2.2 Graduation from Phase I to Phase II

A crucial step in SEARS is to graduate doses from phase I to phase II without suspending either phase. This is realized by continuously applying a graduation rule that transitions doses with promising efficacy and low toxicity.

We first introduce the notation and a simple model for efficacy response. We assume that a binary efficacy response can be measured within the same time frame of the toxicity response. More discussion is given later on this assumption. Let Y_i denote the number of efficacy responses among n_i patients treated at dose i . Let q_i denote the true efficacy

probability of dose i . We assume that q_i are independent and follow Jeffreys prior $Beta(0.5, 0.5)$. Then the posterior distribution of q_i is $Beta(0.5 + Y_i, 0.5 + n_i - Y_i)$. Note that in mTPI, the toxicity probability p_i follows a $Beta(1, 1)$ prior, which is different from the prior for q_i . The use of prior $Beta(1, 1)$ for p_i is mainly due to the setup in the mTPI design. Here we choose Jeffrey's prior for q_i due to its invariant and noninformative property.

The proposed graduation rule is based on posterior probabilities of p_i and q_i . Mathematically dose i that satisfies

$$Pr(p_i < \bar{\pi}_T | data) > p^*, \quad \text{and} \quad Pr(q_i > \underline{\pi}_E | data) > q^* \quad (2)$$

is considered safe with promising efficacy, and will graduate from phase I and join phase II for further evaluation. Here p^* and q^* are two fixed cutoff probabilities, and $\bar{\pi}_T$ and $\underline{\pi}_E$ are physician-specified upper toxicity and lower efficacy probability thresholds, respectively. For example, $\bar{\pi}_T = p_T$, and $\underline{\pi}_E$ could be the historical response rate of the standard treatment. In words, the graduation rule posits that if a dose exhibits low toxicity and reasonable efficacy with high posterior probability, it will graduate to phase II.

An immediate impact after a dose graduates to phase II is that there will be one fewer dose in phase I dose escalation and one more dose in phase II. However, continuing phase I with one fewer dose is unproblematic. Consider an arbitrary example in which dose 3 has just graduated to phase II. Remaining in phase I are doses 1, 2, 4, ..., which can simply be relabeled as doses 1, 2, 3, ... And dose escalation continues based on mTPI using the decision rule in (1). Since *a posteriori*, the toxicity probabilities p_i 's are independent, posterior inference and decision based on \mathbf{B}_i remains the same. Therefore, phase I dose escalation proceeds as usual under mTPI with the new dose labels.

2.3 Phase II Adaptive Randomization

For phase II, we apply an adaptive randomization scheme similar to that in Huang et al. (2007). To take advantage of the seamless feature, we include the efficacy response data from phase I in computing the adaptive randomization probabilities.

Adaptive randomization (AR) procedures aim to assign larger numbers of patients to more efficacious dose arms. Bayesian AR procedures continuously update the randomization probability for arm i according to the observed response data. A common approach is to randomize patients to dose arm i with a probability proportional to $Pr(q_i > \underline{\pi}_E | data)$. However, Huang et al. (2007) pointed out that this approach may not perform well if all of the true response rates are much higher or lower than the threshold value $\underline{\pi}_E$. Instead, they proposed an AR probability proportional to

$$Pr(q_i > \max\{q_j, j \neq i\} | data), \quad (3)$$

which gives high probabilities to doses with *relatively* high efficacy rates. We will use the same AR probability to assign patients in the phase II stage of the SEARS design.

2.4 SEARS Design

We combine the aforementioned procedures in phase I, dose graduation, and phase II into a single seamless design. In addition, we introduce four practical rules as gate keepers in case of overly high toxicity or futility.

Safety rule 1 (early termination of trial)—Suppose that dose 1 has been used to treat patients. If $Pr(p_1 > p_T | data) > \xi_1$ for ξ_1 close to 1 (say, $\xi_1=0.95$), then terminate the trial due to excessive toxicity. That is, the trial is terminated if dose 1 is deemed to be too toxic.

Safety rule 2 (phase I toxicity dose exclusion)—Suppose that the decision is to escalate from dose i to $(i + 1)$. If $Pr(p_{i+1} > p_T | data) > \xi_2$, for ξ_2 close to 1 (say, $\xi_2 = 0.95$), then treat the next cohort of patients at dose i and exclude doses $(i + 1)$ and higher doses from the trial, i.e., these doses will not be used again in the trial. This rule excludes doses that are deemed too toxic before the end the trial. When no patients have been enrolled at dose $(i + 1)$, this rule should not be applied.

Safety rule 3 (phase II toxicity dose exclusion)—For any graduated dose i in phase II, if $Pr(p_i > p_T | data) > \xi_2$, for ξ_2 close to 1 (say, $\xi_2 = 0.95$), then exclude doses i and higher doses from the trial.

Futility rule (futility dose exclusion)—Let f^* be a small probability cutoff, e.g., $f^* = 0.2$. For dose i , if $Pr(q_i > q_0 | data) < f^*$, dose i will be excluded from the trial. This rule ensures that doses with low efficacy are excluded from the trial before it ends.

The SEARS design is summarized as follows.

Trial Initiation Patients of the first cohort are treated at the lowest dose level.

Onset of Phase I Phase I dose finding starts after the first cohort is enrolled. Dose escalation proceeds based on the mTPI design.

Graduation Monitoring Once phase I starts, the graduation rule (2) is continuously applied to all the doses in phase I. Any dose satisfying (2) will graduate to phase II immediately.

Onset of Phase II Once a dose graduates, phase II starts. Patients will be randomized to the graduated doses and a control arm. For arm i , the randomization probability is proportional to (3).

Practical rules Apply *Safety rules 1 & 2* to Phase I, and apply the *Safety rule 3 and Futility rule* to Phase II.

Phase I Termination Phase I is terminated if 1) there are no doses left in phase I or 2) a prespecified maximum sample size N_1 (e.g., $N_1 = 30$) for phase I has been reached.

Trial termination The trial is terminated when any of the three conditions is true: 1) *Safety rule 1* is invoked; 2) there are no doses left in either phase; or 3) a prespecified maximum sample size is reached.

At the end of trial, the final dose is selected according to the posterior probabilities of toxicity and efficacy. We propose to select a dose if

$$Pr(p_i < \bar{\pi}_T) > p^{**} \quad \text{and} \quad Pr(q_i > \underline{\pi}_E) > q^{**}. \quad (4)$$

This is similar to the graduation rule (2), and we recommend using higher values of p^{**} and q^{**} as more stringent thresholds.

To implement the SEARS design, cutoff probabilities $\xi_s, f^*, p^*, q^*, p^{**}$ and q^{**} will be calibrated. Typically the calibration is done by a clinical team, with sufficient discussion and communication between statisticians, study PIs, pharmacologists, and other research staff. The calibration for ξ_s, f^*, p^* and q^* is straightforward and can be carried out through consultation with physicians. Specifically, hypothetical trial data can be presented to physicians who will provide corresponding medical decisions. Then the cutoff values can be elicited to match the physicians' decisions, based on the hypothetical data. For example, $\xi_1 = 0.95$ would terminate the trial if three toxicity events are observed from three patients treated at dose 1, but not if two toxicity events are observed from three patients. In our experience, this would agree with physicians' judgments as well. Thus we set the default value of $\xi_1 = 0.95$. In addition, ξ_2 could be calibrated to allow for more or less safety control. A large ξ_2 value (say, 0.8) makes a dose easier to be excluded due to observed toxicity events. For untried doses, Safety rule 3 should be applied based on calculation using the prior distribution for p_i , and a reasonable prior should not exclude any untried doses. Calibration of p^{**} and q^{**} can be conducted by varying their values and examining the final dose selection percentages in the simulation. This is a standard practice in most adaptive designs.

3 Simulation Studies

3.1 Simulation Setup

We performed extensive simulation studies to assess the operating characteristics of the SEARS design. The toxicity rate of the MTD was set at $p_T = 0.17$. We compared the SEARS design with the design in Xie et al. (2012), which also combines phase I dose finding and phase II dose expansion into a single design. In addition, we compared SEARS to a conventional design that conducts two trials for dose finding and dose expansion. The three designs are denoted "SEARS", "XJT", and "Conventional", respectively.

The XJT design starts with a simple phase I scheme that allows dose escalation if no more than one out of six patients experience toxicity. For doses deemed safe, patients are randomized to the dose and its corresponding placebo with a 2:1 ratio. The number of patients to be randomized is determined adaptively under the XJT design, in which a small or large number of patients will be enrolled to the dose if it shows moderate or high efficacy, respectively. The Conventional design consists of a phase I dose-escalation using the 3+3 design and a subsequent phase II dose-finding study using a parallel-group design. The two phases are conducted sequentially and separately. The 3+3 design identifies the MTD in phase I. After completion of phase I, the estimated MTD and all the lower dose levels enter

phase II. In phase II, the parallel-group design is used to randomize patients equally among available dose arms, including the control, until the maximum sample size is reached. More detailed information about the XJT design and Conventional design can be found in Xie et al. (2012).

We assumed that five doses of a regimen were under investigation. According to the trial setup in Xie et al (2012), we set the maximum sample size to be 180 for all three designs. We construct simulation scenarios by specifying the true toxicity and efficacy rates for these doses. First, we considered three sets of toxicity response rates as:

equal toxicity with $p_i = 0.05$ for all $i = 1, \dots, 5$ doses;

increasing toxicity rates with all safe doses, i.e., $p_i = 0.03, 0.06, 0.09, 0.12, 0.15$ for dose $i = 1, \dots, 5$;

increasing toxicity rates with some unsafe doses, i.e., $p_i = 0.03, 0.06, 0.17, 0.3, 0.5$ for $i = 1, \dots, 5$. Here, doses 4 and 5 are not safe since their toxicity rates are higher than $p_T = 0.17$.

Second, we constructed six sets of efficacy rates for the five doses. The efficacy scenarios were labeled as *fnull*, *increasing*, *decreasing*, *n-shaped*, *u-shaped*, *plateaug* to describe the shapes of the response curves. They are given in Table 1. These scenarios cover a wide range of dose-response curves, including the n-shaped and u-shaped cases that are rare in most pharmacological settings. Lastly, we considered two different response rates for the control arm in phase II, with $q_0 = 0.2$ or 0.5 . Combining the three toxicity sets, six efficacy sets for the doses, and two efficacy rates for the control arm, we obtained a total of 36 scenarios. We determined the true probabilities of toxicity and efficacy in an *ad-hoc* fashion. Alternatively, one could consider simulating the true values using a dose-response model, such as a logistic regression. Response data X_i and Y_i were simulated based on independent binomial distributions using the true values of p_i and q_i as success rates, respectively. For each scenario, we simulated 1,000 trials based on the SEARS design. The cutoff probabilities were $p^* = 0.8$, $q^* = 0.6$ in (2), and $f^* = 0.2$ in the *Futility rule*, and p^{**} and q^{**} in (4) are calibrated according to the procedure below to ensure a fair comparison with the XJT and Conventional designs.

3.2 Results of Comparison

We compared the three designs, SEARS, XJT, and Conventional, in terms of sample size and trial power for all 36 scenarios. A dose is considered *desirable* if it has a true toxicity rate smaller than p_T and an efficacy rate larger than that of the control arm (q_0). The trial power is then defined as the probability of selecting at least one truly desirable dose in non-Null scenarios (scenarios 2–6 in Table 1). For fair comparisons, we first calibrate the SEARS and XJT designs so that they achieve the same trial type I error rate as the Conventional design, explained as follows.

At the end of the trial, the conventional design should recommend a desirable dose for further investigation. Following Xie et al. (2012), we regard the dose selection as a hypothesis testing problem. Consider a hypothesis $H_{0i} : q_i \leq q_0$ versus $H_{1i} : q_i > q_0$ for every dose $i = 1, \dots, 5$. Here q_0 represents an efficacy rate above which dose j would be

considered promising therapeutically. Given the trial data at each dose, one can conduct a hypothesis test at a specified level α . We define the *trial Type I error rate* as the probability of incorrectly rejecting H_{0i} in favor of H_{1i} for at least one dose i under the Null scenario setting in Table 1. The Null scenario represents the case where all $q_i = q_0$ and thus no doses are desirable. Selection of any dose is therefore considered a type I error. Due to multiple comparisons, the trial type I error rate will be larger than the level α of each individual test. If needed, one could enforce a family-wise type I error control on the trial performance by lowering the value of α so that the trial Type I error rate is no more than a desirable threshold, say 0.05. The trial power is the probability of correctly rejecting at least one H_{0i} in favor of H_{1i} under other non-Null scenarios for desirable doses i . Here a “correct rejection” means that 1) dose i has a true efficacy probability larger than q_0 and a true toxicity probability smaller than p_T , and 2) H_{0i} is rejected in favor of H_{1i} .

To compare the three designs, we first applied a chi-square test for each dose at the end of the trials under the Conventional design, with a one-sided $\alpha = 0.025$. Specifically, we simulated 1,000 trials under the Null scenario and applied the Conventional design to each simulated trial. Performing the chi-square test at the end of each trial, we calculated the trial type I error rate as the proportion of trials that falsely rejected any H_{0i} for dose i .

We then calibrated the two adaptive designs, XJT and SEARS, so that they achieved the same trial type I error rate. For SEARS, we kept $p^{**} = p^* = 0.8$ and calibrated q^{**} in (4) using a Bayesian test. Specifically, H_{0i} is rejected if posterior probability $Prob(q_1 > q_0 | data) > q^{**}$. Note that q^{**} was a function of q_0 . We found that setting $q^{**} = 0.97$ for $q_0 = 0.2$ and 0.925 for $q_0 = 0.5$ resulted in a matching type I error rate between SEARS and the Conventional design. Calibration of XJT required only tuning one parameter and we used the same result in Xie et al. (2012). After the calibration, all three designs had the same trial type I error rate under the Null scenarios. We then fixed the calibrated parameter values for all the designs, and simulated 1,000 trials under each of the remaining non-Null scenarios. This allowed us to fairly compare trial power and sample size of the three designs.

Figures 2 and 3 summarize results for 12 scenarios, with the true toxicity rates set at (0.03, 0.06, 0.17, 0.3, 0.5), control efficacy rate at either 0.2 or 0.5, and efficacy rates at each of the six efficacy scenarios. Superior performance is observed for SEARS. Similar results are found for the remaining scenarios, which are provided in the supplementary material. First notice that in Scenario 1 and Scenario 7, the two Null scenarios, all three designs have the same trial type I error rate due to the calibration procedure mentioned above. Examining the remaining non-Null scenarios, the mean total sample size using the SEARS design is considerably lower in all scenarios than under the XJT and Conventional designs. The Conventional design, in general, uses a little fewer than 180 patients for almost all the cases, where 180 is the specified maximum sample size. The XJT design has a smaller sample size than the Conventional design. However, relative to the XJT and the Conventional approaches, sample size reduction of SEARS nearly reaches 50% with no cost to the power (Scenario 11). Similar reduction in sample size is observed in other scenarios as well, without losing power. In summary, the SEARS design is more powerful than the XJT design in all cases, and exhibits comparable power to the Conventional design in all cases while requiring much fewer patients.

Smaller sample sizes and the seamless transition from phase I to phase II will also save time in drug development. Smaller sample sizes implies shorter enrollment periods. Seamless phase I-II eliminates the lengthy process of paper work between the two phases, therefore expediting the process and reducing administrative burden.

The improved performance of SEARS over the XJT and Conventional designs is encouraging, highlighting the importance of allowing seamless graduation of doses and information sharing across phase I and phase II. Below we provide more results demonstrating the performance of SEARS.

3.3 Operating Characteristics of SEARS

In Table 2 we present the percentage of selecting a dose and the average number of patients treated at each dose using SEARS. Selection of a dose implies that the dose can be recommended for phase III trials, and is based on the rule in (4). Since no dose, one dose, or more than one dose could be selected in a trial, the selection percentages across five doses do not need to sum up to one.

Scenarios 2 and 8 represent situations in which both toxicity and efficacy increase with dose level. In both scenarios, dose levels 4 and 5 are above the MTD. Desirable doses, doses 2 and 3 have the highest selection percentages and receive the largest numbers of patients. Scenarios 3 and 9 represent cases of decreasing efficacy response rates, in which SEARS behaves well; it allocates more patients to the lower level doses. For instance, in Scenario 9, about 33 and 20 patients on average are allocated to the first two dose levels, which accounts for 63% of the total sample size. In both scenarios, dose 1 is the most desirable and is selected with a frequency of over 90%. Scenarios 4 and 10 are unconventional cases in that the efficacy response rates form an n-shaped pattern, that is, initially increasing then decreasing. The SEARS design consistently chooses doses 2 and 3 as the most efficacious and safe doses. U-shaped response curves are present in Scenarios 5 and 11. Doses 4 and 5 are once again too toxic. In these cases a reasonable design should allocate most patients to the lowest dose level and allocate few patients to the mid-ranged doses. In scenario 5, we can see that SEARS allocates approximately 34 and 18 patients to doses 1 and 2, respectively, exhibiting a desirable behavior. Lastly, in Scenarios 6 and 12, dose efficacy responses plateaus at high dose levels. Doses 2 and 3 are desirable, and they are selected with the highest frequencies and receive the largest numbers of patients.

3.4 A Trial Monitoring Demo

We carried out a hypothetical trial on computer using the SEARS design. A slide file was produced to demonstrate the trial conduct based on SEARS. The slides went through an entire trial and reported interim decisions made throughout.

Briefly, patient response data were simulated according to Scenario 8 in Table 2 except that the true efficacy rate of dose 1 was set as 0.3, which was below the efficacy rate 0.5 of the control arm. SEARS started with a dose escalation in phase I, in which dose 4 was excluded from the study due to high toxicity, while doses 3 and 2 subsequently graduated to the phase II stage. The trial proceeded with more patients assigned to dose 1 for the Phase I stage and

simultaneously randomized patients in phase II to graduated doses. Eventually, dose 1 was excluded due to futility and the trial ended after the maximum sample size was reached.

The slide file is available online at <http://www.northshore.org/research/investigators/yuan-ji-phd/>.

4 Discussion

We have developed a seamless phase I/II design in which formal dose escalation, dose graduation, and adaptive randomization are coherently combined into a single framework. The SEARS design is truly gap-free in that doses can graduate directly from phase I to phase II and that the two phases are conducted simultaneously with information between them across for decision making.

The performance of the SEARS design is superior to the Convention and XJT designs. Compared with XJT, SEARS achieves higher power with smaller sample sizes due to its efficient adaptive methods and the seamless rules. Specifically, mTPI allows flexibility in graduating doses during the course of phase I dose escalation, and the adaptive randomization allocates more patients to more effective doses. Furthermore, the graduation rule and the stopping rules together provides a framework to promote the use of promising doses early while removing undesirable doses when necessary.

The calibration of SEARS is easy to implement compared with other adaptive designs proposed in the literature. The only parameters to be tuned are the probability cutoffs ξ 's p 's, and q 's in the graduation and stopping rules. Because they are probabilities, calibration of these parameters is intuitive and requires little effort. A prespecified maximum sample size is needed to implement SEARS. That maximum is usually bounded on logistic restriction and other factors. A simple choice is the maximum sample size under the conventional design, which was set to 180 in our simulation studies.

A limitation of the SEARS design is the assumption that efficacy and toxicity responses can be measured within the same time frame, e.g., after one cycle of therapy. This is usually achievable by using surrogate responses for efficacy in practice. For example, in cancer studies, tumor size shrinkage or biomarker abundance changes are often used as short-term responses to treatments. We will consider different types of responses and extend SEARS. Methodologies proposed in the literature, e.g., Bekele et al. (2007) and Ji and Bekele (2009) [12] [13], could be applied in the extension.

When both phase I and phase II are open to enrollment, a practical question is which phase to allocate a new patient to. We recommend that priority be given to phase I enrollment so that it could complete faster. For example, one could randomly assign patients to phase I and phase II with a 2:1 ratio. Another practical consideration is to account for potential time trends in the patient characteristics during the trial. Due to the sequential nature of the phase I trials, it is difficult to mitigate the potential bias due to time trends in the design. However, the bias can be reduced by setting up and executing unified inclusion criterion for eligible patients.

Finally, we note that there are many types of phase I and phase II trials in practice, such as health volunteer phase I trials in non-oncology settings. We focused on oncology phase I and phase II trials in this manuscript.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Dr. Yuan Ji's research is partially supported by NIH R01 132897. Dr. Jielai Xia is partially supported by Research Grants 81072390 and 81001290 from the National Science Foundation of China.

References

1. Braun TM. The bivariate continual reassessment method: extending the CRM to phase I trials of two competing outcomes. *ContrClinTrials*. 2002; 23:240–256.
2. Thall P, Cook J. Dose-finding based on Efficacy-Toxicity tradeoffs. *Biometrics*. 2004; 60:683–693.
3. Bekele B, Shen Y. A Bayesian approach to jointly modeling toxicity and biomarker expression in a phase I/II dose-finding trial. *Biometrics*. 2005; 61:343–354. [PubMed: 16011680]
4. Yin G, Li Y, Ji Y. Bayesian dose-finding in phase I/II dose-finding trials using toxicity and efficacy odds ratios. *Biometrics*. 2006; 62:777–787. [PubMed: 16984320]
5. Zhang W, Sargent D, Mandreka S. An adaptive dose-finding design incorporating both toxicity and efficacy. *Stat Med*. 2006; 25:2365–2383. [PubMed: 16220478]
6. Seegers V, Chevret S, Resche-Rigon M. Bayesian dose -finding design driven by efficacy in onco-hematology phase I/II trials. *Stat Med*. 2010; 30:1574–1583. [PubMed: 21394754]
7. Huang X, Biswas S, Oki Y, Issa JP, Berry D. A parallel phase I/II dose-finding trial design for combination therapies. *Biometrics*. 2007; 63:429–436. [PubMed: 17688495]
8. Berry S, Spinelli W, Liang J, Fardipour P, Berry D, Lewis R, et al. A Bayesian dose-finding trial with adaptive dose expansion to flexibly assess efficacy and safety of an investigational drug. *Clinical Trials*. 2010; 7:121–135. [PubMed: 20338905]
9. Xie F, Ji Y, Tremmel L. A Bayesian adaptive design for multi-dose, randomized, placebo-controlled phase I/II trials. *Contemporary Clinical Trials*. 2012; 34:739–748. [PubMed: 22426247]
10. Ji Y, Liu P, Li Y, Bekele B. A modified toxicity probability interval method for dose-finding trials. *Clinical Trials*. 2010; 7:653–663. [PubMed: 20935021]
11. Ji Y, S-J W. The mTPI Design: A Safer and More Reliable Method than the 3+3 Design for Practical Phase I Trials. *The Journal of Clinical Oncology*. 2013; 31:1785–1791.
12. Bekele B, Ji Y, Shen Y, Thall P. Monitoring late-onset toxicities in phase I trials using predicted risks. *Biostatistics*. 2008; 9:442–457. [PubMed: 18084008]
13. Ji Y, Bekele B. Adaptive randomization for multi-arm comparative clinical trials based on joint efficacy/toxicity outcomes. *Biometrics*. 2009; 65:876–884. [PubMed: 19173694]

SEARS Design Schema

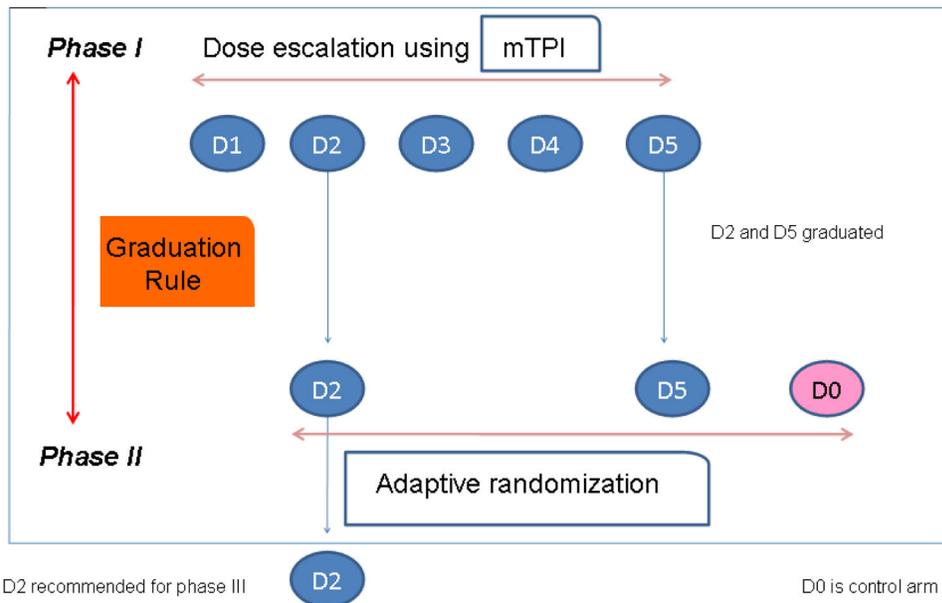


Figure 1. Schema of the SEARS design. Five doses are compared. In Phase I, dose escalation is based on the mTPI design and toxicity outcome. Doses can graduate from Phase I to Phase II under a graduation rule. Graduated doses are compared with each other as patients are randomized in Phase II.

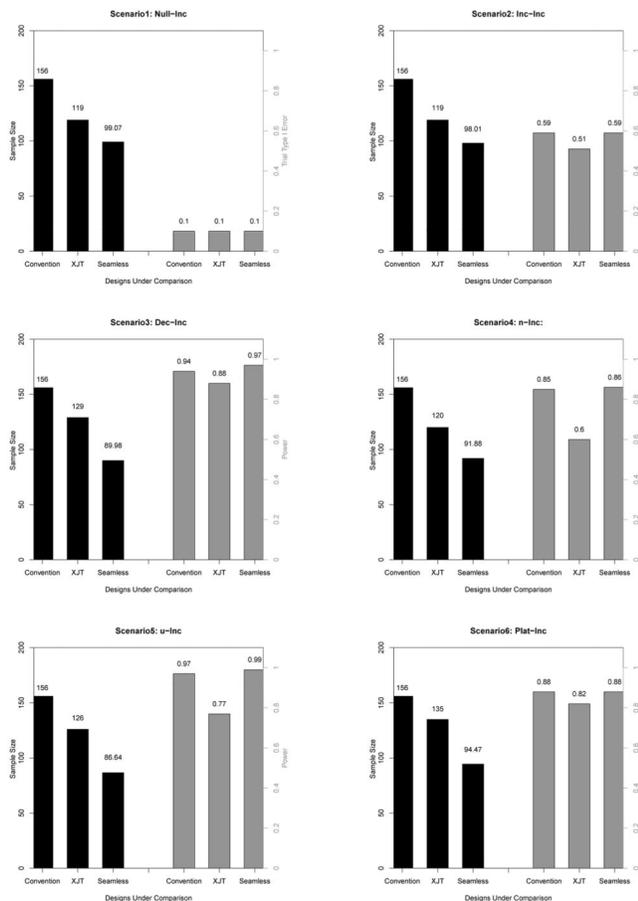


Figure 2. Comparison of the Conventional, XJT, and SEARS design, when the placebo efficacy rate $q_0 = 0.2$. Scenarios labels are in the format of EffShape-ToxShape representing the shapes of dose response curves for the efficacy and toxicity outcomes. Trial type I error rate is matched for the Null-Inc scenario for all three designs, and trial power and sample size are compared for the remaining scenarios.

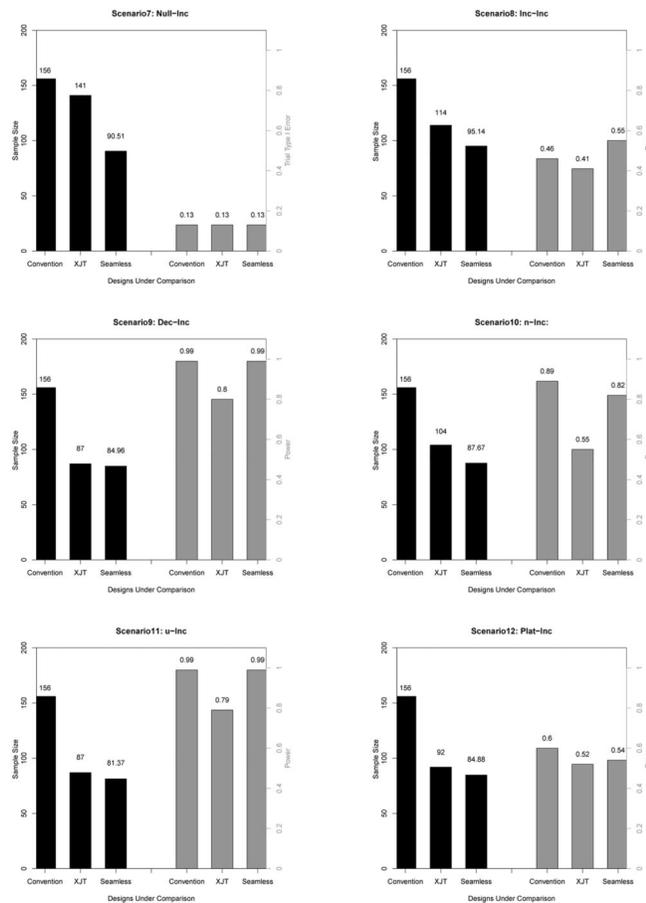


Figure 3. Comparison of the Conventional, XJT, and SEARS design, when the placebo efficacy rate $q_0 = 0.5$. Scenarios labels are in the format of EffShape-ToxShape representing the shapes of dose response curves for the efficacy and toxicity outcomes. Trial type I error rate is matched for the Null-Inc scenario for all three designs, and trial power and sample size are compared for the remaining scenarios.

Table 1

True dose response rates (q_i) to be used in the simulation. Six scenarios are constructed regarding the true efficacy rates for the five doses. They represent dose response shapes: Null, Increasing, Decreasing, n-shaped, u-shaped, and Plateau. Scenarios 2–6 are the non-Null scenarios.

Scenario Label	Shape	Control arm efficacy $q_0 = 0.2$	Control arm efficacy $q_0 = 0.5$
Null	Null	0.2, 0.2, 0.2, 0.2, 0.2	0.5, 0.5, 0.5, 0.5, 0.5
Inc	Increasing	0.2, 0.3, 0.5, 0.7, 0.8	0.5, 0.6, 0.7, 0.8, 0.9
Dec	Decreasing	0.8, 0.7, 0.5, 0.3, 0.2	0.9, 0.8, 0.7, 0.6, 0.5
n	n-shaped	0.2, 0.4, 0.8, 0.4, 0.2	0.5, 0.7, 0.9, 0.7, 0.5
u	u-shaped	0.8, 0.4, 0.2, 0.4, 0.8	0.9, 0.7, 0.5, 0.7, 0.9
Plat	Plateau	0.2, 0.4, 0.5, 0.5, 0.5	0.3, 0.6, 0.7, 0.7, 0.7

Table 2

Operating characteristics of SEARS over 1,000 simulated trials. Scenarios labels are in the format of EffShape-ToxShape representing the shapes of dose response curves for the efficacy and toxicity outcomes. The true efficacy response rates are given in the first row of each scenario block. Doses marked with * are toxic with probability of toxicity higher than that of the MTD. Tox% refers to the percentage of patients experienced DLTs in all 1,000 simulated trials.

Scenario (eff-tox)	Control Arm Efficacy $q_0 = 0.2$							Tox%
	Dose	1	2	3	4*	5*	Control	
	True Tox. Prob.	0.03	0.06	0.17	0.30	0.50	0.50	# Pts
1. Null-Inc	True Eff. Prob.	0.2	0.2	0.2	0.2	0.2	0.2	
	Sel %	3.4	4.7	1.9	0.4	0		
	# Pts	24.19	24.39	19.15	8.73	2.57	20.05	12
2. Inc-Inc	True Eff. Prob.	0.2	0.3	0.4	0.5	0.6		
	Sel %	3.1	34.2	34.9	4.2	0.1		
	# Pts	20.70	26.68	21.32	9.52	3.31	16.29	13
3. Dec-Inc	True Eff. Prob.	0.6	0.5	0.4	0.3	0.2		
	Sel %	94.6	75.1	33.5	4.5	0		
	# Pts	31.51	21.79	14.23	8.81	4.17	8.98	12
4. n-Inc	True Eff. Prob.	0.2	0.4	0.6	0.4	0.2		
	Sel %	4.8	63.1	48.4	5.1	0		
	# Pts	16.90	24.98	23.28	8.53	3.79	12.92	13
5. u-Inc	True Eff. Prob.	0.6	0.4	0.2	0.4	0.6		
	Sel %	96.9	52.7	3.9	6.7	0.4		
	# Pts	34.03	17.95	12.78	8.61	3.61	9.73	11
6. Plat-Inc	True Eff. Prob.	0.2	0.4	0.5	0.5	0.5		
	Sel %	4.1	68.4	47.8	5.6	0.1		
	# Pts	18.15	27.83	21.99	8.94	3.41	13.83	13

Scenario (eff-tox)	Control Arm Efficacy $q_0 = 0.5$					Control # Pts	Tox%
	Dose	1	2	3	4*		
	True Tox. Prob.	0.03	0.06	0.17	0.30	0.50	
7. Null-Inc	True Eff. Prob.	0.5	0.5	0.5	0.5	0.5	
	Sel %	4.5	5.0	1.9	0.1	0	
	# Pts	22.65	22.14	18.14	8.19	2.23	18.17
8. Inc-Inc	True Eff. Prob.	0.5	0.6	0.7	0.8	0.9	
	Sel %	3.8	25.2	32.3	4.0	0.0	
	# Pts	20.65	25.68	21.22	9.66	3.29	15.20
9. Dec-Inc	True Eff. Prob.	0.9	0.8	0.7	0.6	0.5	
	Sel %	96.9	67.1	24.2	2.6	0	
	# Pts	33.33	19.66	13.62	8.35	4.37	7.05
10. n-Inc	True Eff. Prob.	0.5	0.7	0.9	0.7	0.5	
	Sel %	4.6	56.0	51.1	3.0	0	
	# Pts	16.06	23.86	23.83	8.24	3.63	11.19
11. u-Inc	True Eff. Prob.	0.9	0.7	0.5	0.7	0.9	
	Sel %	98.4	41.0	1.2	5.3	0.1	
	# Pts	34.79	15.71	11.72	7.80	3.63	7.57
12. Plat-Inc	True Eff. Prob.	0.3	0.6	0.7	0.7	0.7	
	Sel %	0	26.1	34.9	2.3	0	
	# Pts	11.74	25.72	21.07	8.33	2.44	14.51