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

Haitao Pan\textsuperscript{1,2}, Fang Xie\textsuperscript{4}, Ping Liu\textsuperscript{5}, Jielai Xia\textsuperscript{1,*}, Yuan Ji\textsuperscript{3,*}

March 7, 2012

\textsuperscript{1}Department of Health Statistics, Fourth Military Medical University, Xi’an 710032, China.
\textsuperscript{2}School of Statistics, Xi’an University of Finance and Economics, Xi’an 710061, China.
\textsuperscript{3}Center for Clinical and Research Informatics, NorthShore University HealthSystem, Evanston IL 60201, USA
\textsuperscript{4}Teva Pharmaceutical, USA.
\textsuperscript{5}Abbott Laboratory, USA.

*Corresponding to: Yuan Ji and Jielai Xia

Email: ylji@northshore.org
Background Standard drug development conducts phase I dose finding and phase II dose expansion sequentially and separately. Information between the two phases are rarely shared and 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 (Ji et al., 2010) 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 proceeds 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 a movie demo of SEARS can be obtained at http://odin.mdacc.tmc.edu/~yuanj/soft.html

Limitation We assume that the binary efficacy and toxicity responses can be measurement 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 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 in between the two phases might cause biased inference due to the changes in patient characteristics. Lastly, administratively investigators often need to prepare separate protocols and go through multiple review processes, which increases the duration and cost for the entire drug development process. 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 works include \cite{Braun2002}, \cite{ThallCook2004}, \cite{BekeleShen2005}, \cite{Yin2006}, \cite{Zhang2006}, \cite{Seegers2010}, among others. \cite{Huang2007} and \cite{Berry2010} proposed two-stage designs, in which, after completion of phase I studies, phase II studies subsequently borrow information from previous phase to improve decision making. More recently \cite{Xie2012} proposed to combine the toxicity dose finding in phase I with dose expansion in phase II. Specifically they allow doses passing safety criteria in phase I to be compared to placebo by randomizing patients between them. In addition, they did not insist that phase II can only start after phase I is completed. 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 the phase I part.

Motivated by these works, we propose a seamless design, SEARS, aiming to combine phase I dose finding with phase II dose comparison in one trial. SEARS is characterised 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 is considered safe and with promising 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 sharing information across phases.

In a nutshell, SEARS possesses two Bayesian adaptive methods and a graduation rule to realize these features. In phase I, the 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. Thanks to the flexibility of mTPI, when doses are taken out of the 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. It can be easily seen the dynamics of the seamless extension from phase I to phase II.

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 an on-line movie 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, 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 \( p = (p_1, \ldots, p_d) \) denote the toxicity probabilities for doses \( i = 1, \ldots, d \), where \( d \) is the total number of candidate doses in the trial. The observed data include \( n_i \) patients treated at dose \( i \) and the corresponding \( x_i \) experiencing toxicity. The likelihood function for data \( \{(x_i, n_i), i = 1, \ldots, d\} \) is a product of binomial densities,

\[
l(p) \propto \prod_{i=1}^{d} p_i^{x_i}(1 - p_i)^{n_i-x_i}. \tag{1}
\]

Statistical inference is sequentially applied to estimate \( p_i \)'s 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 modified toxicity probability interval (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 unit probability mass (UPM) of three intervals corresponding to under-, proper-, and over-dosing in terms of toxicity. Here, under-, proper-, or over-dosing refers to whether a dose is lower, around, or higher than the
MTD, respectively. Specifically, the under-dosing interval is defined as $(0, p_T - \epsilon_1)$, the overdosing interval $(p_T + \epsilon_2, 1)$, and the proper-dosing interval $(p_T - \epsilon_1, p_T + \epsilon_2)$, where $\epsilon_1$ and $\epsilon_2$ 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, overdosing 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, given by

$$\text{UPM}(D, i) = \frac{P(p_i - p_T > \epsilon_2 \mid \text{data})}{1 - p_T - \epsilon_2} \text{ for under dosing},$$

$$\text{UPM}(S, i) = \frac{P(-\epsilon_1 \leq p_i - p_T \leq \epsilon_2 \mid \text{data})}{\epsilon_2 + \epsilon_1} \text{ for proper dosing},$$

$$\text{UPM}(E, i) = \frac{P(p_i - p_T < -\epsilon_1 \mid \text{data})}{p_T - \epsilon_1} \text{ for over dosing}.$$

A dose-assignment rule $B_i$ based on these three UPMs chooses the decision with the largest UPM, that is,

$$B_i = \arg \max_{m \in \{D, S, E\}} \text{UPM}(m, i).$$

(2)

Ji et al. (2010) showed that the decision $B_i$ 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 the $\text{Beta}$ density proportional to $x^{a-1}(1-x)^{b-1}$. Combined with the likelihood in (1), the posterior of $p_i$ follows independent $\text{Beta}(1 + x_i, 1 + n_i - x_i)$, for $i = 1, \ldots, d$. When strong prior information on the toxicity of the candidate doses are available, informative beta priors can
replace the vague priors. We recommend the readers referring to Ji et al. (2010) for details of 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 transition doses with promising efficacy and low toxicity.

We first introduce 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

$$Pr(p_i < \bar{\pi}_T \mid data) > p^*,$$

and

$$Pr(q_i > \bar{\pi}_E \mid data) > q^*$$

(3)

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 $\bar{\pi}_E$ are physician-specified upper toxicity and lower efficacy probability thresholds, respectively. For example, $\bar{\pi}_T = p_T$, and $\bar{\pi}_E$ could be the historical response rate of the standard treatment. In words, the graduation rule says 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. So a natural question is how
to proceed phase I with one fewer dose. The answer is to do almost nothing! Consider an arbitrary example in which dose 3 has just graduated to phase II. Then left in phase I are doses 1, 2, 4, ... left in phase I. Simply relabel them as doses 1, 2, 3, ... and continue dose escalation based on mTPI using the decision rule in (2). Since \( a \text{ posteriori} \), the toxicity probabilities \( p_i \)'s are independent, posterior inference and decision based on \( B_i \) 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 the 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 at assigning 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 \( Pr(q_i > \pi_E \mid data) \). 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 \( \pi_E \). Instead, they propose an AR probability proportional to

\[
Pr(q_i > \max\{q_j, j \neq i\} \mid data),
\]

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 three practical rules as gate keepers in
case of over toxicity or futility.

Safety rule 1 (early termination of trial). Suppose that dose 1 has been used to treat patients. If $\text{Pr}(p_1 > p_T \mid \text{data}) > \xi_1$ for $\xi_1$ 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 (toxicity dose exclusion). Suppose that the decision is E according to mTPI, to escalate from dose $i$ to $(i + 1)$. If $\text{Pr}(p_{i+1} > p_T \mid \text{data}) > \xi_2$, for $\xi_2$ 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 from the trial, i.e., these doses will not be used again in the trial. In words, this rule excludes doses that are deemed too toxic before the end the trial.

Futility rule (futility dose exclusion). For dose $i$, if $\text{Pr}(q_i > q_0 \mid \text{data}) < f^*$, 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 (3) is continuously applied to all the doses in phase I. Any dose satisfying (3) 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 (4).

**Practical rules** Apply Safety rules 1 & 2 to Phase I, and apply the Futility rule to Phase II.
Phase I Termination Phase I is terminated if 1) there is no dose left in phase I or 2) a prespecified maximum sample size $N_1$ (e.g., $N_1 = 30$) for phase I has reached.

Trial termination The trial is terminated when either of the three conditions is true: 1) Safety rule 1 is invoked; 2) no dose is left in both phases; or 3) a prespecified maximum sample size is reached.

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

$$Pr(p_i < \bar{\pi}_T) > p^{**} \quad \text{and} \quad Pr(q_i > \bar{\pi}_E) > q^{**}.$$  \hspace{1cm} (5)

This is similar to the graduation rule (3), 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. The calibration for $\xi$’s, $f^*$, $p^*$ and $q^*$ is straightforward and can be easily carried out by consulting with physicians. Specifically, hypothetical trial data can be presented to physicians who will provide corresponding medical decisions. Then the cutoff values can be easily elicited to match the decisions provided by the physicians based on the hypothetical data. For example, $\xi_1 = 0.95$ would terminate the trial if 3 toxicity events are observed from 3 patients treated at dose 1, but not if 2 toxicity events are observed from 3 patients. In our experience, this would agree with physicians’ judgement as well. Thus we set the default value of $\xi_1 = 0.95$. Calibration of $p^{**}$ and $q^{**}$ can be easily conducted by varying their values and examining the final dose selection percentages in the simulation. It is a standard practice in most adaptive designs.
3 Numerical Examples

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 to a conventional design that would conduct 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 6 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 the 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. We set up 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 $i = 1, \cdots, 5$ doses;

**increasing toxicity rates with all doses safe**, i.e., $p_i = 0.03, 0.06, 0.09, 0.12, 0.15$ for
dose $i = 1, \ldots, 5$; increasing toxicity rates with some unsafe doses, i.e., $p_i = 0.03, 0.06, 0.17, 0.3, 0.5$ for $i = 1, \ldots, 5$. Here, doses 4 and 6 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 \{null, increasing, decreasing, n-shaped, u-shaped, equal\} to describe the shape of the response curve. They are given in Table 1. 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. For each scenario, we simulated 2,000 trials based on the SEARS design. The cutoff probabilities were $p^* = 0.8, q^* = 0.6$ in (3), and $f^* = 0.2$ in the Futility rule, and $p^{**}$ and $q^{**}$ in (5) are calibrated according to the procedure below to ensure a fair comparison with the XJT and Conventional designs.

### 3.2 Simulation Results Comparing Three Designs

We compared the three designs, SEARS, XJT, and Conventional, in terms of sample size and trial power for all 36 scenarios. We call a dose desirable if it has a true toxicity rate smaller than $p_T$ and efficacy rate larger than that of the control arm ($p_0$). The trial power is then defined as the probability of selecting truly desirable doses 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.

Consider a simple hypothesis $H_{0j} : q_j \leq q_0$ versus $H_{1j} : q_j > q_0$ for every dose $j = 1, \ldots, 5$. 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 testing at a specified level $\alpha$. We define the trial Type I error rate as the probability of rejecting $H_{0j}$ in favor of $H_{1j}$ incorrectly for at least one dose $j$ under the Null scenario setting in Table 1. The
Null scenario represents the case where all $q_j = q_0$ and thus no doses are desirable. Selection of any dose is therefore considered as a type I error. Due to multiple comparison, the trial type I error rate will be larger than the level $\alpha$ of each individual test. The trial power is the probability of correctly rejecting $H_{0j}$ in favor of $H_{1j}$ under other non-Null scenarios for desirable doses $j$. 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 $\alpha = 0.025$. Specifically, we simulated 2,000 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 $H_{0j}$ for any dose $j$. 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 (5) using a Bayesian test. Specifically, $H_{0j}$ is rejected if posterior probability $\text{Prob}(q_1 > q_0 | \text{data}) > q^{**}$. Note that $q^{**}$ was a function of $q_0$. We found that setting $q^{**} = 0.97$ and 0.925 for $q_0 = 0.2$ and $q_0 = 0.5$ resulted in a matching type I error rate between SEARS and the Conventional design. Calibration of XJT only required 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. Results for all the other scenarios are provided in the supplementary material with similar findings. 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 the scenarios than the XJT and Conventional designs. The Con-
ventional design in general uses a little less than 180 patients for almost all the cases, where 180 is the specified maximum sample size for the Conventional design. The XJT design has a smaller sample size than Conventional. However, relative to the XJT and the Conventional approaches, sample size reduction of SEARS goes up to 50%. Importantly, this is at no cost to the power! The SEARS design is more powerful than the XJT design in all cases and is comparable to the Conventional design. As a further investigation, we carried out additional simulations (results not shown) and found that the power of SEARS would even be higher than Conventional if we used a maximum of 120 patients for the Conventional design. In those cases, the SEARS design still had smaller sample sizes.

The improvement of the SEARS over the XJT and Conventional designs is encouraging. It shows that 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.

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 (5). Since more than one dose could be selected in a trial, the sum of selection percentages across five doses might be larger than 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. Therefore, dose 3 is the most desirable dose. In both scenarios dose 3 has the highest selection percentages and receives the largest number of patients. Scenarios 3 and 9 represent the cases of decreasing efficacy response rates, in which we observe a good characteristic of the design that allocates more patients to the lower level doses. For instance, in Scenario 9, on average about 33 and 15 patients are allocated to the first two dose level, which accounts for 65% of total sample size. In both scenarios, dose 1 is the most desirable and is selected with the over 90% frequency. Scenarios 4 and 10 are unconventional cases in that the efficacy response
rates form an n-shaped pattern, that is, initially increasing then decreasing. Consistently the SEARS design chooses dose 3 as the most efficacious and safe dose. For example, in scenario 10, most patients (28.69) are allocated to dose 3, which exhibits a dose selection percentage of 54.5%. U-shaped response curve is present in Scenarios 5 and 11. Also, doses 4 and 5 are too toxic. In these cases a good design should allocate most patients to the lowest dose level and allocate few patients to the middle of the dose range. In scenario 5, we can see that SEARS allocates approximately 34 and 18 patients to the doses 1 and 2, respectively, exhibiting a desirable behavior. Lastly, in Scenarios 6 and 12, all doses are more efficacious than the control, and doses 5 and 6 are too toxic. Doses 1 and 2 are desirable, and they are selected with the highest frequencies and receive the largest numbers of patients.

3.3 A Trial Monitoring Demo

We carried out a hypothetical trial on computer using the SEARS design. A movie file was produced to demonstrate the trial conduct based on SEARS. The movie 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 by 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 trial ended after the maximum sample size was reached.

The movie file is available on-line at [http://odin.mdacc.tmc.edu/~yuanj/soft.html](http://odin.mdacc.tmc.edu/~yuanj/soft.html).


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 shared across for decision making.

The performance of the SEARS design is superior. Compared with a recently developed adaptive design in Xie et al. (2012), SEARS achieves higher power with smaller sample sizes. This is due to the efficient adaptive methods and the seamless rules employed in SEARS. Specifically, mTPI allows the flexibility of 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 and at the same time remove 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 limitation of the SEARS is that it assumes that the 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 a short-term response 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), could be applied in the extension.
Acknowledgement

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


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 patients randomized in Phase II.
Figure 2: Comparison of the conventional, XJT, and SEARS design. Trial type I error rate is matched for the Null scenario 1, and trial power and sample size are compared for the remaining scenarios.
Figure 3: Comparison of the conventional, XJT, and SEARS design. Trial type I error rate is matched for the Null scenario 7, 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 are named Null, Increasing, Decreasing, n-shaped, u-shaped, and Equal, reflecting the shape of efficacy response pattern. Scenarios 2-6 are the non-Null scenarios.

<table>
<thead>
<tr>
<th>Index</th>
<th>Scenario name</th>
<th>Control arm efficacy $q_0 = 0.2$</th>
<th>Control arm efficacy $q_0 = 0.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Null</td>
<td>0.2,0.2,0.2,0.2,0.2</td>
<td>0.5,0.5,0.5,0.5,0.5</td>
</tr>
<tr>
<td>2</td>
<td>Increasing</td>
<td>0.2,0.3,0.5,0.7,0.8</td>
<td>0.5,0.6,0.7,0.8,0.9</td>
</tr>
<tr>
<td>3</td>
<td>Decreasing</td>
<td>0.8,0.7,0.5,0.3,0.2</td>
<td>0.9,0.8,0.7,0.6,0.5</td>
</tr>
<tr>
<td>4</td>
<td>n-shaped</td>
<td>0.2,0.4,0.8,0.4,0.2</td>
<td>0.5,0.7,0.9,0.7,0.5</td>
</tr>
<tr>
<td>5</td>
<td>u-shaped</td>
<td>0.8,0.4,0.2,0.4,0.8</td>
<td>0.9,0.7,0.5,0.7,0.9</td>
</tr>
<tr>
<td>6</td>
<td>Equal</td>
<td>0.5,0.5,0.5,0.5,0.5</td>
<td>0.8,0.8,0.8,0.8,0.8</td>
</tr>
</tbody>
</table>
Table 2: Dose selection percentages and average numbers patients allocated under SEARS. The true toxicity rates for the five doses are (0.03,0.06,0.17,0.30,0.50).

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Control Arm Efficacy $q_0 = 0.2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sel %</td>
</tr>
<tr>
<td>1. Null</td>
<td>0.2 0.2 0.2 0.2 0.2</td>
</tr>
<tr>
<td>Sel %</td>
<td>7.1 6.9 4.0 0.5 0</td>
</tr>
<tr>
<td># Pts</td>
<td>21.52 11.96 3.67</td>
</tr>
<tr>
<td>2. Increasing</td>
<td>0.2 0.3 0.4 0.5 0.6</td>
</tr>
<tr>
<td>Sel %</td>
<td>5.8 36.4 42.1 6.3 0.1</td>
</tr>
<tr>
<td># Pts</td>
<td>25.27 14.55 4.34</td>
</tr>
<tr>
<td>3. Decreasing</td>
<td>0.6 0.5 0.4 0.3 0.2</td>
</tr>
<tr>
<td>Sel %</td>
<td>95.6 78.9 32.9 5.5 0.05</td>
</tr>
<tr>
<td># Pts</td>
<td>15.48 10.21 4.89</td>
</tr>
<tr>
<td>4. n-shaped</td>
<td>0.2 0.4 0.6 0.4 0.2</td>
</tr>
<tr>
<td>Sel %</td>
<td>4.6 62.6 55.2 7.6 0.05</td>
</tr>
<tr>
<td># Pts</td>
<td>28.84 11.65 4.2</td>
</tr>
<tr>
<td>5. u-shaped</td>
<td>0.6 0.4 0.2 0.4 0.6</td>
</tr>
<tr>
<td>Sel %</td>
<td>97.4 54.0 3.4 7.8 0.6</td>
</tr>
<tr>
<td># Pts</td>
<td>12.94 11.07 4.9</td>
</tr>
<tr>
<td>6. Equal</td>
<td>0.5 0.5 0.5 0.5 0.5</td>
</tr>
<tr>
<td>Sel %</td>
<td>83.1 81.9 49.6 8.8 0.3</td>
</tr>
<tr>
<td># Pts</td>
<td>21.68 13.08 5.67</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Control Arm Efficacy $q_0 = 0.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sel %</td>
</tr>
<tr>
<td>7. Null</td>
<td>0.5 0.5 0.5 0.5 0.5</td>
</tr>
<tr>
<td>Sel %</td>
<td>6.4 5.5 2.6 0.3 0.05</td>
</tr>
<tr>
<td># Pts</td>
<td>16.72 9.32 2.96</td>
</tr>
<tr>
<td>---</td>
<td>---------------</td>
</tr>
<tr>
<td>Sel %</td>
<td>0.5 0.6 0.7 0.8 0.9</td>
</tr>
<tr>
<td># Pts</td>
<td>4.9 26.2 38.1 5.9 0.1</td>
</tr>
</tbody>
</table>