Journal of Probability and Statistics

Adaptive and Sequential Methods for Clinical Trials

Guest Editors: Yichuan Zhao, Zhengjia Chen, Xuelin Huang, and Mourad Tighiouart

Adaptive and Sequential Methods for Clinical Trials

Journal of Probability and Statistics

Adaptive and Sequential Methods for Clinical Trials

Guest Editors: Yichuan Zhao, Zhengjia Chen, Xuelin Huang, and Mourad Tighiouart

Editorial Board

Contents

Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2013, Article ID 386058, 2 pages
http://dx.doi.org/10.1155/2013/386058

Editorial
Adaptive and Sequential Methods for Clinical Trials

Yichuan Zhao,[1] Zhengjia Chen,[2] Xuelin Huang,[3] and Mourad Tighiouart[4]

[1] *Department of Mathematics & Statistics, Georgia State University, Atlanta, GA 30303, USA*
[2] *Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA 30322, USA*
[3] *Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA*
[4] *Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA*

Correspondence should be addressed to Yichuan Zhao; yichuan@gsu.edu

Received 31 October 2012; Accepted 31 October 2012

The clinical trial, a prospective evaluation of the effect of interventions in humans under prespecified conditions, is a standard and integral part of modern medicine. Many adaptive and sequential approaches have been proposed for clinical trials, which allow for modifications to an ongoing trial without undermining the validity and integrity of the trial. The application of adaptive and sequential methods in clinical trials has significantly improved the flexibility, efficiency, therapeutic effect, and validity of such trials.

This special issue describes state-of-the-art statistical research in adaptive and sequential methods and the application of such methods in clinical trials. It provides 1 review article and 5 research articles contributed by some of the leading experts in this field. The review article gives a comprehensive overview of the outstanding methodology in the current literature that is related to adaptive and sequential clinical trials, while each of the 5 research articles addresses specific critical issues in contemporary clinical trials, as summarized below.

In the review paper "*Methodology and Application of Adaptive and Sequential Approaches in Contemporary Clinical Trials*," by Z. Chen et al., the most distinguished and applicable adaptive and sequential approaches, especially novel designs, are reviewed, compared, and contrasted according to the phase of clinical trial (phases I, II, and III) to which they are applied. The future directions of the related areas of research are also explored and discussed.

The research article entitled "*Exact group sequential methods for estimating a binomial proportion*," by Z. Chen and X. Chen, first reviews existing sequential methods for estimating a binomial proportion. A new family of group sequential sampling schemes is then proposed for estimating a binomial proportion with a prescribed margin of error, which achieves unprecedented efficiency while guaranteeing prespecified confidence levels.

The research article entitled "*Two-stage adaptive optimal design with fixed first-stage sample size*," by A. Lane and N. Flournoy, proposes a two-stage adaptive optimal design with a fixed first-stage sample size, as applied to a small pilot study of fixed size that is to be followed by a much larger experiment. The authors study the large sample behavior of their design by assuming a nonlinear regression model with normal errors and explicitly deriving the asymptotic distribution of the maximum likelihood estimate.

In the research paper "*Escalation with overdose control using ordinal toxicity grades for cancer phase I clinical trials*," by M. Tighiouart et al., the authors extend a Bayesian adaptive phase I clinical trial by introducing an intermediate-grade toxicity and show that the efficiency and safety of the trial are maintained and fewer patients are overdosed.

The research paper "*Incorporating a patient dichotomous characteristic in cancer phase I clinical trials using escalation with overdose control*," by M. Tighiouart et al., describes a design for phase I clinical trials in cancer that takes into account heterogeneity among patients, which is thought to be related to treatment susceptibility, and reduces the number of patients being overdosed.

In the research paper "*Number of patients per cohort and sample size considerations using dose escalation with overdose control*," M. Tighiouart and A. Rogatko compare the safety and efficiency of trials designed with three or only one patient per cohort and present the number of patients needed to design a trial to achieve a given accuracy of the estimate of the maximum tolerated dose.

As the editors of this special issue, we hope that readers of this special journal issue will find these articles representative of the contributions of this important research field to clinical trials, in terms of research methodology and its many practical applications. We thank the authors and reviewers for their significant contributions to the articles in this special issue. We also extend our thanks to the Hindis Publishing Corporation for their professional and efficient service.

Yichuan Zhao
Zhengjia Chen
Xuelin Huang
Mourad Tighiouart

Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2013, Article ID 603297, 24 pages
http://dx.doi.org/10.1155/2013/603297

Research Article

Exact Group Sequential Methods for Estimating a Binomial Proportion

Zhengjia Chen[1] and Xinjia Chen[2]

[1] *Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA 30322, USA*
[2] *Department of Electrical Engineering, Southern University and A&M College, Baton Rouge, LA 70813, USA*

Correspondence should be addressed to Zhengjia Chen; zchen38@emory.edu

Received 23 May 2012; Revised 1 October 2012; Accepted 9 October 2012

Academic Editor: Xuelin Huang

We first review existing sequential methods for estimating a binomial proportion. Afterward, we propose a new family of group sequential sampling schemes for estimating a binomial proportion with prescribed margin of error and confidence level. In particular, we establish the uniform controllability of coverage probability and the asymptotic optimality for such a family of sampling schemes. Our theoretical results establish the possibility that the parameters of this family of sampling schemes can be determined so that the prescribed level of confidence is guaranteed with little waste of samples. Analytic bounds for the cumulative distribution functions and expectations of sample numbers are derived. Moreover, we discuss the inherent connection of various sampling schemes. Numerical issues are addressed for improving the accuracy and efficiency of computation. Computational experiments are conducted for comparing sampling schemes. Illustrative examples are given for applications in clinical trials.

1. Introduction

Estimating a binomial proportion is a problem of ubiquitous significance in many areas of engineering and sciences. For economical reasons and other concerns, it is important to use as fewer as possible samples to guarantee the required reliability of estimation. To achieve this goal, sequential sampling schemes can be very useful. In a sequential sampling scheme, the total number of observations is not fixed in advance. The sampling process is continued stage by stage until a prespecified stopping rule is satisfied. The stopping rule is evaluated with accumulated observations. In many applications, for administrative feasibility, the sampling experiment is performed in a group fashion. Similar to group sequential tests [1, Section 8], [2], an estimation method based on taking samples by groups and evaluating them sequentially is referred to as a group sequential estimation method. It should be noted that group sequential estimation methods are general enough to include fixed-sample-size and fully sequential procedures as special cases. Particularly, a fixed-sample-size method can be viewed as a group sequential procedure of only one stage. If the increment between the sample sizes of consecutive stages is equal to 1, then the group sequential method is actually a fully sequential method.

It is a common contention that statistical inference, as a unique science to quantify the uncertainties of inferential statements, should avoid errors in the quantification of uncertainties, while minimizing the sampling cost. That is, a statistical inferential method is expected to be exact and efficient. The conventional notion of exactness is that no approximation is involved, except the round-off error due to finite word length of computers. Existing sequential methods for estimating a binomial proportion are dominantly of asymptotic nature (see, e.g., [3–7] and the references therein). Undoubtedly, asymptotic techniques provide approximate solutions and important insights for the relevant problems. However, any asymptotic method inevitably introduces unknown error in the resultant approximate solution due to the necessary use of a finite number of samples. In the direction of nonasymptotic sequential estimation, the primary goal is to ensure that the true coverage probability is above the prespecified confidence level for any value of

the associated parameter, while the required sample size is as low as possible. In this direction, Mendo and Hernando [8] developed an inverse binomial sampling scheme for estimating a binomial proportion with relative precision. Tanaka [9] developed a rigorous method for constructing fixed-width sequential confidence intervals for a binomial proportion. Although no approximation is involved, Tanaka's method is very conservative due to the bounding techniques employed in the derivation of sequential confidence intervals. Franzén [10] studied the construction of fixed-width sequential confidence intervals for a binomial proportion. However, no effective method for defining stopping rules is proposed in [10]. In his later paper [11], Franzén proposed to construct fixed-width confidence intervals based on sequential probability ratio tests (SPRTs) invented by Wald [12]. His method can generate fixed-sample-size confidence intervals based on SPRTs. Unfortunately, he made a fundamental flaw by mistaking that if the width of the fixed-sample-size confidence interval decreases to be smaller than the prespecified length as the number of samples is increasing, then the fixed-sample-size confidence interval at the termination of sampling process is the desired fixed-width sequential confidence interval guaranteeing the prescribed confidence level. More recently, Frey published a paper [13] in The American Statistician (TAS) on the classical problem of sequentially estimating a binomial proportion with prescribed margin of error and confidence level. Before Frey submitted his original manuscript to TAS in July 2009, a general framework of multistage parameter estimation had been established by Chen [14–18], which provides exact methods for estimating parameters of common distributions with various error criterion. This framework is also proposed in [19]. The approach of Frey [13] is similar to that of Chen [14–18] for the specific problem of estimating a binomial proportion with prescribed margin of error and confidence level.

In this paper, our primary interests are in the exact sequential methods for the estimation of a binomial proportion with prescribed margin of error and confidence level. We first introduce the exact approach established in [14–18]. In particular, we introduce the inclusion principle proposed in [18] and its applications to the construction of concrete stopping rules. We investigate the connection among various stopping rules. Afterward, we propose a new family of stopping rules which are extremely simple and accommodate some existing stopping rules as special cases. We provide rigorous justification for the feasibility and asymptotic optimality of such stopping rules. We prove that the prescribed confidence level can be guaranteed uniformly for all values of a binomial proportion by choosing appropriate parametric values for the stopping rule. We show that as the margin of error tends to be zero, the sample size tends to the attainable minimum as if the binomial proportion were exactly known. We derive analytic bounds for distributions and expectations of sample numbers. In addition, we address some critical computational issues and propose methods to improve the accuracy and efficiency of numerical calculation. We conduct extensive numerical experiment to study the performance of various stopping rules. We determine parametric values for the proposed stopping rules to achieve unprecedentedly efficiency while guaranteeing prescribed confidence levels. We attempt to make our proposed method as user-friendly as possible so that it can be immediately applicable even for layer persons.

The remainder of the paper is organized as follows. In Section 2, we introduce the exact approach proposed in [14–18]. In Section 3, we discuss the general principle of constructing stopping rules. In Section 4, we propose a new family of sampling schemes and investigate their feasibility, optimality, and analytic bounds of the distribution and expectation of sample numbers. In Section 5, we compare various computational methods. In particular, we illustrate why the natural method of evaluating coverage probability based on gridding parameter space is neither rigorous nor efficient. In Section 6, we present numerical results for various sampling schemes. In Section 7, we illustrate the applications of our group sequential method in clinical trials. Section 8 is the conclusion. The proofs of theorems are given in appendices. Throughout this paper, we shall use the following notations. The empty set is denoted by $\varnothing$. The set of positive integers is denoted by $\mathbb{N}$. The ceiling function is denoted by $\lceil\cdot\rceil$. The notation $\Pr\{E \mid \theta\}$ denotes the probability of the event E associated with parameter θ. The expectation of a random variable is denoted by $\mathbb{E}[\cdot]$. The standard normal distribution is denoted by $\Phi(\cdot)$. For $\alpha \in (0,1)$, the notation $\mathcal{Z}_\alpha$ denotes the critical value such that $\Phi(\mathcal{Z}_\alpha) = 1-\alpha$. For $n \in \mathbb{N}$, in the case that $X_1, \ldots, X_n$ are i.i.d. samples of X, we denote the sample mean $(\sum_{i=1}^{n} X_i)/n$ by $\overline{X}_n$, which is also called the relative frequency when X is a Bernoulli random variable. The other notations will be made clear as we proceed.

2. How Can It Be Exact?

In many areas of scientific investigation, the outcome of an experiment is of dichotomy nature and can be modeled as a Bernoulli random variable X, defined in probability space $(\Omega, \Pr, \mathcal{F})$, such that

$$\Pr\{X=1\} = 1 - \Pr\{X=0\} = p \in (0,1), \tag{1}$$

where p is referred to as a binomial proportion. In general, there is no analytic method for evaluating the binomial proportion p. A frequently used approach is to estimate p based on i.i.d. samples $X_1, X_2, \ldots$ of X. To reduce the sampling cost, it is appropriate to estimate p by a multistage sampling procedure. More formally, let $\epsilon \in (0,1)$ and $1-\delta$, with $\delta \in (0,1)$, be the prespecified margin of error and confidence level, respectively. The objective is to construct a sequential estimator $\widehat{\mathbf{p}}$ for p based on a multistage sampling scheme such that

$$\Pr\{|\widehat{\mathbf{p}} - p| < \epsilon \mid p\} \geq 1-\delta, \tag{2}$$

for any $p \in (0,1)$. Throughout this paper, the probability $\Pr\{|\widehat{\mathbf{p}} - p| < \epsilon \mid p\}$ is referred to as the *coverage probability*. Accordingly, the probability $\Pr\{|\widehat{\mathbf{p}} - p| \geq \epsilon \mid p\}$ is referred to as the *complementary coverage probability*. Clearly, a complete construction of a multistage estimation scheme needs to determine the number of stages, the sample sizes for all stages, the stopping rule, and the estimator for p. Throughout this paper, we let s denote the number of stages and let n_ℓ denote the number of samples at the ℓth stages. That is, the sampling process consists of s stages with sample sizes $n_1 < n_2 < \cdots < n_s$. For $\ell = 1, 2, \ldots, s$, define $K_\ell = \sum_{i=1}^{n_\ell} X_i$ and $\widehat{\mathbf{p}}_\ell = K_\ell / n_\ell$. The stopping rule is to be defined in terms of $\widehat{\mathbf{p}}_\ell$, $\ell = 1, \ldots, s$. Of course, the index of stage at the termination of the sampling process, denoted by $\mathbf{l}$, is a random number. Accordingly, the number of samples at the termination of the experiment, denoted by $\mathbf{n}$, is a random number which equals $n_{\mathbf{l}}$. Since for each ℓ, $\widehat{\mathbf{p}}_\ell$ is a maximum-likelihood and minimum-variance unbiased estimator of p, the sequential estimator for p is taken as

$$\widehat{\mathbf{p}} = \widehat{\mathbf{p}}_{\mathbf{l}} = \frac{\sum_{i=1}^{n_{\mathbf{l}}} X_i}{n_{\mathbf{l}}} = \frac{\sum_{i=1}^{\mathbf{n}} X_i}{\mathbf{n}}. \tag{3}$$

In the above discussion, we have outlined the general characteristics of a multistage sampling scheme for estimating a binomial proportion. It remains to determine the number of stages, the sample sizes for all stages, and the stopping rule so that the resultant estimator $\widehat{\mathbf{p}}$ satisfies (2) for any $p \in (0,1)$.

Actually, the problem of sequential estimation of a binomial proportion has been treated by Chen [14–18] in a general framework of multistage parameter estimation. The techniques of [14–18] are sufficient to offer exact solutions for a wide range of sequential estimation problems, including the estimation of a binomial proportion as a special case. The central idea of the approach in [14–18] is the control of coverage probability by a single parameter ζ, referred to as the *coverage tuning parameter*, and the adaptive rigorous checking of coverage guarantee by virtue of bounds of coverage probabilities. It is recognized in [14–18] that, due to the discontinuity of the coverage probability on parameter space, the conventional method of evaluating the coverage probability for a finite number of parameter values is neither rigorous not computationally efficient for checking the coverage probability guarantee.

As mentioned in the introduction, Frey published an article [13] in TAS on the sequential estimation of a binomial proportion with prescribed margin of error and confidence level. For clarity of presentation, the comparison of the works of Chen and Frey is given in Section 5.4. In the remainder of this section, we shall only introduce the idea and techniques of [14–18], which had been precedentially developed by Chen before Frey submitted his original manuscript to TAS in July 2009. We will introduce the approach of [14–18] with a focus on the special problem of estimating a binomial proportion with prescribed margin of error and confidence level.

2.1. Four Components Suffice. The exact methods of [14–18] for multistage parameter estimation have four main components as follows.

(i) Stopping rules parameterized by the coverage tuning parameter $\zeta > 0$ such that the associated coverage probabilities can be made arbitrarily close to 1 by choosing $\zeta > 0$ to be a sufficiently small number.

(ii) Recursively computable lower and upper bounds for the complementary coverage probability for a given ζ and an interval of parameter values.

(iii) Adapted branch and bound algorithm.

(iv) Bisection coverage tuning.

Without looking at the technical details, one can see that these four components are sufficient for constructing a sequential estimator so that the prescribed confidence level is guaranteed. The reason is as follows. As lower and upper bounds for the complementary coverage probability are available, the global optimization technique, branch and bound (B&B) algorithm [20], can be used to compute exactly the maximum of complementary coverage probability on the whole parameter space. Thus, it is possible to check rigorously whether the coverage probability associated with a given ζ is no less than the prespecified confidence level. Since the coverage probability can be controlled by ζ, it is possible to determine ζ as large as possible to guarantee the desired confidence level by a bisection search. This process is referred to as bisection coverage tuning in [14–18]. Since a critical subroutine needed for bisection coverage tuning is to check whether the coverage probability is no less than the prespecified confidence level, it is not necessary to compute exactly the maximum of the complementary coverage probability. Therefore, Chen revised the standard B&B algorithm to reduce the computational complexity and called the improved algorithm as the adapted B&B Algorithm. The idea is to adaptively partition the parameter space as many subintervals. If for all subintervals, the upper bounds of the complementary coverage probability are no greater than δ, then declare that the coverage probability is guaranteed. If there exists a subinterval for which the lower bound of the complementary coverage probability is greater than δ, then declare that the coverage probability is not guaranteed. Continue partitioning the parameter space if no decision can be made. The four components are illustrated in the sequel under the headings of stopping rules, interval bounding, adapted branch and bound, and bisection coverage tuning.

2.2. Stopping Rules. The first component for the exact sequential estimation of a binomial proportion is the stopping rule for constructing a sequential estimator such that the coverage probability can be controlled by the coverage tuning

parameter ζ. For convenience of describing some concrete stopping rules, define

$$\mathscr{M}(z,\theta)=\begin{cases} z\ln\dfrac{\theta}{z} & \\ \quad +(1-z)\ln\dfrac{1-\theta}{1-z} & \text{for } z\in(0,1),\\ & \theta\in(0,1),\\ \ln(1-\theta) & \text{for } z=0,\ \theta\in(0,1),\\ \ln\theta & \text{for } z=1,\ \theta\in(0,1),\\ -\infty & \text{for } z\in[0,1],\\ & \theta\notin(0,1),\end{cases}$$

$$S(k,l,n,p)=\begin{cases}\displaystyle\sum_{i=k}^{l}\binom{n}{i}p^{i}(1-p)^{n-i} & \text{for } p\in(0,1),\\ 0 & \text{for } p\notin(0,1),\end{cases} \tag{4}$$

where k and l are integers such that $0\le k\le l\le n$. Assume that $0<\zeta\delta<1$. For the purpose of controlling the coverage probability $\Pr\{|\widehat{\mathbf{p}}-\mathrm{p}|<\epsilon\mid p\}$ by the coverage tuning parameter, Chen has proposed four stopping rules as follows.

Stopping Rule A. Continue sampling until $\mathscr{M}((1/2)-|(1/2)-\widehat{\mathbf{p}}_\ell|,(1/2)-|(1/2)-\widehat{\mathbf{p}}_\ell|+\epsilon)\le(\ln(\zeta\delta))/n_\ell$ for some $\ell\in\{1,\dots,s\}$.

Stopping Rule B. Continue sampling until $(|\widehat{\mathbf{p}}_\ell-(1/2)|-(2/3)\epsilon)^2\ge(1/4)+(\epsilon^2 n_\ell/2\ln(\zeta\delta))$ for some $\ell\in\{1,\dots,s\}$.

Stopping Rule C. Continue sampling until $S(K_\ell,n_\ell,n_\ell,\widehat{\mathbf{p}}_\ell-\epsilon)\le\zeta\delta$ and $S(0,K_\ell,n_\ell,\widehat{\mathbf{p}}_\ell+\epsilon)\le\zeta\delta$ for some $\ell\in\{1,\dots,s\}$.

Stopping Rule D. Continue sampling until $n_\ell\ge\widehat{\mathbf{p}}_\ell(1-\widehat{\mathbf{p}}_\ell)(2/\epsilon^2)\ln(1/\zeta\delta)$ for some $\ell\in\{1,\dots,s\}$.

Stopping Rule A was first proposed in [14, Theorem 7] and restated in [15, Theorem 16]. Stopping Rule B was first proposed in [16, Theorem 1] and represented as the third stopping rule in [21, Section 4.1.1]. Stopping Rule C originated from [17, Theorem 1] and was restated as the first stopping rule in [21, Section 4.1.1]. Stopping Rule D was described in the remarks following Theorem 7 of [22]. All these stopping rules can be derived from the general principles proposed in [18, Section 3] and [19, Section 2.4].

Given that a stopping rule can be expressed in terms of $\widehat{\mathbf{p}}_\ell$ and n_ℓ for $\ell=1,\dots,s$, it is possible to find a bivariate function $\mathscr{D}(\cdot,\cdot)$ on $\{(z,n):z\in[0,1],\ n\in\mathbb{N}\}$, taking values from $\{0,1\}$, such that the stopping rule can be stated as the following: continue sampling until $\mathscr{D}(\widehat{\mathbf{p}}_\ell,n_\ell)=1$ for some $\ell\in\{1,\dots,s\}$. It can be checked that such representation applies to Stopping Rules A, B, C, and D. For example, Stopping Rule B can be expressed in this way by virtue of function $\mathscr{D}(\cdot,\cdot)$ such that

$$\mathscr{D}(z,n)=\begin{cases}1 & \text{if } \left(\left|z-\dfrac{1}{2}\right|-\dfrac{2}{3}\epsilon\right)^2\ge\dfrac{1}{4}+\dfrac{\epsilon^2 n}{2\ln(\zeta\delta)},\\ 0 & \text{otherwise.}\end{cases} \tag{5}$$

The motivation of introducing function $\mathscr{D}(\cdot,\cdot)$ is to parameterize the stopping rule in terms of design parameters. Function $\mathscr{D}(\cdot,\cdot)$ determines the form of the stopping rule and, consequently, the sample sizes for all stages can be chosen as functions of design parameters. Specifically, let

$$N_{\min}=\min\left\{n\in\mathbb{N}:\mathscr{D}\left(\frac{k}{n},n\right)=1 \text{ for some nonnegative integer } k \text{ not exceeding } n\right\}, \tag{6}$$

$$N_{\max}=\min\left\{n\in\mathbb{N}:\mathscr{D}\left(\frac{k}{n},n\right)=1 \text{ for all nonnegative integer } k \text{ not exceeding } n\right\}. \tag{7}$$

To avoid unnecessary checking of the stopping criterion and thus reduce administrative cost, there should be a possibility that the sampling process is terminated at the first stage. Hence, the minimum sample size n_1 should be chosen to ensure that $\{\mathbf{n}=n_1\}\neq\varnothing$. This implies that the sample size n_1 for the first stage can be taken as $N_{\min}$. On the other hand, since the sampling process must be terminated at or before the sth stage, the maximum sample size n_s should be chosen to guarantee that $\{\mathbf{n}>n_s\}=\varnothing$. This implies that the sample size n_s for the last stage can be taken as $N_{\max}$. If the number of stages s is given, then the sample sizes for stages in between 1 and s can be chosen as $s-2$ integers between $N_{\min}$ and $N_{\max}$. Particularly, if the group sizes are expected to be approximately equal, then the sample sizes can be taken as

$$n_\ell=\left\lceil N_{\min}+\frac{\ell-1}{s-1}(N_{\max}-N_{\min})\right\rceil,\quad \ell=1,\dots,s. \tag{8}$$

Since the stopping rule is associated with the coverage tuning parameter ζ, it follows that the number of stages s and the sample sizes $n_1,n_2,\dots,n_s$ can be expressed as functions of ζ. In this sense, it can be said that the stopping rule is parameterized by the coverage tuning parameter ζ. The above method of parameterizing stopping rules has been used in [14–17] and proposed in [21, Section 2.1, page 9].

2.3. Interval Bounding. The second component for the exact sequential estimation of a binomial proportion is the method of bounding the complementary coverage probability $\Pr\{|\widehat{\mathbf{p}}-p|\ge\epsilon\mid p\}$ for p in an interval $[a,b]$ contained by interval $(0,1)$. Applying Theorem 8 of [15] to the special case of a Bernoulli distribution immediately yields

$$\begin{aligned}&\Pr\{\widehat{\mathbf{p}}\le a-\epsilon\mid b\}+\Pr\{\widehat{\mathbf{p}}\ge b+\epsilon\mid a\}\\ &\quad\le\Pr\{|\widehat{\mathbf{p}}-p|\ge\epsilon\mid p\}\\ &\quad\le\Pr\{\widehat{\mathbf{p}}\le b-\epsilon\mid a\}+\Pr\{\widehat{\mathbf{p}}\ge a+\epsilon\mid b\},\end{aligned} \tag{9}$$

∇ Let $k \leftarrow 0$, $l_0 \leftarrow \Psi_{\text{lb}}(\mathcal{I}_{\text{init}})$ and $u_0 \leftarrow \Psi_{\text{ub}}(\mathcal{I}_{\text{init}})$.
∇ Let $\mathcal{S}_0 \leftarrow \{\mathcal{I}_{\text{init}}\}$ if $u_0 > \delta$. Otherwise, let $\mathcal{S}_0$ be empty.
∇ While $\mathcal{S}_k$ is nonempty, $l_k < \delta$ and u_k is greater than $\max\{l_k + \eta, \delta\}$, do the following:
⋄ Split each interval in $\mathcal{S}_k$ as two new intervals of equal length.
Let S_k denote the set of all new intervals obtained from this splitting procedure.
⋄ Eliminate any interval $\mathcal{I}$ from S_k such that $\Psi_{\text{ub}}(\mathcal{I}) \leq \delta$.
⋄ Let $\mathcal{S}_{k+1}$ be the set S_k processed by the above elimination procedure.
⋄ Let $l_{k+1} \leftarrow \max_{\mathcal{I} \in \mathcal{S}_{k+1}} \Psi_{\text{lb}}(\mathcal{I})$ and $u_{k+1} \leftarrow \max_{\mathcal{I} \in \mathcal{S}_{k+1}} \Psi_{\text{ub}}(\mathcal{I})$. Let $k \leftarrow k + 1$.
∇ If $\mathcal{S}_k$ is empty and $l_k < \delta$, then declare $\max \Psi(\mathcal{I}_{\text{init}}) \leq \delta$.
Otherwise, declare $\max \Psi(\mathcal{I}_{\text{init}}) > \delta$.

Algorithm 1

for all $p \in [a, b] \subseteq (0, 1)$. The bounds of (9) can be shown as follows. Note that $\Pr\{\hat{\mathbf{p}} \leq a - \epsilon \mid p\} + \Pr\{\hat{\mathbf{p}} \geq b + \epsilon \mid p\} \leq \Pr\{|\hat{\mathbf{p}} - p| \geq \epsilon \mid p\} = \Pr\{\hat{\mathbf{p}} \leq p - \epsilon \mid p\} + \Pr\{\hat{\mathbf{p}} \geq p + \epsilon \mid p\} \leq \Pr\{\hat{\mathbf{p}} \leq b - \epsilon \mid p\} + \Pr\{\hat{\mathbf{p}} \geq a + \epsilon \mid p\}$ for $p \in [a, b] \subseteq (0, 1)$. As a consequence of the monotonicity of $\Pr\{\hat{\mathbf{p}} \geq \vartheta \mid p\}$ and $\Pr\{\hat{\mathbf{p}} \leq \vartheta \mid p\}$ with respect to p, where ϑ is a real number independent of p, the lower and upper bounds of $\Pr\{|\hat{\mathbf{p}} - p| \geq \epsilon \mid p\}$ for $p \in [a, b] \subseteq (0, 1)$ can be given as $\Pr\{\hat{\mathbf{p}} \leq a - \epsilon \mid b\} + \Pr\{\hat{\mathbf{p}} \geq b + \epsilon \mid a\}$ and $\Pr\{\hat{\mathbf{p}} \leq b - \epsilon \mid a\} + \Pr\{\hat{\mathbf{p}} \geq a + \epsilon \mid b\}$, respectively.

In page 15, equation (1) of [15], Chen proposed to apply the recursive method of Schultz et al. [23, Section 2] to compute the lower and upper bounds of $\Pr\{|\hat{\mathbf{p}} - p| \geq \epsilon \mid p\}$ given by (9). It should be pointed out that such lower and upper bounds of $\Pr\{|\hat{\mathbf{p}} - p| \geq \epsilon \mid p\}$ can also be computed by the recursive path-counting method of Franzén [10, page 49].

2.4. Adapted Branch and Bound. The third component for the exact sequential estimation of a binomial proportion is the adapted B&B algorithm, which was proposed in [15, Section 2.8], for quick determination of whether the coverage probability is no less than $1 - \delta$ for any value of the associated parameter. Such a task of checking the coverage probability is also referred to as checking the coverage probability guarantee. Given that lower and upper bounds of the complementary coverage probability on an interval of parameter values can be obtained by the interval bounding techniques, this task can be accomplished by applying the B&B algorithm [20] to compute exactly the maximum of the complementary coverage probability on the parameter space. However, in our applications, it suffices to determine whether the maximum of the complementary coverage probability $\Pr\{|\hat{\mathbf{p}} - p| \geq \epsilon \mid p\}$ with respect to $p \in (0, 1)$ is greater than the confidence parameter δ. For fast checking whether the maximal complementary coverage probability exceeds δ, Chen proposed to reduce the computational complexity by revising the standard B&B algorithm as the Adapted B&B Algorithm in [15, Section 2.8]. To describe this algorithm, let $\mathcal{I}_{\text{init}}$ denote the parameter space $(0, 1)$. For an interval $\mathcal{I} \subseteq \mathcal{I}_{\text{init}}$, let $\max \Psi(\mathcal{I})$ denote the maximum of the complementary coverage probability $\Pr\{|\hat{\mathbf{p}} - p| \geq \epsilon \mid p\}$ with respect to $p \in \mathcal{I}$. Let $\Psi_{\text{lb}}(\mathcal{I})$ and $\Psi_{\text{ub}}(\mathcal{I})$ be, respectively, the lower and upper bounds of $\Psi(\mathcal{I})$, which can be obtained by the interval bounding techniques introduced in Section 2.3. Let $\eta > 0$ be a prespecified tolerance, which is much smaller than δ. The adapted B&B algorithm of [15] is represented with a slight modification as in Algorithm 1.

It should be noted that for a sampling scheme of symmetrical stopping boundary, the initial interval $\mathcal{I}_{\text{init}}$ may be taken as $(0, 1/2)$ for the sake of efficiency. In Section 5.1, we will illustrate why the adapted B&B algorithm is superior than the direct evaluation based on gridding parameter space. As will be seen in Section 5.2, the objective of the adapted B&B algorithm can also be accomplished by the Adaptive Maximum Checking Algorithm due to Chen [21, Section 3.3] and rediscovered by Frey [13, Appendix]. An explanation is given in Section 5.3 for the advantage of working with the complementary coverage probability.

2.5. Bisection Coverage Tuning. The fourth component for the exact sequential estimation of a binomial proportion is Bisection Coverage Tuning. Based on the adaptive rigorous checking of coverage probability, Chen proposed in [14, Section 2.7] and [15, Section 2.6] to apply a bisection search method to determine maximal ζ such that the coverage probability is no less than $1 - \delta$ for any value of the associated parameter. Moreover, Chen has developed asymptotic results in [15, page 21, Theorem 18] for determining the initial interval of ζ needed for the bisection search. Specifically, if the complementary coverage probability $\Pr\{|\hat{\mathbf{p}} - p| \geq \epsilon \mid p\}$ associated with $\zeta = \zeta_0$ tends to δ as $\epsilon \to 0$, then the initial interval of ζ can be taken as $[\zeta_0 2^i, \zeta_0 2^{i+1}]$, where i is the largest integer such that the complementary coverage probability associated with $\zeta = \zeta_0 2^i$ is no greater than δ for all $p \in (0, 1)$. By virtue of a bisection search, it is possible to obtain $\zeta^* \in [\zeta_0 2^i, \zeta_0 2^{i+1}]$ such that the complementary coverage probability associated with $\zeta = \zeta^*$ is guaranteed to be no greater than δ for all $p \in (0, 1)$.

3. Principle of Constructing Stopping Rules

In this section, we shall illustrate the inherent connection between various stopping rules. It will be demonstrated that a lot of stopping rules can be derived by virtue of the inclusion principle proposed by Chen [18, Section 3].

3.1. Inclusion Principle. The problem of estimating a binomial proportion can be considered as a special case of parameter estimation for a random variable X parameterized by $\theta \in \Theta$, where the objective is to construct a sequential estimator $\widehat{\theta}$ for θ such that $\Pr\{|\widehat{\theta} - \theta| < \epsilon \mid \theta\} \geq 1 - \delta$ for any $\theta \in \Theta$. Assume that the sampling process consists of s stages with sample sizes $n_1 < n_2 < \cdots < n_s$. For $\ell = 1, \ldots, s$, define an estimator $\widehat{\theta}_\ell$ for θ in terms of samples $X_1, \ldots, X_{n_\ell}$ of X. Let $[L_\ell, U_\ell]$, $\ell = 1, 2, \ldots, s$ be a sequence of confidence intervals such that for any ℓ, $[L_\ell, U_\ell]$ is defined in terms of $X_1, \ldots, X_{n_\ell}$ and that the coverage probability $\Pr\{L_\ell \leq \theta \leq U_\ell \mid \theta\}$ can be made arbitrarily close to 1 by choosing $\zeta > 0$ to be a sufficiently small number. In Theorem 2 of [18], Chen proposed the following general stopping rule:

$$\text{Continue sampling until } U_\ell - \epsilon \leq \widehat{\theta}_\ell \leq L_\ell + \epsilon \text{ for some } \ell \in \{1, \ldots, s\}. \tag{10}$$

At the termination of the sampling process, a sequential estimator for θ is taken as $\widehat{\theta} = \widehat{\theta}_{\mathbf{l}}$, where $\mathbf{l}$ is the index of stage at the termination of sampling process.

Clearly, the general stopping rule (10) can be restated as follows.

Continue sampling until the confidence interval $[L_\ell, U_\ell]$ is included by interval $[\widehat{\theta}_\ell - \epsilon, \widehat{\theta}_\ell + \epsilon]$ for some $\ell \in \{1, \ldots, s\}$.

The sequence of confidence intervals are parameterized by ζ for purpose of controlling the coverage probability $\Pr\{|\widehat{\theta} - \theta| < \epsilon \mid \theta\}$. Due to the inclusion relationship $[L_\ell, U_\ell] \subseteq [\widehat{\theta}_\ell - \epsilon, \widehat{\theta}_\ell + \epsilon]$, such a general methodology of using a sequence of confidence intervals to construct a stopping rule for controlling the coverage probability is referred to as the *inclusion principle*. It is asserted by Theorem 2 of [18] that

$$\Pr\left\{\left|\widehat{\theta} - \theta\right| < \epsilon \mid \theta\right\} \geq 1 - s\zeta\delta, \quad \forall \theta \in \Theta, \tag{11}$$

provided that $\Pr\{L_\ell < \theta < U_\ell \mid \theta\} \geq 1 - \zeta\delta$ for $\ell = 1, \ldots, s$ and $\theta \in \Theta$. This demonstrates that if the number of stages s is bounded respective to ζ, then the coverage probability $\Pr\{|\widehat{\theta} - \theta| < \epsilon \mid \theta\}$ associated with the stopping rule derived from the inclusion principle can be controlled by ζ. Actually, before explicitly proposing the inclusion principle in [18], Chen had extensively applied the inclusion principle in [14–17] to construct stopping rules for estimating parameters of various distributions such as binomial, Poisson, geometric, hypergeometric, and normal distributions. A more general version of the inclusion principle is proposed in [19, Section 2.4]. For simplicity of the stopping rule, Chen had made effort to eliminate the computation of confidence limits.

In the context of estimating a binomial proportion p, the inclusion principle immediately leads to the following general stopping rule:

$$\text{Continue sampling until } \widehat{\mathbf{p}}_\ell - \epsilon \leq L_\ell \leq U_\ell \leq \widehat{\mathbf{p}}_\ell + \epsilon \text{ for some } \ell \in \{1, \ldots, s\}. \tag{12}$$

Consequently, the sequential estimator for p is taken as $\widehat{\mathbf{p}}$ according to (3). It should be pointed out that the stopping rule (12) had been rediscovered by Frey in Section 2, the 1st paragraph of [13]. The four stopping rules considered in his paper follow immediately from applying various confidence intervals to the general stopping rule (12).

In the sequel, we will illustrate how to apply (12) to the derivation of Stopping Rules A, B, C, and D introduced in Section 2.2 and other specific stopping rules.

3.2. Stopping Rule from Wald Intervals. By virtue of Wald's method of interval estimation for a binomial proportion p, a sequence of confidence intervals $[L_\ell, U_\ell]$, $\ell = 1, \ldots, s$ for p can be constructed such that

$$L_\ell = \widehat{\mathbf{p}}_\ell - \mathcal{Z}_{\zeta\delta}\sqrt{\frac{\widehat{\mathbf{p}}_\ell(1 - \widehat{\mathbf{p}}_\ell)}{n_\ell}}, \quad U_\ell = \widehat{\mathbf{p}}_\ell + \mathcal{Z}_{\zeta\delta}\sqrt{\frac{\widehat{\mathbf{p}}_\ell(1 - \widehat{\mathbf{p}}_\ell)}{n_\ell}}, \quad \ell = 1, \ldots, s, \tag{13}$$

and that $\Pr\{L_\ell \leq p \leq U_\ell \mid p\} \approx 1 - 2\zeta\delta$ for $\ell = 1, \ldots, s$ and $p \in (0, 1)$. Note that, for $\ell = 1, \ldots, s$, the event $\{\widehat{\mathbf{p}}_\ell - \epsilon \leq L_\ell \leq U_\ell \leq \widehat{\mathbf{p}}_\ell + \epsilon\}$ is the same as the event $\{(\widehat{\mathbf{p}}_\ell - 1/2)^2 \geq (1/4) - n_\ell(\epsilon/\mathcal{Z}_{\zeta\delta})^2\}$. So, applying this sequence of confidence intervals to (12) results in the stopping rule "continue sampling until $(\widehat{\mathbf{p}}_\ell - 1/2)^2 \geq (1/4) - n_\ell(\epsilon/\mathcal{Z}_{\zeta\delta})^2$ for some $\ell \in \{1, \ldots, s\}$". Since for any $\zeta \in (0, 1/\delta)$, there exists a unique number $\zeta' \in (0, 1/\delta)$ such that $\mathcal{Z}_{\zeta\delta} = \sqrt{2\ln(1/\zeta'\delta)}$, this stopping rule is equivalent to "Continue sampling until $(\widehat{\mathbf{p}}_\ell - 1/2)^2 \geq (1/4) + (\epsilon^2 n_\ell/2\ln(\zeta\delta))$ for some $\ell \in \{1, \ldots, s\}$." This stopping rule is actually the same as Stopping Rule D, since $\{(\widehat{\mathbf{p}}_\ell - 1/2)^2 \geq (1/4) + (\epsilon^2 n_\ell/2\ln(\zeta\delta))\} = \{n_\ell \geq \widehat{\mathbf{p}}_\ell(1 - \widehat{\mathbf{p}}_\ell)(2/\epsilon^2)\ln(1/\zeta\delta)\}$ for $\ell \in \{1, \ldots, s\}$.

3.3. Stopping Rule from Revised Wald Intervals. Define $\widetilde{\mathbf{p}}_\ell = (n_\ell\,\widehat{\mathbf{p}}_\ell + a)/(n_\ell + 2a)$ for $\ell = 1, \ldots, s$, where a is a positive number. Inspired by Wald's method of interval estimation for p, a sequence of confidence intervals $[L_\ell, U_\ell]$, $\ell = 1, \ldots, s$ can be constructed such that

$$L_\ell = \widehat{\mathbf{p}}_\ell - \mathcal{Z}_{\zeta\delta}\sqrt{\frac{\widetilde{\mathbf{p}}_\ell(1 - \widetilde{\mathbf{p}}_\ell)}{n_\ell}}, \qquad U_\ell = \widehat{\mathbf{p}}_\ell + \mathcal{Z}_{\zeta\delta}\sqrt{\frac{\widetilde{\mathbf{p}}_\ell(1 - \widetilde{\mathbf{p}}_\ell)}{n_\ell}} \tag{14}$$

and that $\Pr\{L_\ell \leq p \leq U_\ell \mid p\} \approx 1 - 2\zeta\delta$ for $\ell = 1, \ldots, s$ and $p \in (0, 1)$. This sequence of confidence intervals was applied by Frey [13] to the general stopping rule (12). As a matter of fact, such idea of revising Wald interval $[\,\overline{X}_n - \mathcal{Z}_{\zeta\delta}\sqrt{(\overline{X}_n(1 - \overline{X}_n))/n},\ \overline{X}_n + \mathcal{Z}_{\zeta\delta}\sqrt{(\overline{X}_n(1 - \overline{X}_n))/n}\,]$ by replacing the relative frequency $\overline{X}_n = (\sum_{i=1}^n X_i)/n$ involved in the confidence limits with $\widetilde{p}_a = (n\overline{X}_n + a)/(n + 2a)$ had been proposed by Chen [24, Section 4].

As can be seen from Section 2, page 243, of Frey [13], applying (12) with the sequence of revised Wald intervals yields the stopping rule "Continue sampling until

$(\widetilde{\mathbf{p}}_\ell - 1/2)^2 \geq (1/4) + (\epsilon^2 n_\ell / 2\ln(\zeta\delta))$ for some $\ell \in \{1, \ldots, s\}$." Clearly, replacing $\widehat{\mathbf{p}}_\ell$ in Stopping Rule D with $\widetilde{\mathbf{p}}_\ell = (a + n_\ell \widehat{\mathbf{p}}_\ell)/(n_\ell + 2a)$ also leads to this stopping rule.

3.4. Stopping Rule from Wilson's Confidence Intervals. Making use of the interval estimation method of Wilson [25], one can obtain a sequence of confidence intervals $[L_\ell, U_\ell]$, $\ell = 1, \ldots, s$ for p such that

$$L_\ell = \max\left\{0, \frac{\widehat{\mathbf{p}}_\ell + \left(\mathcal{Z}^2_{\zeta\delta}/2n_\ell\right)}{1 + \left(\mathcal{Z}^2_{\zeta\delta}/n_\ell\right)} - \frac{\mathcal{Z}_{\zeta\delta}\sqrt{\left(\widehat{\mathbf{p}}_\ell\left(1 - \widehat{\mathbf{p}}_\ell\right)/n_\ell\right) + \left(\mathcal{Z}_{\zeta\delta}/2n_\ell\right)^2}}{1 + \left(\mathcal{Z}^2_{\zeta\delta}/n_\ell\right)}\right\},$$

$$U_\ell = \min\left\{1, \frac{\widehat{\mathbf{p}}_\ell + \left(\mathcal{Z}^2_{\zeta\delta}/2n_\ell\right)}{1 + \left(\mathcal{Z}^2_{\zeta\delta}/n_\ell\right)} + \frac{\mathcal{Z}_{\zeta\delta}\sqrt{\left(\widehat{\mathbf{p}}_\ell\left(1 - \widehat{\mathbf{p}}_\ell\right)/n_\ell\right) + \left(\mathcal{Z}_{\zeta\delta}/2n_\ell\right)^2}}{1 + \left(\mathcal{Z}^2_{\zeta\delta}/n_\ell\right)}\right\} \tag{15}$$

and that $\Pr\{L_\ell \leq p \leq U_\ell \mid p\} \approx 1 - 2\zeta\delta$ for $\ell = 1, \ldots, s$ and $p \in (0, 1)$. It should be pointed out that the sequence of Wilson's confidence intervals has been applied by Frey [13, Section 2, page 243] to the general stopping rule (12) for estimating a binomial proportion.

Since a stopping rule directly involves the sequence of Wilson's confidence intervals is cumbersome, it is desirable to eliminate the computation of Wilson's confidence intervals in the stopping rule. For this purpose, we need to use the following result.

Theorem 1. *Assume that* $0 < \zeta\delta < 1$ *and* $0 < \epsilon < 1/2$. *Then, Wilson's confidence intervals satisfy* $\{\widehat{\mathbf{p}}_\ell - \epsilon \leq L_\ell \leq U_\ell \leq \widehat{\mathbf{p}}_\ell + \epsilon\} = \{(|\widehat{\mathbf{p}}_\ell - 1/2| - \epsilon)^2 \geq (1/4) - n_\ell(\epsilon/\mathcal{Z}_{\zeta\delta})^2\}$ *for* $\ell = 1, \ldots, s$.

See Appendix A for a proof. As a consequence of Theorem 1 and the fact that for any $\zeta \in (0, 1/\delta)$, there exists a unique number $\zeta' \in (0, 1/\delta)$ such that $\mathcal{Z}_{\zeta\delta} = \sqrt{2\ln(1/\zeta'\delta)}$, applying the sequence of Wilson's confidence intervals to (12) leads to the following stopping rule.

Continue sampling until

$$\left(\left|\widehat{\mathbf{p}}_\ell - \frac{1}{2}\right| - \epsilon\right)^2 \geq \frac{1}{4} + \frac{\epsilon^2 n_\ell}{2\ln(\zeta\delta)}, \tag{16}$$

for some $\ell \in \{1, \ldots, s\}$.

3.5. Stopping Rule from Clopper-Pearson Confidence Intervals. Applying the interval estimation method of Clopper-Pearson [26], a sequence of confidence intervals $[L_\ell, U_\ell]$, $\ell = 1, \ldots, s$ for p can be obtained such that $\Pr\{L_\ell \leq p \leq U_\ell \mid p\} \geq 1 - 2\zeta\delta$ for $\ell = 1, \ldots, s$ and $p \in (0, 1)$, where the upper confidence limit U_ℓ satisfies the equation $S(0, K_\ell, n_\ell, U_\ell) = \zeta\delta$ if $K_\ell < n_\ell$; and the lower confidence limit L_ℓ satisfies the equation $S(K_\ell, n_\ell, n_\ell, L_\ell) = \zeta\delta$ if $K_\ell > 0$. The well-known equation (10.8) in [27, page 173] implies that $S(0, k, n, p)$, with $0 \leq k < n$, is decreasing with respect to $p \in (0, 1)$ and that $S(k, n, n, p)$, with $0 < k \leq n$, is increasing with respect to $p \in (0, 1)$. It follows that

$$\begin{aligned}
\left\{\widehat{\mathbf{p}}_\ell - \epsilon \leq L_\ell\right\} &= \left\{0 < \widehat{\mathbf{p}}_\ell - \epsilon \leq L_\ell\right\} \cup \left\{\widehat{\mathbf{p}}_\ell \leq \epsilon\right\} \\
&= \left\{\widehat{\mathbf{p}}_\ell > \epsilon,\ S\left(K_\ell, n_\ell, n_\ell, \widehat{\mathbf{p}}_\ell - \epsilon\right) \leq \zeta\delta\right\} \\
&\quad \cup \left\{\widehat{\mathbf{p}}_\ell \leq \epsilon\right\} \\
&= \left\{\widehat{\mathbf{p}}_\ell > \epsilon,\ S\left(K_\ell, n_\ell, n_\ell, \widehat{\mathbf{p}}_\ell - \epsilon\right) \leq \zeta\delta\right\} \\
&\quad \cup \left\{\widehat{\mathbf{p}}_\ell \leq \epsilon,\ S\left(K_\ell, n_\ell, n_\ell, \widehat{\mathbf{p}}_\ell - \epsilon\right) \leq \zeta\delta\right\} \\
&= \left\{S\left(K_\ell, n_\ell, n_\ell, \widehat{\mathbf{p}}_\ell - \epsilon\right) \leq \zeta\delta\right\}, \\
\left\{\widehat{\mathbf{p}}_\ell + \epsilon \geq U_\ell\right\} &= \left\{1 > \widehat{\mathbf{p}}_\ell + \epsilon \geq U_\ell\right\} \cup \left\{\widehat{\mathbf{p}}_\ell \geq 1 - \epsilon\right\} \\
&= \left\{\widehat{\mathbf{p}}_\ell < 1 - \epsilon,\ S\left(0, K_\ell, n_\ell, \widehat{\mathbf{p}}_\ell + \epsilon\right) \leq \zeta\delta\right\} \\
&\quad \cup \left\{\widehat{\mathbf{p}}_\ell \geq 1 - \epsilon\right\} \\
&= \left\{\widehat{\mathbf{p}}_\ell < 1 - \epsilon,\ S\left(0, K_\ell, n_\ell, \widehat{\mathbf{p}}_\ell + \epsilon\right) \leq \zeta\delta\right\} \\
&\quad \cup \left\{\widehat{\mathbf{p}}_\ell \geq 1 - \epsilon,\ S\left(0, K_\ell, n_\ell, \widehat{\mathbf{p}}_\ell + \epsilon\right) \leq \zeta\delta\right\} \\
&= \left\{S\left(0, K_\ell, n_\ell, \widehat{\mathbf{p}}_\ell + \epsilon\right) \leq \zeta\delta\right\},
\end{aligned} \tag{17}$$

for $\ell = 1, \ldots, s$. Consequently,

$$\begin{aligned}
&\left\{\widehat{\mathbf{p}}_\ell - \epsilon \leq L_\ell \leq U_\ell \leq \widehat{\mathbf{p}}_\ell + \epsilon\right\} \\
&\quad = \left\{S\left(K_\ell, n_\ell, n_\ell, \widehat{\mathbf{p}}_\ell - \epsilon\right) \leq \zeta\delta,\ S\left(0, K_\ell, n_\ell, \widehat{\mathbf{p}}_\ell + \epsilon\right) \leq \zeta\delta\right\},
\end{aligned} \tag{18}$$

for $\ell = 1, \ldots, s$. This demonstrates that applying the sequence of Clopper-Pearson confidence intervals to the general stopping rule (12) gives Stopping Rule C.

It should be pointed out that Stopping Rule C was rediscovered by Frey as the third stopping rule in Section 2, page 243 of his paper [13].

3.6. Stopping Rule from Fishman's Confidence Intervals. By the interval estimation method of Fishman [28], a sequence of confidence intervals $[L_\ell, U_\ell]$, $\ell = 1, \ldots, s$ for p can be obtained such that

$$L_\ell = \begin{cases} 0 & \text{if } \widehat{\mathbf{p}}_\ell = 0, \\ \left\{\theta_\ell \in \left(0, \widehat{\mathbf{p}}_\ell\right) : \mathcal{M}\left(\widehat{\mathbf{p}}_\ell, \theta_\ell\right) = \dfrac{\ln(\zeta\delta)}{n_\ell}\right\} & \text{if } \widehat{\mathbf{p}}_\ell > 0, \end{cases}$$

$$U_\ell = \begin{cases} 1 & \text{if } \widehat{\mathbf{p}}_\ell = 1, \\ \left\{\theta_\ell \in \left(\widehat{\mathbf{p}}_\ell, 1\right) : \mathcal{M}\left(\widehat{\mathbf{p}}_\ell, \theta_\ell\right) = \dfrac{\ln(\zeta\delta)}{n_\ell}\right\} & \text{if } \widehat{\mathbf{p}}_\ell < 1. \end{cases} \tag{19}$$

Under the assumption that $0 < \zeta\delta < 1$ and $0 < \epsilon < 1/2$, by similar techniques as the proof of Theorem 7 of [22], it can be shown that $\{\widehat{\mathbf{p}}_\ell - \epsilon \le L_\ell \le U_\ell \le \widehat{\mathbf{p}}_\ell + \epsilon\} = \{\mathscr{M}((1/2) - |(1/2) - \widehat{\mathbf{p}}_\ell|, (1/2) - |(1/2) - \widehat{\mathbf{p}}_\ell| + \epsilon) \le (\ln(\zeta\delta))/n_\ell\}$ for $\ell = 1, \ldots, s$. Therefore, applying the sequence of confidence intervals of Fishman to the general stopping rule (12) gives Stopping Rule A.

It should be noted that Fishman's confidence intervals are actually derived from the Chernoff bounds of the tailed probabilities of the sample mean of Bernoulli random variable. Hence, Stopping Rule A is also referred to as the stopping rule from Chernoff bounds in this paper.

3.7. Stopping Rule from Confidence Intervals of Chen et al. Using the interval estimation method of Chen et al. [29], a sequence of confidence intervals $[L_\ell, U_\ell]$, $\ell = 1, \ldots, s$ for p can be obtained such that

$$L_\ell = \max\left\{0, \widehat{\mathbf{p}}_\ell + \frac{3}{4} \times \frac{1 - 2\widehat{\mathbf{p}}_\ell - \sqrt{1 + (9n_\ell/2\ln(1/\zeta\delta))\widehat{\mathbf{p}}_\ell(1-\widehat{\mathbf{p}}_\ell)}}{1 + (9n_\ell/8\ln(1/\zeta\delta))}\right\},$$

$$U_\ell = \min\left\{1, \widehat{\mathbf{p}}_\ell + \frac{3}{4} \times \frac{1 - 2\widehat{\mathbf{p}}_\ell + \sqrt{1 + (9n_\ell/2\ln(1/\zeta\delta))\widehat{\mathbf{p}}_\ell(1-\widehat{\mathbf{p}}_\ell)}}{1 + (9n_\ell/8\ln(1/\zeta\delta))}\right\} \tag{20}$$

and that $\Pr\{L_\ell \le p \le U_\ell \mid p\} \ge 1 - 2\zeta\delta$ for $\ell = 1, \ldots, s$ and $p \in (0, 1)$. Under the assumption that $0 < \zeta\delta < 1$ and $0 < \epsilon < 1/2$, by similar techniques as the proof of Theorem 1 of [30], it can be shown that $\{\widehat{\mathbf{p}}_\ell - \epsilon \le L_\ell \le U_\ell \le \widehat{\mathbf{p}}_\ell + \epsilon\} = \{(|\widehat{\mathbf{p}}_\ell - 1/2| - (2/3)\epsilon)^2 \ge (1/4) + (\epsilon^2 n_\ell/2\ln(\zeta\delta))\}$ for $\ell = 1, \ldots, s$. This implies that applying the sequence of confidence intervals of Chen et al. to the general stopping rule (12) leads to Stopping Rule B.

Actually, the confidence intervals of Chen et al. [29] are derived from Massart's inequality [31] on the tailed probabilities of the sample mean of Bernoulli random variable. For this reason, Stopping Rule B is also referred to as the stopping rule from Massart's inequality in [21, Section 4.1.1].

4. Double-Parabolic Sequential Estimation

From Sections 2.2, 3.2, and 3.7, it can be seen that, by introducing a new parameter $\rho \in [0, 1]$ and letting ρ take values 2/3 and 0, respectively, Stopping Rules B and D can be accommodated as special cases of the following general stopping rule.

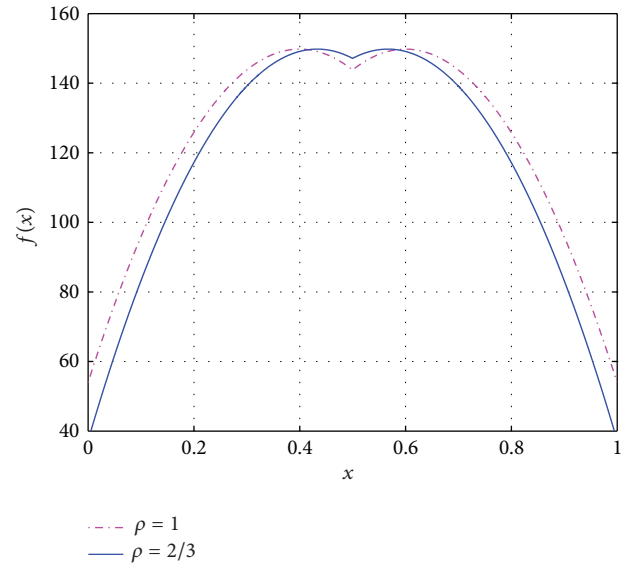


Figure 1: Double-parabolic sampling.

Continue the sampling process until

$$\left(\left|\widehat{\mathbf{p}}_\ell - \frac{1}{2}\right| - \rho\epsilon\right)^2 \ge \frac{1}{4} + \frac{\epsilon^2 n_\ell}{2\ln(\zeta\delta)} \tag{21}$$

for some $\ell \in \{1, 2, \ldots, s\}$, where $\zeta \in (0, 1/\delta)$.

Moreover, as can be seen from (16), the stopping rule derived from applying Wilson's confidence intervals to (12) can also be viewed as a special case of such general stopping rule with $\rho = 1$.

From the stopping condition (21), it can be seen that the stopping boundary is associated with the double-parabolic function $f(x) = (2/\epsilon^2)\ln(\zeta\delta)[(1/4) - (|x - 1/2| - \rho\epsilon)^2]$ such that x and $f(x)$ correspond to the sample mean and sample size, respectively. For $\epsilon = 0.1$, $\delta = 0.05$, and $\zeta = 1$, stopping boundaries with various ρ are shown by Figure 1.

For fixed ϵ and δ, the parameters ρ and ζ affect the shape of the stoping boundary in a way as follows. As ρ increases, the span of stopping boundary is increasing in the axis of sample mean. By decreasing ζ, the stopping boundary can be dragged toward the direction of increasing sample size. Hence, the parameter ρ is referred to as the *dilation coefficient*. The parameter ζ is referred to as the *coverage tuning parameter*. Since the stopping boundary consists of two parabolas, this approach of estimating a binomial proportion is referred to as the *double-parabolic sequential estimation* method.

4.1. Parametrization of the Sampling Scheme. In this section, we shall parameterize the double-parabolic sequential sampling scheme by the method described in Section 2.2. From the stopping condition (21), the stopping rule can be restated as follows. Continue sampling until $\mathscr{D}(\widehat{\mathbf{p}}_\ell, n_\ell) = 1$

for some $\ell \in \{1, \ldots, s\}$, where the function $\mathscr{D}(\cdot,\cdot)$ is defined by

$$\mathscr{D}(z,n) = \begin{cases} 1 & \text{if } \left(\left|z-\frac{1}{2}\right| - \rho\epsilon\right)^2 \geq \frac{1}{4} + \frac{\epsilon^2 n}{2\ln(\zeta\delta)}, \\ 0 & \text{otherwise.} \end{cases} \quad (22)$$

Clearly, the function $\mathscr{D}(\cdot,\cdot)$ associated with the double-parabolic sequential sampling scheme depends on the design parameters ρ, ζ, ϵ and δ. Applying the function $\mathscr{D}(\cdot,\cdot)$ defined by (22) to (6) yields

$$N_{\min} = \min\left\{ n \in \mathbb{N} : \left(\left|\frac{k}{n}-\frac{1}{2}\right| - \rho\epsilon\right)^2 \geq \frac{1}{4} + \frac{\epsilon^2 n}{2\ln(\zeta\delta)} \text{ for some nonnegative integer } k \text{ not exceeding } n \right\}. \quad (23)$$

Since ϵ is usually small in practical applications, we restrict ϵ to satisfy $0 < \rho\epsilon \leq 1/4$. As a consequence of $0 \leq \rho\epsilon \leq 1/4$ and the fact that $|z - 1/2| \leq 1/2$ for any $z \in [0,1]$, it must be true that $(|z-1/2| - \rho\epsilon)^2 \leq ((1/2) - \rho\epsilon)^2$ for any $z \in [0,1]$. It follows from (23) that $((1/2) - \rho\epsilon)^2 \geq (1/4) + (\epsilon^2 N_{\min}/2\ln(\zeta\delta))$, which implies that the minimum sample size can be taken as

$$N_{\min} = \left\lceil 2\rho\left(\frac{1}{\epsilon} - \rho\right)\ln\frac{1}{\zeta\delta} \right\rceil. \quad (24)$$

On the other hand, applying the function $\mathscr{D}(\cdot,\cdot)$ defined by (22) to (7) gives

$$N_{\max} = \min\left\{ n \in \mathbb{N} : \left(\left|\frac{k}{n}-\frac{1}{2}\right| - \rho\epsilon\right)^2 \geq \frac{1}{4} + \frac{\epsilon^2 n}{2\ln(\zeta\delta)} \text{ for all nonnegative integer } k \text{ not exceeding } n \right\}. \quad (25)$$

Since $(|z-1/2| - \rho\epsilon)^2 \geq 0$ for any $z \in [0,1]$, it follows from (25) that $(1/4) + (\epsilon^2 N_{\max}/2\ln(\zeta\delta)) \leq 0$, which implies that maximum sample size can be taken as

$$N_{\max} = \left\lceil \frac{1}{2\epsilon^2}\ln\frac{1}{\zeta\delta} \right\rceil. \quad (26)$$

Therefore, the sample sizes $n_1, \ldots, n_s$ can be chosen as functions of ρ, ζ, ϵ, and δ which satisfy the following constraint:

$$N_{\min} \leq n_1 < \cdots < n_{s-1} < N_{\max} \leq n_s. \quad (27)$$

In particular, if the number of stages s is given and the group sizes are expected to be approximately equal, then the sample sizes, $n_1, \ldots, n_s$, for all stages can be obtained by substituting $N_{\min}$ defined by (24) and $N_{\max}$ defined by (26) into (8). For example, if the values of design parameters are $\epsilon = 0.05$, $\delta = 0.05$, $\rho = 3/4$, $\zeta = 2.6759$ and $s = 7$, then the sample sizes of this sampling scheme are calculated as

$$n_1 = 59, \quad n_2 = 116, \quad n_3 = 173, \quad n_4 = 231, \quad n_5 = 288, \quad n_6 = 345, \quad n_7 = 403. \quad (28)$$

The stopping rule is completely determined by substituting the values of design parameters into (21).

4.2. Uniform Controllability of Coverage Probability. Clearly, for prespecified ϵ, δ, and ρ, the coverage probability $\Pr\{|\widehat{\mathbf{p}} - p| < \epsilon \mid p\}$ depends on the parameter ζ, the number of stages s, and the sample sizes $n_1, \ldots, n_s$. As illustrated in Section 4.1, the number of stages s and the sample sizes $n_1, \ldots, n_s$ can be defined as functions of $\zeta \in (0, 1/\delta)$. That is, the stopping rule can be parameterized by ζ. Accordingly, for any $p \in (0,1)$, the coverage probability $\Pr\{|\widehat{\mathbf{p}} - p| < \epsilon \mid p\}$ becomes a function of ζ. The following theorem shows that it suffices to choose $\zeta \in (0, 1/\delta)$ small enough to guarantee the prespecified confidence level.

Theorem 2. *Let $\epsilon, \delta \in (0,1)$ and $\rho \in (0,1]$ be fixed. Assume that the number of stages s and the sample sizes $n_1, \ldots, n_s$ are functions of $\zeta \in (0, 1/\delta)$ such that the constraint* (27) *is satisfied. Then,* $\Pr\{|\widehat{\mathbf{p}} - p| < \epsilon \mid p\}$ *is no less than* $1 - \delta$ *for any* $p \in (0,1)$ *provided that*

$$0 < \zeta \leq \frac{1}{\delta}\exp\left(\frac{\ln(\delta/2) + \ln\left[1 - \exp\left(-2\epsilon^2\right)\right]}{4\epsilon\rho(1-\rho\epsilon)}\right). \quad (29)$$

See Appendix B for a proof. For Theorem 2 to be valid, the choice of sample sizes is very flexible. Particularly, the sample sizes can be arithmetic or geometric progressions or any others, as long as the constraint (27) is satisfied. It can be seen that for the coverage probability to be uniformly controllable, the dilation coefficient ρ must be greater than 0. Theorem 2 asserts that there exists $\zeta > 0$ such that the coverage probability is no less than $1 - \delta$, regardless of the associated binomial proportion p. For the purpose of reducing sampling cost, we want to have a value of ζ as large as possible such that the prespecified confidence level is guaranteed for any $p \in (0,1)$. This can be accomplished by the technical components introduced in Sections 2.1, 2.3, 2.4, and 2.5. Clearly, for every value of ρ, we can obtain a corresponding value of ζ (as large as possible) to ensure the desired confidence level. However, the performance of resultant stopping rules are different. Therefore, we can try a number of values of ρ and pick the best resultant stopping rule for practical use.

4.3. Asymptotic Optimality of Sampling Schemes. Now we shall provide an important reason why we propose the sampling scheme of that structure by showing its asymptotic optimality. Since the performance of a group sampling scheme will be close to its fully sequential counterpart, we investigate the optimality of the fully sequential sampling

scheme. In this scenario, the sample sizes $n_1, n_2, \ldots, n_s$ are consecutive integers such that

$$\left\lceil 2\rho\left(\frac{1}{\epsilon}-\rho\right)\ln\frac{1}{\zeta\delta}\right\rceil = n_1 < n_2 < \cdots < n_{s-1} < n_s = \left\lceil \frac{1}{2\epsilon^2}\ln\frac{1}{\zeta\delta}\right\rceil. \tag{30}$$

The fully sequential sampling scheme can be viewed as a special case of a group sampling scheme of $s = n_s - n_1 + 1$ stages and group size 1. Clearly, if δ, ζ and ρ are fixed, the sampling scheme is dependent only on ϵ. Hence, for any $p \in (0,1)$, if we allow ϵ to vary in $(0,1)$, then the coverage probability $\Pr\{|\widehat{\mathbf{p}} - p| < \epsilon \mid p\}$ and the average sample number $\mathbb{E}[\mathbf{n}]$ are functions of ϵ. We are interested in knowing the asymptotic behavior of these functions as $\epsilon \to 0$, since ϵ is usually small in practical situations. The following theorem provides us the desired insights.

Theorem 3. *Assume that $\delta \in (0,1)$, $\zeta \in (0, 1/\delta)$ and $\rho \in (0,1]$ are fixed. Define $N(p,\epsilon,\delta,\zeta) = (2p(1-p)\ln(1/\zeta\delta))/\epsilon^2$ for $p \in (0,1)$ and $\epsilon \in (0,1)$. Then,*

$$\Pr\left\{\lim_{\epsilon\to 0}\frac{\mathbf{n}}{N(p,\epsilon,\delta,\zeta)} = 1 \mid p\right\} = 1, \tag{31}$$

$$\lim_{\epsilon\to 0}\Pr\{|\widehat{\mathbf{p}} - p| < \epsilon \mid p\} = 2\Phi\left(\sqrt{2\ln\frac{1}{\zeta\delta}}\right) - 1, \tag{32}$$

$$\lim_{\epsilon\to 0}\frac{\mathbb{E}[\mathbf{n}]}{N(p,\epsilon,\delta,\zeta)} = 1, \tag{33}$$

for any $p \in (0,1)$.

See Appendix C for a proof. From (32), it can be seen that $\lim_{\epsilon\to 0}\Pr\{|\widehat{\mathbf{p}} - p| < \epsilon \mid p\} = 1 - \delta$ for any $p \in (0,1)$ if $\zeta = (1/\delta)\exp(-(1/2)\mathscr{Z}^2_{\delta/2})$. Such value can be taken as an initial value for the coverage tuning parameter ζ. In addition to providing guidance on the coverage tuning techniques, Theorem 3 also establishes the optimality of the sampling scheme. To see this, let $\mathscr{N}(p,\epsilon,\delta)$ denote the minimum sample size n required for a fixed-sample-size procedure to guarantee that $\Pr\{|\overline{X}_n - p| < \epsilon \mid p\} \geq 1 - \delta$ for any $p \in (0,1)$, where $\overline{X}_n = (\sum_{i=1}^n X_i)/n$. It is well known that from the central limit theorem,

$$\lim_{\epsilon\to 0}\frac{\mathscr{N}(p,\epsilon,\delta)}{p(1-p)\left(\mathscr{Z}_{\delta/2}/\epsilon\right)^2} = 1. \tag{34}$$

Applying (33), (34), and letting $\zeta = (1/\delta)\exp(-(1/2)\mathscr{Z}^2_{\delta/2})$, we have $\lim_{\epsilon\to 0}(\mathscr{N}(p,\epsilon,\delta)/N(p,\epsilon,\delta,\zeta)) = 1$ for $p \in (0,1)$ and $\delta \in (0,1)$, which implies the asymptotic optimality of the double-parabolic sampling scheme. By virtue of (33), an approximate formula for computing the average sample number is given as follows:

$$\mathbb{E}[\mathbf{n}] \approx N(p,\epsilon,\delta,\zeta) = \frac{2p(1-p)\ln(1/\zeta\delta)}{\epsilon^2}, \tag{35}$$

for $p \in (0,1)$ and $\epsilon \in (0,1)$. From (34), one obtains $\mathscr{N}(p,\epsilon,\delta) \approx p(1-p)(\mathscr{Z}_{\delta/2}/\epsilon)^2$, which is a well-known result in statistics. In situations that no information of p is available, one usually uses

$$N_{\text{normal}} \stackrel{\text{def}}{=} \left\lceil \frac{1}{4}\left(\frac{\mathscr{Z}_{\delta/2}}{\epsilon}\right)^2\right\rceil \tag{36}$$

as the sample size for estimating the binomial proportion p with prescribed margin of error ϵ and confidence level $1-\delta$. Since the sample size formula (36) can lead to under-coverage, researchers in many areas are willing to use a more conservative but rigorous sample size formula

$$N_{\text{ch}} \stackrel{\text{def}}{=} \left\lceil \frac{\ln(2/\delta)}{2\epsilon^2}\right\rceil, \tag{37}$$

which is derived from the Chernoff-Hoeffding bound [32, 33]. Comparing (35) and (37), one can see that under the premise of guaranteeing the prescribed confidence level $1-\delta$, the double-parabolic sampling scheme can lead to a substantial reduction of sample number when the unknown binomial proportion p is close to 0 or 1.

4.4. Bounds on Distribution and Expectation of Sample Number. We shall derive analytic bounds for the cumulative distribution function and expectation of the sample number $\mathbf{n}$ associated with the double-parabolic sampling scheme. In this direction, we have obtained the following results.

Theorem 4. *Let $p \in (0, 1/2]$. Define $a_\ell = (1/2) - \rho\epsilon - \sqrt{(1/4) + (\epsilon^2 n_\ell/2\ln(\zeta\delta))}$ for $\ell = 1, \ldots, s$. Let τ denote the index of stage such that $a_{\tau-1} \leq p < a_\tau$. Then, $\Pr\{\mathbf{n} > n_\ell \mid p\} \leq \exp(n_\ell \mathscr{M}(a_\ell, p))$ for $\tau \leq \ell < s$. Moreover, $\mathbb{E}[\mathbf{n}] \leq n_\tau + \sum_{\ell=\tau}^{s-1}(n_{\ell+1} - n_\ell)\exp(n_\ell \mathscr{M}(a_\ell, p))$.*

See Appendix D for a proof. By the symmetry of the double-parabolic sampling scheme, similar analytic bounds for the distribution and expectation of the sample number can be derived for the case that $p \in [1/2, 1)$.

5. Comparison of Computational Methods

In this section, we shall compare various computational methods. First, we will illustrate why a frequently used method of evaluating the coverage probability based on gridding the parameter space is not rigorous and is less efficient as compared to the adapted B&B algorithm. Second, we will introduce the Adaptive Maximum Checking Algorithm of [21] which has better computational efficiency as compared to the adapted B&B algorithm. Third, we will explain that it is more advantageous in terms of numerical accuracy to work with the complementary coverage probability as compared to direct evaluation of the coverage probability. Finally, we will compare the computational methods of Chen [14–18] and Frey [13] for the design of sequential procedures for estimating a binomial proportion.

∇ Choose initial step size $d > \eta$.
∇ Let $F \leftarrow 0$, $T \leftarrow 0$ and $b \leftarrow \overline{\theta}$.
∇ While $F = T = 0$, do the following:
  ⋄ Let st $\leftarrow 0$ and $\ell \leftarrow 2$;
  ⋄ While st = 0, do the following:
    ⋆ Let $\ell \leftarrow \ell - 1$ and $d \leftarrow d2^{\ell}$.
    ⋆ If $b - d > \underline{\theta}$, then let $a \leftarrow b - d$ and $T \leftarrow 0$.
      Otherwise, let $a \leftarrow \underline{\theta}$ and $T \leftarrow 1$.
    ⋆ If $\overline{C}(a, b) < \delta$, then let st $\leftarrow 1$ and $b \leftarrow a$.
    ⋆ If $d < \eta$, then let st $\leftarrow 1$ and $F \leftarrow 1$.
∇ Return F.

Algorithm 2

5.1. Verifying Coverage Guarantee without Gridding Parameter Space. For purpose of constructing a sampling scheme so that the prescribed confidence level $1 - \delta$ is guaranteed, an essential task is to determine whether the coverage probability $\Pr\{|\widehat{\mathbf{p}} - p| < \epsilon \mid p\}$ associated with a given stopping rule is no less than $1 - \delta$. In other words, it is necessary to compare the infimum of coverage probability with $1 - \delta$. To accomplish such a task of checking coverage guarantee, a natural method is to evaluate the infimum of coverage probability as follows:

(i) choose m grid points $p_1, \ldots, p_m$ from parameter space $(0, 1)$;

(ii) compute $c_j = \Pr\{|\widehat{\mathbf{p}} - p| < \epsilon \mid p_j\}$ for $j = 1, \ldots, m$;

(iii) Take $\min\{c_1, \ldots, c_m\}$ as $\inf_{p\in(0,1)}\Pr\{|\widehat{\mathbf{p}} - p| < \epsilon \mid p\}$.

This method can be easily mistaken as an exact approach and has been frequently used for evaluating coverage probabilities in many problem areas.

It is not hard to show that if the sample size $\mathbf{n}$ of a sequential procedure has a support $\mathcal{S}$, then the coverage probability $\Pr\{|\widehat{\mathbf{p}} - p| < \epsilon \mid p\}$ is discontinuous at $p \in \mathcal{P} \cap (0, 1)$, where $\mathcal{P} = \{(k/n) \pm \epsilon : k$ is a nonnegative integer no greater than $n \in \mathcal{S}\}$. The set $\mathcal{P}$ typically has a large number of parameter values. Due to the discontinuity of the coverage probability as a function of p, the coverage probabilities can differ significantly for two parameter values which are extremely close. This implies that an intolerable error can be introduced by taking the minimum of coverage probabilities of a finite number of parameter values as the infimum of coverage probability on the whole parameter space. So, if one simply uses the minimum of the coverage probabilities of a finite number of parameter values as the infimum of coverage probability to check the coverage guarantee, the sequential estimator $\widehat{\mathbf{p}}$ of the resultant stopping rule will fail to guarantee the prescribed confidence level.

In addition to the lack of rigorousness, another drawback of checking coverage guarantee based on the method of gridding parameter space is its low efficiency. A critical issue is on the choice of the number, m, of grid points. If the number m is too small, the induced error can be substantial. On the other hand, choosing a large number for m results in high computational complexity.

In contrast to the method based on gridding parameter space, the adapted B&B algorithm is a rigorous approach for checking coverage guarantee as a consequence of the mechanism for comparing the bounds of coverage probability with the prescribed confidence level. The algorithm is also efficient due to the mechanism of pruning branches.

5.2. Adaptive Maximum Checking Algorithm. As illustrated in Section 2, the techniques developed in [14–18] are sufficient to provide exact solutions for a wide range of sequential estimation problems. However, one of the four components, the adapted B&B algorithm, requires computing both the lower and upper bounds of the complementary coverage probability. To further reduce the computational complexity, it is desirable to have a checking algorithm which needs only one of the lower and upper bounds. For this purpose, Chen had developed the Adaptive Maximum Checking Algorithm (AMCA) in [21, Section 3.3] and [19, Section 2.7]. In the following introduction of the AMCA, we shall follow the description of [21]. The AMCA can be applied to a wide class of computational problems dependent on the following critical subroutine.

Determine whether a function $C(\theta)$ is smaller than a prescribed number δ for every value of θ contained in interval $[\underline{\theta}, \overline{\theta}]$.

Particularly, for checking the coverage guarantee in the context of estimating a binomial proportion, the parameter θ is the binomial proportion p and the function $C(\theta)$ is actually the complementary coverage probability. In many situations, it is impossible or very difficult to evaluate $C(\theta)$ for every value of θ in interval $[\underline{\theta}, \overline{\theta}]$, since the interval may contain infinitely many or an extremely large number of values. Similar to the adapted B&B algorithm, the purpose of AMCA is to reduce the computational complexity associated with the problem of determining whether the maximum of $C(\theta)$ over $[\underline{\theta}, \overline{\theta}]$ is less than δ. The only assumption required for AMCA is that, for any interval $[a, b] \subseteq [\underline{\theta}, \overline{\theta}]$, it is possible to compute an upper bound $\overline{C}(a, b)$ such that $C(\theta) \leq \overline{C}(a, b)$ for any $\theta \in [a, b]$ and that the upper bound converges to $C(\theta)$ as the interval width $b - a$ tends to 0. The backward AMCA proceeds as in Algorithm 2.

The output of the backward AMCA is a binary variable F such that "$F = 0$" means "$C(\theta) < \delta$" and "$F = 1$" means "$C(\theta) \geq \delta$." An intermediate variable T is introduced in the description of AMCA such that "$T = 1$" means that the left endpoint of the interval is reached. The backward AMCA starts from the right endpoint of the interval (i.e., $b = \overline{\theta}$) and attempts to find an interval $[a, b]$ such that $\overline{C}(a, b) < \delta$. If such an interval is available, then, attempt to go backward to find the next consecutive interval with twice width. If doubling the interval width fails to guarantee $\overline{C}(a, b) < \delta$, then try to repeatedly cut the interval width in half to ensure that $\overline{C}(a, b) < \delta$. If the interval width becomes smaller than a prescribed tolerance η, then AMCA declares that "$F = 1$." For our relevant statistical problems, if $C(\theta) \geq \delta$ for some $\theta \in [\underline{\theta}, \overline{\theta}]$, it is sure that "$F = 1$" will be declared. On the other hand, it is possible that "$F = 1$" is declared even though $C(\theta) < \delta$ for any $\theta \in [\underline{\theta}, \overline{\theta}]$. However, such situation can be made extremely rare and immaterial if we choose η to be a very small number. Moreover, this will only introduce negligible conservativeness in the evaluation of $C(\theta)$ if η is chosen to be sufficiently small (e.g., $\eta = 10^{-15}$). Clearly, the backward AMCA can be easily modified as forward AMCA. Moreover, the AMCA can also be easily modified as Adaptive Minimum Checking Algorithm (forward and backward). For checking the maximum of complementary coverage probability $\Pr\{|\hat{\mathbf{p}} - p| \geq \epsilon \mid p\}$, one can use the AMCA with $C(p) = \Pr\{|\hat{\mathbf{p}} - p| \geq \epsilon \mid p\}$ over interval $[0, 1/2]$. We would like to point out that, in contrast to the adapted B&B algorithm, it seems difficult to generalize the AMCA to problems involving multidimensional parameter spaces.

5.3. Working with Complementary Coverage Probability. We would like to point out that, instead of evaluating the coverage probability as in [13], it is better to evaluate the complementary coverage probability for purpose of reducing numerical error. The advantage of working on the complementary coverage probability can be explained as follows. Note that, in many cases, the coverage probability is very close to 1 and the complementary coverage probability is very close to 0. Since the absolute precision for computing a number close to 1 is much lower than the absolute precision for computing a number close to 0, the method of directly evaluating the coverage probability will lead to intolerable numerical error for problems involving small δ. As an example, consider a situation that the complementary coverage probability is in the order of 10^{-5}. Direct computation of the coverage probability can easily lead to an absolute error of the order of 10^{-5}. However, the absolute error of computing the complementary coverage probability can be readily controlled at the order of 10^{-9}.

5.4. Comparison of Approaches of Chen and Frey. As mentioned in the introduction, Frey published a paper [13] in The American Statistician (TAS) on the sequential estimation of a binomial proportion with prescribed margin of error and confidence level. The approaches of Chen and Frey are based on the same strategy as follows. First, construct a family of stopping rules parameterized by γ (and possibly other design parameters) so that the associated coverage probability $\Pr\{|\hat{\mathbf{p}} - p| < \epsilon \mid p\}$ can be controlled by parameter γ in the sense that the coverage probability can be made arbitrarily close to 1 by increasing γ. Second, apply a bisection search method to determine the parameter γ so that the coverage probability is no less than the prescribed confidence level $1 - \delta$ for any $p \in (0, 1)$.

For the purpose of controlling the coverage probability, Frey [13] applied the inclusion principle previously proposed in [18, Section 3] and used in [14–17]. As illustrated in Section 3, the central idea of inclusion principle is to use a sequence of confidence intervals to construct stopping rules so that the sampling process is continued until a confidence interval is included by an interval defined in terms of the estimator and margin of error. Due to the inclusion relationship, the associated coverage probability can be controlled by the confidence coefficients of the sequence of confidence intervals. The critical value γ used by Frey plays the same role for controlling coverage probabilities as that of the coverage tuning parameter ζ used by Chen. Frey [13] stated stopping rules in terms of confidence limits. This way of expressing stopping rules is straightforward and insightful, since one can readily see the principle behind the construction. For convenience of practical use, Chen proposed to eliminate the necessity of computing confidence limits.

Similar to the AMCA proposed in [21, Section 3.3], the algorithm of Frey [13, Appendix] for checking coverage guarantee adaptively scans the parameter space based on interval bounding. The adaptive method used by Frey for updating step size is essentially the same as that of the AMCA. Ignoring the number 0.01 in Frey's expression "$\varepsilon_i = \min\{0.01, 2(p_{i-1} - p_{i-2})\}$," which has very little impact on the computational efficiency, Frey's step size ε_i can be identified as the adaptive step size d in the AMCA. The operation associated with "$\varepsilon_i = \min\{0.01, 2(p_{i-1} - p_{i-2})\}$" has a similar function as that of the command "Let st $\leftarrow 0$ and $\ell \leftarrow 2$" in the outer loop of the AMCA. The operation associated with Frey's expression "$p_{i-1} + \varepsilon_i/2^j,\ j \geq 0$" is equivalent to that of the command "Let $\ell \leftarrow \ell - 1$ and $d \leftarrow d2^{\ell}$" in the inner loop of the AMCA. Frey proposed to declare a failure of coverage guarantee if "the distance from p_{i-1} to the candidate value for p_i falls below 10^{-14}." The number "10^{-14}" actually plays the same role as "η" in the AMCA, where "$\eta = 10^{-15}$" is recommended by [21].

6. Numerical Results

In this section, we shall illustrate the proposed double-parabolic sampling scheme through examples. As demonstrated in Sections 2.2 and 4, the double-parabolic sampling scheme can be parameterized by the dilation coefficient ρ and the coverage tuning parameter ζ. Hence, the performance of the resultant stopping rule can be optimized with respect to $\rho \in (0, 1]$ and ζ by choosing various values of ρ from interval $(0, 1]$ and determining the corresponding values of

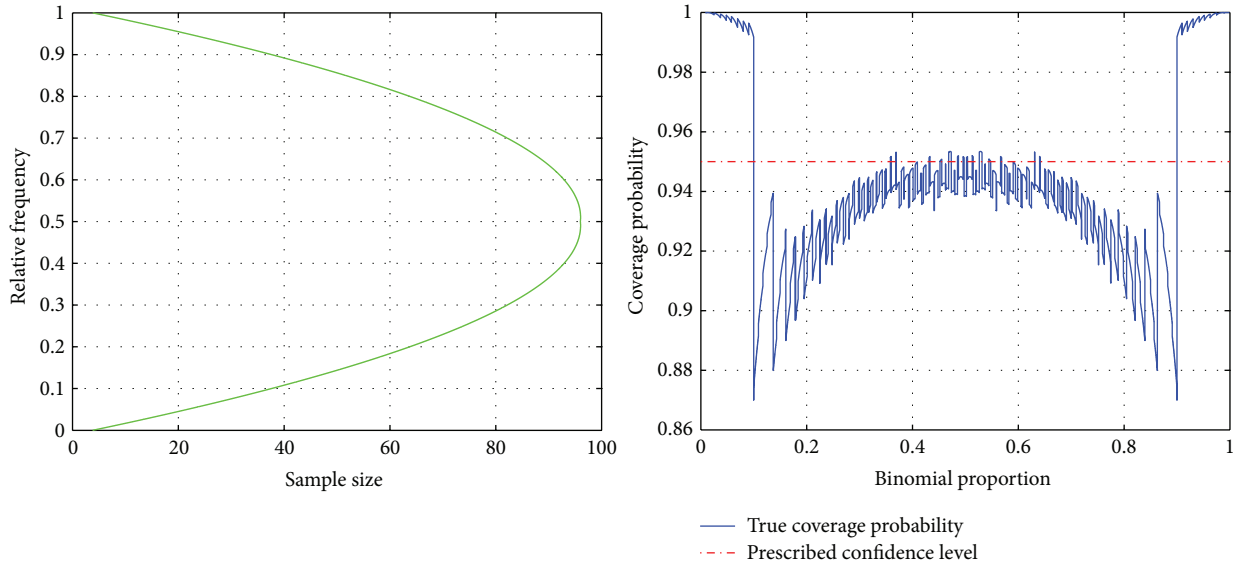


Figure 2: Double-parabolic sampling with $\epsilon = 0.1$, $\delta = 0.05$, $\rho = 1/10$, and $\zeta = 2.93$.

Table 1: Coverage tuning parameter.

ϵ	δ	ζ	ϵ	δ	ζ	ϵ	δ	ζ
0.1	0.1	2.0427	0.1	0.05	2.4174	0.1	0.01	3.0608
0.05	0.1	2.0503	0.05	0.05	2.5862	0.05	0.01	3.3125
0.02	0.1	2.1725	0.02	0.05	2.5592	0.02	0.01	3.4461
0.01	0.1	2.1725	0.01	0.05	2.5592	0.01	0.01	3.4461

ζ by the computational techniques introduced in Section 2 to guarantee the desired confidence interval.

6.1. Asymptotic Analysis May Be Inadequate. For fully sequential cases, we have evaluated the double-parabolic sampling scheme with $\epsilon = 0.1$, $\delta = 0.05$, $\rho = 0.1$, and $\zeta = (1/\delta)\exp(-(1/2)\mathcal{Z}^2_{\delta/2}) \approx 2.93$. The stopping boundary is displayed in the left side of Figure 2. The function of coverage probability with respect to the binomial proportion is shown in the right side of Figure 2, which indicates that the coverage probabilities are generally substantially lower than the prescribed confidence level $1 - \delta = 0.05$. By considering $\epsilon = 0.1$ as a small number and applying the asymptotic theory, the coverage probability associated with the sampling scheme is expected to be close to 0.95. This numerical example demonstrates that although the asymptotic method is insightful and involves virtually no computation, it may not be adequate.

In general, the main drawback of an asymptotic method is that there is no guarantee of coverage probability. Although an asymptotical method asserts that if the margin of error ϵ tends to 0, the coverage probability will tend to the prespecified confidence level $1 - \delta$, it is difficult to determine how small the margin of error ϵ is sufficient for the asymptotic method to be applicable. Note that $\epsilon \to 0$ implies the average sample size tends to ∞. However, in reality, the sample sizes must be finite. Consequently, an asymptotic method inevitably introduces unknown statistical error. Since an asymptotic method does not necessarily guarantee the prescribed confidence level, it is not fair to compare its associated sample size with that of an exact method, which guarantees the prespecified confidence level.

This example also indicates that, due to the discrete nature of the problem, the coverage probability is a discontinuous and erratic function of p, which implies that Monte Carlo simulation is not suitable for evaluating the coverage performance.

6.2. Parametric Values of Fully Sequential Schemes. For fully sequential cases, to allow direct application of our double-parabolic sequential method, we have obtained values of coverage tuning parameter ζ, which guarantee the prescribed confidence levels, for double-parabolic sampling schemes with $\rho = 3/4$ and various combinations of (ϵ, δ) as shown in Table 1. We used the computational techniques introduced in Section 2 to obtain this table.

To illustrate the use of Table 1, suppose that one wants a fully sequential sampling procedure to ensure that $\Pr\{|\widehat{\mathbf{p}} - p| < 0.1 \mid p\} > 0.95$ for any $p \in (0, 1)$. This means that one can choose $\epsilon = 0.1$, $\delta = 0.05$ and the range of sample size is given by (30). From Table 1, it can be seen that the value of ζ corresponding to $\epsilon = 0.1$ and $\delta = 0.05$ is 2.4174. Consequently, the stopping rule is completely determined by substituting the values of design parameters $\epsilon = 0.1$,

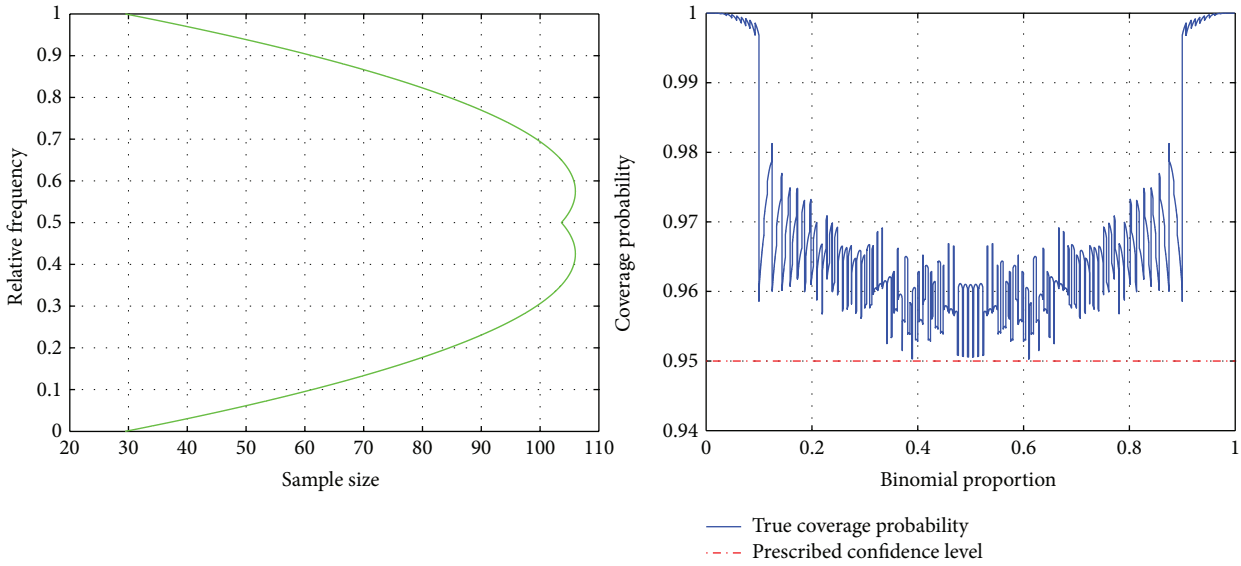


Figure 3: Double-parabolic sampling with $\epsilon = 0.1$, $\delta = 0.05$, $\rho = 3/4$, and $\zeta = 2.4174$.

Table 2: Coverage tuning parameter.

	$s = 3$	$s = 4$	$s = 5$	$s = 6$	$s = 7$	$s = 8$	$s = 9$	$s = 10$
$\epsilon = 0.1$	2.6583	2.6583	2.5096	2.5946	2.4459	2.6512	2.5096	2.4459
$\epsilon = 0.05$	2.6759	2.6759	2.6759	2.6759	2.6759	2.6759	2.6759	2.6759
$\epsilon = 0.02$	2.6725	2.6725	2.6725	2.6725	2.6725	2.6725	2.6725	2.6725
$\epsilon = 0.01$	2.6796	2.6796	2.6796	2.6796	2.6796	2.5875	2.6796	2.6796

$\delta = 0.05$, $\rho = 3/4$, and $\zeta = 2.4174$ into its definition. The stopping boundary of this sampling scheme is displayed in the left side of Figure 3. The function of coverage probability with respect to the binomial proportion is shown in the right side of Figure 3.

6.3. Parametric Values of Group Sequential Schemes. In many situations, especially in clinical trials, it is desirable to use group sequential sampling schemes. In Tables 2 and 3, assuming that sample sizes satisfy (8) for the purpose of having approximately equal group sizes, we have obtained parameters for concrete schemes by the computational techniques introduced in Section 2.

For dilation coefficient $\rho = 3/4$ and confidence parameter $\delta = 0.05$, we have obtained values of coverage tuning parameter ζ, which guarantee the prescribed confidence level 0.95, for double-parabolic sampling schemes, with the number of stages s ranging from 3 to 10, as shown in Table 2.

For dilation coefficient $\rho = 3/4$ and confidence parameter $\delta = 0.01$, we have obtained values of coverage tuning parameter ζ, which guarantee the prescribed confidence level 0.99, for double-parabolic sampling schemes, with the number of stages s ranging from 3 to 10, as shown in Table 3.

To illustrate the use of these tables, suppose that one wants a ten-stage sampling procedure of approximately equal group sizes to ensure that $\Pr\{|\widehat{\mathbf{p}} - p| < 0.01 \mid p\} > 0.99$ for any $p \in (0, 1)$. This means that one can choose $\epsilon = \delta = 0.01$, $s = 10$ and sample sizes satisfying (8). To obtain appropriate parameter values for the sampling procedure, one can look at Table 3 to find the coverage tuning parameter ζ corresponding to $\epsilon = 0.01$ and $s = 10$. From Table 3, it can be seen that ζ can be taken as 3.5753. Consequently, the stopping rule is completely determined by substituting the values of design parameters $\epsilon = 0.01$, $\delta = 0.01$, $\rho = 3/4$, $\zeta = 3.5753$, and $s = 10$ into its definition and (8). The stopping boundary of this sampling scheme and the function of coverage probability with respect to the binomial proportion are displayed, respectively, in the left and right sides of Figure 4.

6.4. Comparison of Sampling Schemes. We have conducted numerical experiments to investigate the impact of dilation coefficient ρ on the performance of our double-parabolic sampling schemes. Our computational experiences indicate that the dilation coefficient $\rho = 3/4$ is frequently a good choice in terms of average sample number and coverage probability. For example, consider the case that the margin of error is given as $\epsilon = 0.1$ and the prescribed confidence level is $1 - \delta$ with $\delta = 0.05$. For the double-parabolic sampling scheme with the dilation coefficient ρ chosen as $2/3, 3/4$, and 1, we have determined that, to ensure the prescribed confidence level $1 - \delta = 0.95$, it suffices to set the coverage tuning parameter ζ as 2.1, 2.4 and 2.4, respectively. The average sample numbers of these sampling schemes and the coverage probabilities as functions of the binomial proportion are shown, respectively, in the left and right sides

Table 3: Coverage tuning parameter.

	$s = 3$	$s = 4$	$s = 5$	$s = 6$	$s = 7$	$s = 8$	$s = 9$	$s = 10$
$\epsilon = 0.1$	3.3322	3.3322	3.3322	3.3322	3.3322	3.2709	3.0782	3.3322
$\epsilon = 0.05$	3.5074	3.5074	3.5074	3.5074	3.5074	3.5074	3.5074	3.5074
$\epsilon = 0.02$	3.5430	3.5430	3.5430	3.5430	3.5430	3.5430	3.5430	3.5430
$\epsilon = 0.01$	3.5753	3.5753	3.5753	3.5753	3.5753	3.5753	3.5753	3.5753

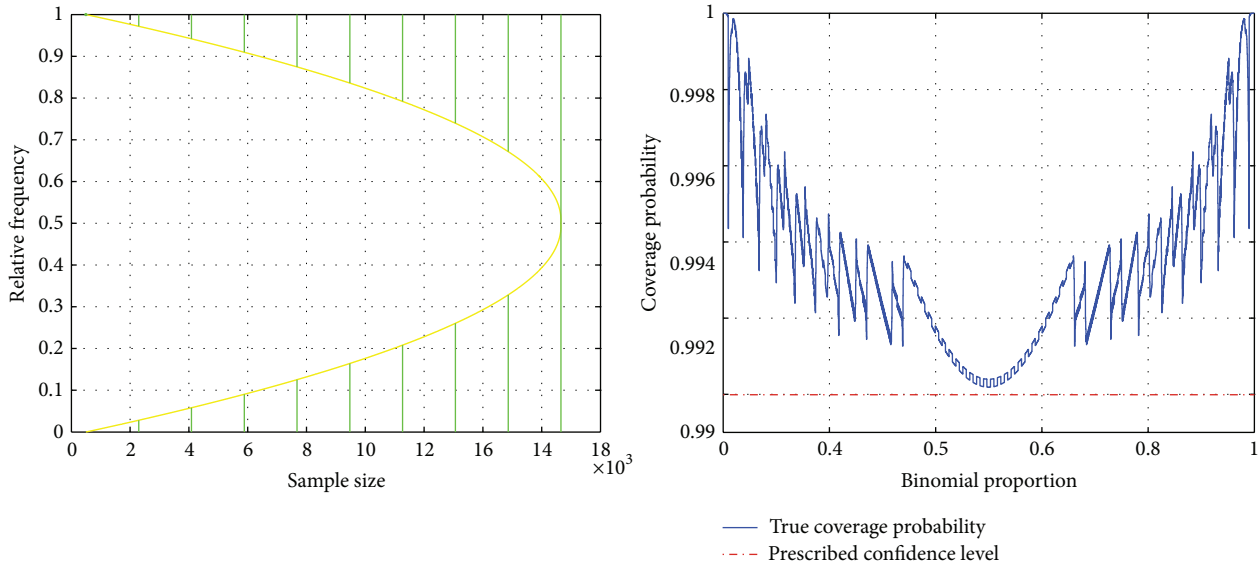


Figure 4: Double-parabolic sampling with $\epsilon = \delta = 0.01$, $s = 10$, $\rho = 3/4$, and $\zeta = 3.5753$.

of Figure 5. From Figure 5, it can be seen that a double-parabolic sampling scheme with dilation coefficient $\rho = 3/4$ has better performance in terms of average sample number and coverage probability as compared to that of the double-parabolic sampling scheme with smaller or larger values of dilation coefficient.

We have investigated the impact of confidence intervals on the performance of fully sequential sampling schemes constructed from the inclusion principle. We have observed that the stopping rule derived from Clopper-Pearson intervals generally outperforms the stopping rules derived from other types of confidence intervals. However, via appropriate choice of the dilation coefficient, the double-parabolic sampling scheme can perform uniformly better than the stopping rule derived from Clopper-Pearson intervals. To illustrate, consider the case that $\epsilon = 0.1$ and $\delta = 0.05$. For stopping rules derived from Clopper-Pearson intervals, Fishman's intervals, Wilson's intervals, and revised Wald intervals with $a = 4$, we have determined that to guarantee the prescribed confidence level $1 - \delta = 0.95$, it suffices to set the coverage tuning parameter ζ as $0.5, 1, 2.4$, and 0.37, respectively. For the stopping rule derived from Wald intervals, we have determined $\zeta = 0.77$ to ensure the confidence level, under the condition that the minimum sample size is taken as $\lceil (1/\epsilon) \ln(1/\zeta\delta) \rceil$. Recall that for the double-parabolic sampling scheme with $\rho = 3/4$, we have obtained $\zeta = 2.4$ for purpose of guaranteeing the confidence level. The average sample numbers of these sampling schemes are shown in Figure 6. From these plots, it can be seen that as compared to the stopping rule derived from Clopper-Pearson intervals, the stopping rule derived from the revised Wald intervals performs better in the region of p close to 0 or 1, but performs worse in the region of p in the middle of $(0, 1)$. The performance of stopping rules from Fishman's intervals (i.e., from Chernoff bound) and Wald intervals are obviously inferior as compared to that of the stopping rule derived from Clopper-Pearson intervals. It can be observed that the double-parabolic sampling scheme uniformly outperforms the stopping rule derived from Clopper-Pearson intervals.

6.5. Estimation with High Confidence Level. In some situations, we need to estimate a binomial proportion with a high confidence level. For example, one might want to construct a sampling scheme such that, for $\epsilon = 0.05$ and $\delta = 10^{-10}$, the resultant sequential estimator $\widehat{\mathbf{p}}$ satisfies $\Pr\{|\widehat{\mathbf{p}} - p| < \epsilon \mid p\} > 1 - \delta$ for any $p \in (0, 1)$. By working with the complementary coverage probability, we determined that it suffices to let the dilation coefficient $\rho = 3/4$ and the coverage tuning parameter $\zeta = 7.65$. The stopping boundary and the function of coverage probability with respect to the binomial proportion are displayed, respectively, in the left and right sides of Figure 7. As addressed in Section 5.3, it should be noted that it is impossible to obtain such a sampling scheme without working with the complementary coverage probability.

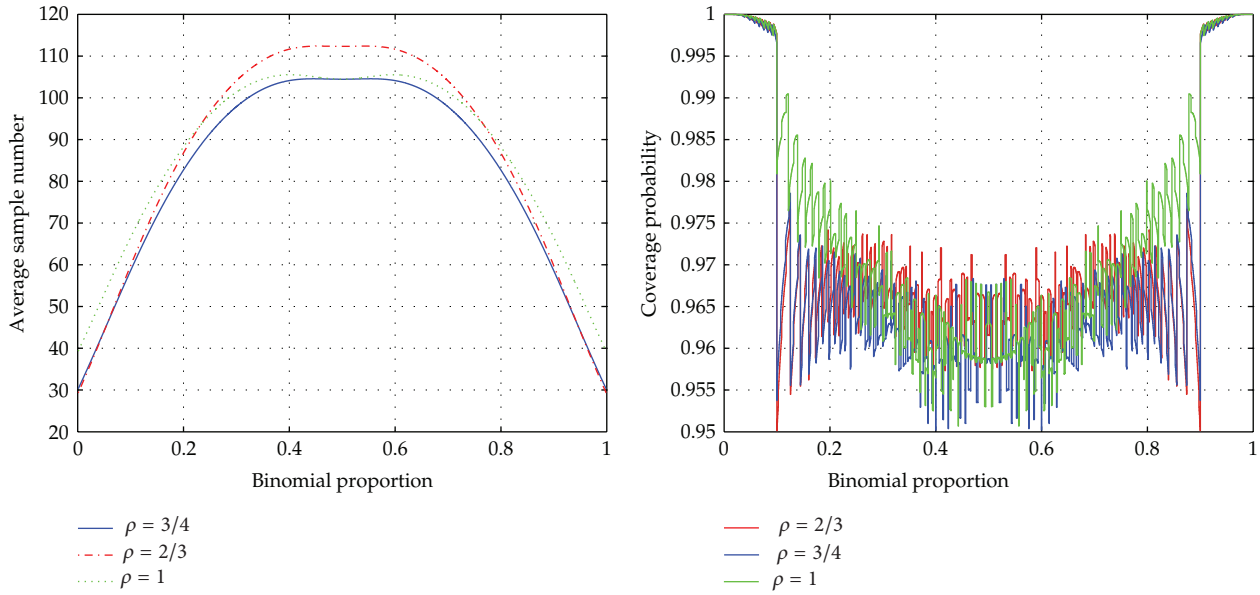


Figure 5: Double-parabolic sampling with various dilation coefficients.

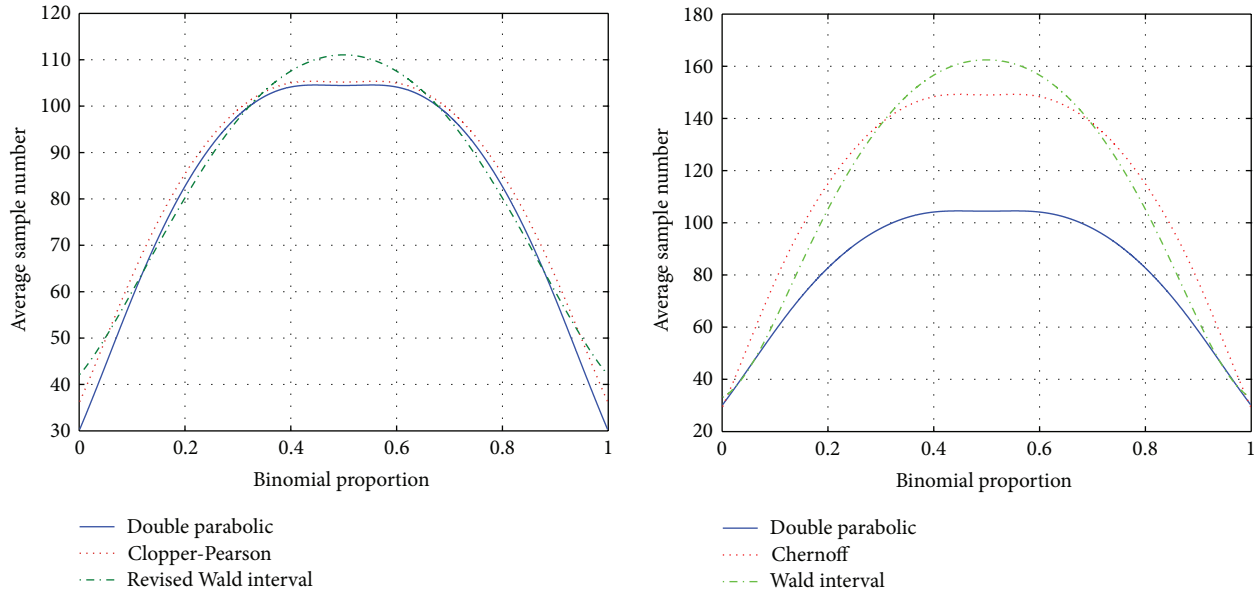


Figure 6: Comparison of average sample numbers.

7. Illustrative Examples for Clinical Trials

In this section, we shall illustrate the applications of our double-parabolic group sequential estimation method in clinical trials.

An example of our double-parabolic sampling scheme can be illustrated as follows. Assume that $\epsilon = \delta = 0.05$ is given and that the sampling procedure is expected to have 7 stages with sample sizes satisfying (8). Choosing $\rho = 3/4$, we have determined that it suffices to take $\zeta = 2.6759$ to guarantee that the coverage probability is no less than $1 - \delta = 0.95$ for all $p \in (0, 1)$. Accordingly, the sample sizes of this sampling scheme are calculated as 59, 116, 173, 231, 288, 345, and 403. This sampling scheme, with a sample path, is shown in the left side of Figure 8. In this case, the stopping rule can be equivalently described by virtue of Figure 8 as the following: continue sampling until $(\widehat{\mathbf{p}}_\ell, n_\ell)$ hit a green line at some stage. The coverage probability is shown in the right side of Figure 8.

To apply this estimation method in a clinical trial for estimating the proportion p of a binomial response with margin of error 0.05 and confidence level 95%, we can have seven groups of patients with group sizes 59, 57, 57, 58, 57, 57, and 58. In the first stage, we conduct experiment with

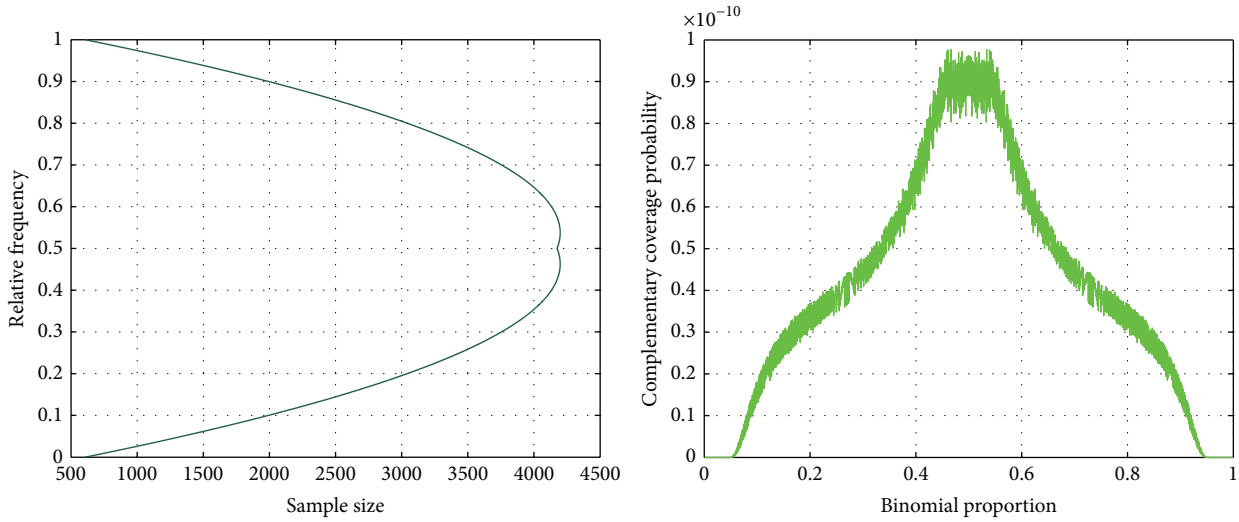


Figure 7: Double-parabolic sampling with $\epsilon = 0.05$, $\delta = 10^{-10}$, $\rho = 3/4$, and $\zeta = 7.65$.

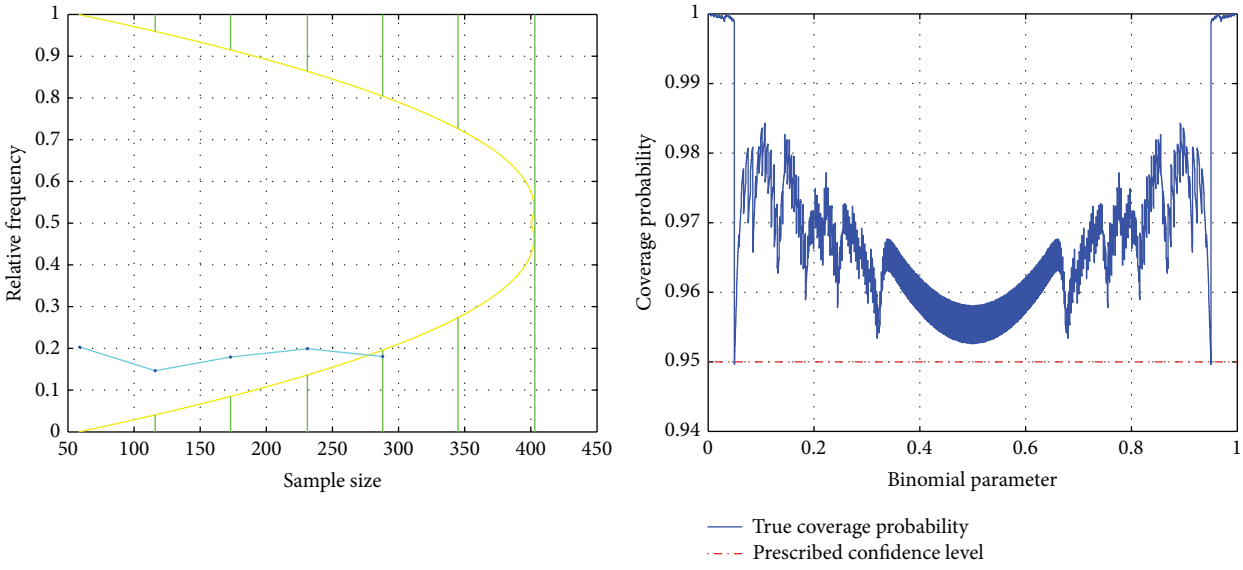


Figure 8: Double-parabolic sampling with $\epsilon = \delta = 0.05$, $s = 7$, $\rho = 3/4$, and $\zeta = 2.6759$.

the 59 patients of the first group. We observe the relative frequency of response and record it as $\widehat{\mathbf{p}}_1$. Suppose that there are 12 patients having positive responses, then the relative frequency at the first stage is $\widehat{\mathbf{p}}_1 = 12/59 = 0.2034$. With the values of $(\widehat{\mathbf{p}}_1, n_1) = (0.2034, 59)$, we check if the stopping rule is satisfied. This is equivalent to see if the point $(\widehat{\mathbf{p}}_1, n_1)$ hits a green line at the first stage. For such value of $(\widehat{\mathbf{p}}_1, n_1)$, it can be seen that the stopping condition is not fulfilled. So, we need to conduct the second stage of experiment with the 57 patients of the second group. We observe the response of these 57 patients. Suppose we observe that 5 patients among this group have positive responses. Then, we add 5 with 12, the number of positive responses before the second stage, to obtain 17 positive responses among $n_2 = 59 + 57 = 116$ patients. So, at the second stage, we get the relative frequency $\widehat{\mathbf{p}}_2 = 17/116 = 0.1466$. Since the stopping rule is not satisfied with the values of $(\widehat{\mathbf{p}}_2, n_2) = (0.1466, 116)$, we need to conduct the third stage of experiment with the 57 patients of the third group. Suppose we observe that 14 patients among this group have positive responses. Then, we add 14 with 17, the number of positive responses before the third stage, to get 31 positive responses among $n_3 = 59 + 57 + 57 = 173$ patients. So, at the third stage, we get the relative frequency $\widehat{\mathbf{p}}_3 = 31/173 = 0.1792$. Since the stopping rule is not satisfied with the values of $(\widehat{\mathbf{p}}_3, n_3) = (0.1792, 173)$, we need to conduct the fourth stage of experiment with the 58 patients of the fourth group. Suppose we observe that 15 patients among this group have positive responses. Then, we add 15 with 31, the number of positive responses before the fourth stage, to get 46 positive responses among $n_4 = 59 + 57 + 57 + 58 = 231$ patients. So, at the fourth stage, we get the relative frequency

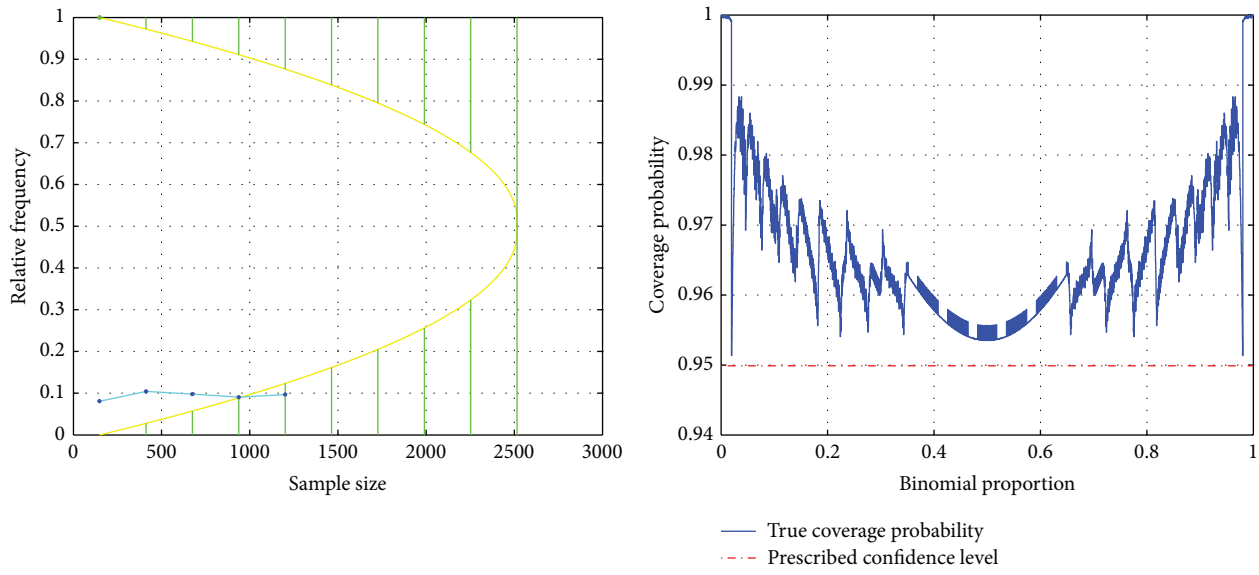


Figure 9: Double-parabolic sampling with $\epsilon = 0.02$, $\delta = 0.05$, $s = 10$, $\rho = 3/4$, and $\zeta = 2.6725$.

$\widehat{\mathbf{p}}_4 = 46/231 = 0.1991$. Since the stopping rule is not satisfied with the values of $(\widehat{\mathbf{p}}_4, n_4) = (0.1991, 231)$, we need to conduct the fifth stage of experiment with the 57 patients of the fifth group. Suppose we observe that 6 patients among this group have positive responses. Then, we add 6 with 46, the number of positive responses before the fifth stage, to get 52 positive responses among $n_5 = 59 + 57 + 57 + 58 + 57 = 288$ patients. So, at the fifth stage, we get the relative frequency $\widehat{\mathbf{p}}_5 = 52/288 = 0.1806$. It can be seen that the stopping rule is satisfied with the values of $(\widehat{\mathbf{p}}_5, n_5) = (0.1806, 288)$. Therefore, we can terminate the sampling experiment and take $\widehat{\mathbf{p}} = 52/288 = 0.1806$ as an estimate of the proportion of the whole population having positive responses. With a 95% confidence level, one can believe that the difference between the true value of p and its estimate $\widehat{\mathbf{p}} = 0.1806$ is less than 0.05.

In this experiment, we only use 288 samples to obtain the estimate for p. Except the round-off error, there is no other source of error for reporting statistical accuracy, since no asymptotic approximation is involved. As compared to fixed-sample-size procedure, we achieved a substantial save of samples. To see this, one can check that using the rigorous formula (37) gives a sample size 738, which is overly conservative. From the classical approximate formula (35), the sample size is determined as 385, which has been known to be insufficient to guarantee the prescribed confidence level 95%. The exact method of [34] shows that at least 391 samples are needed. As compared to the best-fixed-sample size obtained by the method of [34], the reduction of sample sizes resulted from our double-parabolic sampling scheme is $391 - 288 = 103$. It can be seen that the fixed-sample-size procedure wastes $103/288 = 35.76\%$ samples as compared to our group sequential method, which is also an exact method. This percentage may not be serious if it were a save of a number of simulation runs. However, as the number count is for patients, the reduction of samples is important for ethical and economical reasons. Using our group sequential method, the worst-case sample size is equal to 403, which is only 12 more than the minimum sample size of fixed-sample procedure. However, a lot of samples can be saved in the average case.

As ϵ or δ become smaller, the reduction of samples is more significant. For example, let $\epsilon = 0.02$ and $\delta = 0.05$, we have a double-parabolic sample scheme with 10 stages. The sampling scheme, with a sample path, is shown in the left side of Figure 9. The coverage probability is shown in the right side of Figure 9.

8. Conclusion

In this paper, we have reviewed recent development of group sequential estimation methods for a binomial proportion. We have illustrated the inclusion principle and its applications to various stopping rules. We have introduced computational techniques in the literature, which suffice for determining parameters of stopping rules to guarantee desired confidence levels. Moreover, we have proposed a new family of sampling schemes with stopping boundary of double-parabolic shape, which are parameterized by the coverage tuning parameter and the dilation coefficient. These parameters can be determined by the exact computational techniques to reduce the sampling cost, while ensuring prescribed confidence levels. The new family of sampling schemes are extremely simple in structure and asymptotically optimal as the margin of error tends to 0. We have established analytic bounds for the distribution and expectation of the sample number at the termination of the sampling process. We have obtained parameter values via the exact computational techniques for the proposed sampling schemes such that the confidence levels are guaranteed and that the sampling schemes are generally more efficient as compared to existing ones.

Appendices

A. Proof of Theorem 1

Consider function $g(x,z) = (x-z)^2/x(1-x)$ for $x \in (0,1)$ and $z \in [0,1]$. It can be checked that $\partial g(x,z)/\partial x = (x-z)[z(1-x)+x(1-z)][x(1-x)]^{-2}$, which shows that for any fixed $z \in [0,1]$, $-g(x,z)$ is a unimodal function of $x \in (0,1)$, with a maximum attained at $x = z$. By such a property of $g(x,z)$ and the definition of Wilson's confidence intervals, we have

$$\begin{aligned}
\left\{\widehat{\mathbf{p}}_\ell - \epsilon \le L_\ell\right\} &= \left\{0 < \widehat{\mathbf{p}}_\ell - \epsilon \le L_\ell\right\} \cup \left\{\widehat{\mathbf{p}}_\ell \le \epsilon\right\} \\
&= \left\{0 < \widehat{\mathbf{p}}_\ell - \epsilon \le L_\ell \le \widehat{\mathbf{p}}_\ell,\ g\left(L_\ell, \widehat{\mathbf{p}}_\ell\right) = \frac{\mathcal{Z}_{\zeta\delta}^2}{n_\ell}\right\} \\
&\quad \cup \left\{\widehat{\mathbf{p}}_\ell \le \epsilon\right\} \\
&= \left\{\widehat{\mathbf{p}}_\ell > \epsilon,\ \frac{\epsilon^2}{\left(\widehat{\mathbf{p}}_\ell - \epsilon\right)\left[1-\left(\widehat{\mathbf{p}}_\ell - \epsilon\right)\right]} \ge \frac{\mathcal{Z}_{\zeta\delta}^2}{n_\ell}\right\} \\
&\quad \cup \left\{\widehat{\mathbf{p}}_\ell \le \epsilon\right\}, \\
\left\{\widehat{\mathbf{p}}_\ell + \epsilon \ge U_\ell\right\} &= \left\{1 > \widehat{\mathbf{p}}_\ell + \epsilon \ge U_\ell\right\} \cup \left\{\widehat{\mathbf{p}}_\ell + \epsilon \ge 1\right\} \\
&= \left\{1 > \widehat{\mathbf{p}}_\ell + \epsilon \ge U_\ell \ge \widehat{\mathbf{p}}_\ell,\ g\left(U_\ell, \widehat{\mathbf{p}}_\ell\right) = \frac{\mathcal{Z}_{\zeta\delta}^2}{n_\ell}\right\} \\
\cup\left\{\widehat{\mathbf{p}}_\ell + \epsilon \ge 1\right\} &= \left\{\widehat{\mathbf{p}}_\ell < 1-\epsilon,\ \frac{\epsilon^2}{\left(\widehat{\mathbf{p}}_\ell + \epsilon\right)\left[1-\left(\widehat{\mathbf{p}}_\ell + \epsilon\right)\right]} \ge \frac{\mathcal{Z}_{\zeta\delta}^2}{n_\ell}\right\} \\
&\quad \cup \left\{\widehat{\mathbf{p}}_\ell \ge 1 - \epsilon\right\},
\end{aligned} \tag{A.1}$$

for $\ell = 1, \ldots, s$, where we have used the fact that $\{\widehat{\mathbf{p}}_\ell > \epsilon\} \subseteq \{L_\ell > 0\}$, $\{\widehat{\mathbf{p}}_\ell < 1-\epsilon\} \subseteq \{U_\ell < 1\}$ and $0 \le L_\ell \le \widehat{\mathbf{p}}_\ell \le U_\ell \le 1$. Recall that $0 < \epsilon < 1/2$. It follows that

$$\begin{aligned}
&\left\{\widehat{\mathbf{p}}_\ell - \epsilon \le L_\ell \le U_\ell \le \widehat{\mathbf{p}}_\ell + \epsilon\right\} \\
&= \left\{\epsilon < \widehat{\mathbf{p}}_\ell < 1-\epsilon,\ \frac{\epsilon^2}{\left(\widehat{\mathbf{p}}_\ell - \epsilon\right)\left[1-\left(\widehat{\mathbf{p}}_\ell - \epsilon\right)\right]} \ge \frac{\mathcal{Z}_{\zeta\delta}^2}{n_\ell},\right. \\
&\qquad \left.\frac{\epsilon^2}{\left(\widehat{\mathbf{p}}_\ell + \epsilon\right)\left[1-\left(\widehat{\mathbf{p}}_\ell + \epsilon\right)\right]} \ge \frac{\mathcal{Z}_{\zeta\delta}^2}{n_\ell}\right\} \\
&\quad \cup \left\{\widehat{\mathbf{p}}_\ell \le \epsilon,\ \frac{\epsilon^2}{\left(\widehat{\mathbf{p}}_\ell + \epsilon\right)\left[1-\left(\widehat{\mathbf{p}}_\ell + \epsilon\right)\right]} \ge \frac{\mathcal{Z}_{\zeta\delta}^2}{n_\ell}\right\} \\
&\quad \cup \left\{\widehat{\mathbf{p}}_\ell \ge 1 - \epsilon,\ \frac{\epsilon^2}{\left(\widehat{\mathbf{p}}_\ell - \epsilon\right)\left[1-\left(\widehat{\mathbf{p}}_\ell - \epsilon\right)\right]} \ge \frac{\mathcal{Z}_{\zeta\delta}^2}{n_\ell}\right\} \\
&= \left\{\epsilon < \widehat{\mathbf{p}}_\ell < 1-\epsilon,\ \left(\left|\widehat{\mathbf{p}}_\ell - \frac{1}{2}\right| - \epsilon\right)^2 \ge \frac{1}{4} - n_\ell\left(\frac{\epsilon}{\mathcal{Z}_{\zeta\delta}}\right)^2\right\} \\
&\quad \cup \left\{\widehat{\mathbf{p}}_\ell \le \epsilon,\ \left(\left|\widehat{\mathbf{p}}_\ell - \frac{1}{2}\right| - \epsilon\right)^2 \ge \frac{1}{4} - n_\ell\left(\frac{\epsilon}{\mathcal{Z}_{\zeta\delta}}\right)^2\right\} \\
&\quad \cup \left\{\widehat{\mathbf{p}}_\ell \ge 1 - \epsilon,\ \left(\left|\widehat{\mathbf{p}}_\ell - \frac{1}{2}\right| - \epsilon\right)^2 \ge \frac{1}{4} - n_\ell\left(\frac{\epsilon}{\mathcal{Z}_{\zeta\delta}}\right)^2\right\} \\
&= \left\{\left(\left|\widehat{\mathbf{p}}_\ell - \frac{1}{2}\right| - \epsilon\right)^2 \ge \frac{1}{4} - n_\ell\left(\frac{\epsilon}{\mathcal{Z}_{\zeta\delta}}\right)^2\right\},
\end{aligned} \tag{A.2}$$

for $\ell = 1, \ldots, s$. This completes the proof of the theorem.

B. Proof of Theorem 2

By the assumption that $n_s \ge (1/2\epsilon^2)\ln(1/\zeta\delta)$, we have $(1/4) + (\epsilon^2 n_s/2\ln(\zeta\delta)) \le 0$ and, consequently, $\Pr\{(|\widehat{\mathbf{p}}_s - 1/2| - \rho\epsilon)^2 \ge (1/4) + (\epsilon^2 n_s/2\ln(\zeta\delta))\} = 1$. It follows from the definition of the sampling scheme that the sampling process must stop at or before the sth stage. In other words, $\Pr\{\mathbf{l} \le s\} = 1$. This allows one to write

$$\begin{aligned}
\Pr\left\{\left|\widehat{\mathbf{p}} - p\right| \ge \epsilon \mid p\right\} &= \sum_{\ell=1}^{s} \Pr\left\{\left|\widehat{\mathbf{p}} - p\right| \ge \epsilon,\ \mathbf{l} = \ell \mid p\right\} \\
&= \sum_{\ell=1}^{s} \Pr\left\{\left|\widehat{\mathbf{p}}_\ell - p\right| \ge \epsilon,\ \mathbf{l} = \ell \mid p\right\} \\
&\le \sum_{\ell=1}^{s} \Pr\left\{\left|\widehat{\mathbf{p}}_\ell - p\right| \ge \epsilon \mid p\right\},
\end{aligned} \tag{B.1}$$

for $p \in (0,1)$. By virtue of the well-known Chernoff-Hoeffding bound [32, 33], we have

$$\Pr\left\{\left|\widehat{\mathbf{p}}_\ell - p\right| \ge \epsilon \mid p\right\} \le 2\exp\left(-2n_\ell\epsilon^2\right), \tag{B.2}$$

for $\ell = 1, \ldots, s$. Making use of (B.1), (B.2), and the fact that $n_1 \ge 2\rho((1/\epsilon) - \rho)\ln(1/\zeta\delta)$ as can be seen from (30), we have

$$\begin{aligned}
&\Pr\left\{\left|\widehat{\mathbf{p}} - p\right| \ge \epsilon \mid p\right\} \\
&\le 2\sum_{\ell=1}^{s} \exp\left(-2n_\ell\epsilon^2\right) \\
&\le 2\sum_{m=n_1}^{\infty} \exp\left(-2m\epsilon^2\right) = \frac{2\exp\left(-2n_1\epsilon^2\right)}{1-\exp\left(-2\epsilon^2\right)} \\
&\le \frac{2\exp\left(-2\epsilon^2 \times 2\rho\left((1/\epsilon) - \rho\right)\ln(1/\zeta\delta)\right)}{1-\exp\left(-2\epsilon^2\right)} \\
&= \frac{2\exp\left(4\epsilon\rho(1-\rho\epsilon)\ln(\zeta\delta)\right)}{1-\exp\left(-2\epsilon^2\right)},
\end{aligned} \tag{B.3}$$

for any $p \in (0,1)$. Therefore, to guarantee that $\Pr\{|\widehat{\mathbf{p}} - p| < \epsilon \mid p\} \ge 1 - \delta$ for any $p \in (0,1)$, it is sufficient to choose ζ

such that $2\exp(4\epsilon\rho(1-\rho\epsilon)\ln(\zeta\delta)) \le \delta[1-\exp(-2\epsilon^2)]$. This inequality can be written as $4\epsilon\rho(1-\rho\epsilon)\ln(\zeta\delta) \le \ln(\delta/2) + \ln[1-\exp(-2\epsilon^2)]$ or, equivalently, $\zeta \le (1/\delta)\exp((\ln(\delta/2) + \ln[1-\exp(-2\epsilon^2)])/4\epsilon\rho(1-\rho\epsilon))$. The proof of the theorem is thus completed.

C. Proof of Theorem 3

First, we need to show that $\Pr\{\lim_{\epsilon\to 0}(\mathbf{n}/N(p,\epsilon,\delta,\zeta)) = 1 \mid p\} = 1$ for any $p \in (0,1)$. Clearly, the sample number $\mathbf{n}$ is a random number dependent on ϵ. Note that for any $\omega \in \Omega$, the sequences $\{\overline{X}_{\mathbf{n}(\omega)}(\omega)\}_{\epsilon\in(0,1)}$ and $\{\overline{X}_{\mathbf{n}(\omega)-1}(\omega)\}_{\epsilon\in(0,1)}$ are subsets of $\{\overline{X}_m(\omega)\}_{m=1}^{\infty}$. By the strong law of large numbers, for almost every $\omega \in \Omega$, the sequence $\{\overline{X}_m(\omega)\}_{m=1}^{\infty}$ converges to p. Since every subsequence of a convergent sequence must converge, it follows that the sequences $\{\overline{X}_{\mathbf{n}(\omega)}(\omega)\}_{\epsilon\in(0,1)}$ and $\{\overline{X}_{\mathbf{n}(\omega)-1}(\omega)\}_{\epsilon\in(0,1)}$ converge to p as $\epsilon \to 0$ provided that $\mathbf{n}(\omega) \to \infty$ as $\epsilon \to 0$. Since it is certain that $\mathbf{n} \ge 2\rho((1/\epsilon) - \rho)\ln(1/\zeta\delta) \to \infty$ as $\epsilon \to 0$, we have that $\{\lim_{\epsilon\to 0}((\mathbf{n}-1)/\mathbf{n}) = 1\}$ is a sure event. It follows that $B = \{\lim_{\epsilon\to 0}\overline{X}_{\mathbf{n}-1} = p,\ \lim_{\epsilon\to 0}\overline{X}_{\mathbf{n}} = p,\ \lim_{\epsilon\to 0}((\mathbf{n}-1)/\mathbf{n}) = 1\}$ is an almost sure event. By the definition of the sampling scheme, we have

$$A = \left\{\left(\left|\overline{X}_{\mathbf{n}-1} - \frac{1}{2}\right| - \rho\epsilon\right)^2 < \frac{1}{4} + \frac{\epsilon^2(\mathbf{n}-1)}{2\ln(\zeta\delta)},\ \left(\left|\overline{X}_{\mathbf{n}} - \frac{1}{2}\right| - \rho\epsilon\right)^2 \ge \frac{1}{4} + \frac{\epsilon^2\mathbf{n}}{2\ln(\zeta\delta)}\right\} \tag{C.1}$$

as a sure event. Hence, $A \cap B$ is an almost sure event. Define $C = \{\lim_{\epsilon\to 0}(\mathbf{n}/N(p,\epsilon,\delta,\zeta)) = 1\}$. We need to show that C is an almost sure event. For this purpose, we let $\omega \in A \cap B$ and expect to show that $\omega \in C$. As a consequence of $\omega \in A \cap B$,

$$\frac{\mathbf{n}(\omega)}{N(p,\epsilon,\delta,\zeta)} < \frac{\mathbf{n}(\omega)}{\mathbf{n}(\omega)-1} \times \frac{\left[(1/4) - \left(\left|\overline{X}_{\mathbf{n}(\omega)-1}(\omega) - (1/2)\right| - \rho\epsilon\right)^2\right]}{p(1-p)},\qquad \lim_{\epsilon\to 0}\overline{X}_{\mathbf{n}(\omega)-1}(\omega) = p,\qquad \lim_{\epsilon\to 0}\frac{\mathbf{n}(\omega)-1}{\mathbf{n}(\omega)} = 1. \tag{C.2}$$

By the continuity of the function $|x-1/2| - \rho\epsilon$ with respect to x and ϵ, we have

$$\begin{aligned}\limsup_{\epsilon\to 0}\frac{\mathbf{n}(\omega)}{N(p,\epsilon,\delta,\zeta)} &\le \lim_{\epsilon\to 0}\frac{\mathbf{n}(\omega)}{\mathbf{n}(\omega)-1} \times \frac{\left[(1/4) - \left(\left|\lim_{\epsilon\to 0}\overline{X}_{\mathbf{n}(\omega)-1}(\omega) - (1/2)\right| - \lim_{\epsilon\to 0}\rho\epsilon\right)^2\right]}{p(1-p)}\\ &= 1.\end{aligned} \tag{C.3}$$

On the other hand, as a consequence of $\omega \in A \cap B$,

$$\frac{\mathbf{n}(\omega)}{N(p,\epsilon,\delta,\zeta)} \ge \frac{\left[(1/4) - \left(\left|\overline{X}_{\mathbf{n}(\omega)}(\omega) - (1/2)\right| - \rho\epsilon\right)^2\right]}{p(1-p)},\qquad \lim_{\epsilon\to 0}\overline{X}_{\mathbf{n}(\omega)}(\omega) = p. \tag{C.4}$$

Making use of the continuity of the function $|x-1/2| - \rho\epsilon$ with respect to x and ϵ, we have

$$\begin{aligned}\liminf_{\epsilon\to 0}\frac{\mathbf{n}(\omega)}{N(p,\epsilon,\delta,\zeta)} &\ge \frac{\left[(1/4) - \left(\left|\lim_{\epsilon\to 0}\overline{X}_{\mathbf{n}(\omega)}(\omega) - (1/2)\right| - \lim_{\epsilon\to 0}\rho\epsilon\right)^2\right]}{p(1-p)}\\ &= 1.\end{aligned} \tag{C.5}$$

Combining (C.3) and (C.5) yields $\lim_{\epsilon\to 0}(\mathbf{n}(\omega)/N(p,\epsilon,\delta,\zeta)) = 1$ and thus $A \cap B \subseteq C$. This implies that C is an almost sure event and thus $\Pr\{\lim_{\epsilon\to 0}(\mathbf{n}/N(p,\epsilon,\delta,\zeta)) = 1 \mid p\} = 1$ for $p \in (0,1)$.

Next, we need to show that $\lim_{\epsilon\to 0}\Pr\{|\widehat{\mathbf{p}} - p| < \epsilon \mid p\} = 2\Phi(\sqrt{2\ln(1/\zeta\delta)}) - 1$ for any $p \in (0,1)$. For simplicity of notations, let $\sigma = \sqrt{p(1-p)}$ and $a = \sqrt{2\ln(1/\zeta\delta)}$. Note that $\Pr\{|\widehat{\mathbf{p}} - p| < \epsilon \mid p\} = \Pr\{|\overline{X}_{\mathbf{n}} - p| < \epsilon \mid p\} = \Pr\{\sqrt{\mathbf{n}}|\overline{X}_{\mathbf{n}} - p|/\sigma < \epsilon\sqrt{\mathbf{n}}/\sigma\}$. Clearly, for any $\eta \in (0,a)$,

$$\begin{aligned}&\Pr\left\{\frac{\sqrt{\mathbf{n}}\left|\overline{X}_{\mathbf{n}} - p\right|}{\sigma} < \frac{\epsilon\sqrt{\mathbf{n}}}{\sigma}\right\}\\ &\le \Pr\left\{\frac{\sqrt{\mathbf{n}}\left|\overline{X}_{\mathbf{n}} - p\right|}{\sigma} < \frac{\epsilon\sqrt{\mathbf{n}}}{\sigma},\ \frac{\epsilon\sqrt{\mathbf{n}}}{\sigma} \in [a-\eta, a+\eta]\right\} + \Pr\left\{\frac{\epsilon\sqrt{\mathbf{n}}}{\sigma} \notin [a-\eta, a+\eta]\right\}\\ &\le \Pr\left\{\frac{\sqrt{\mathbf{n}}\left|\overline{X}_{\mathbf{n}} - p\right|}{\sigma} < a+\eta,\ \frac{\epsilon\sqrt{\mathbf{n}}}{\sigma} \in [a-\eta, a+\eta]\right\} + \Pr\left\{\frac{\epsilon\sqrt{\mathbf{n}}}{\sigma} \notin [a-\eta, a+\eta]\right\}\\ &\le \Pr\left\{\frac{\sqrt{\mathbf{n}}\left|\overline{X}_{\mathbf{n}} - p\right|}{\sigma} < a+\eta\right\} + \Pr\left\{\frac{\epsilon\sqrt{\mathbf{n}}}{\sigma} \notin [a-\eta, a+\eta]\right\},\end{aligned} \tag{C.6}$$

$$\begin{aligned}
&\Pr\left\{\frac{\sqrt{\mathbf{n}}\left|\overline{X}_{\mathbf{n}}-p\right|}{\sigma}<\frac{\epsilon\sqrt{\mathbf{n}}}{\sigma}\right\}\\
&\quad\geq \Pr\left\{\frac{\sqrt{\mathbf{n}}\left|\overline{X}_{\mathbf{n}}-p\right|}{\sigma}<\frac{\epsilon\sqrt{\mathbf{n}}}{\sigma},\ \frac{\epsilon\sqrt{\mathbf{n}}}{\sigma}\in[a-\eta,a+\eta]\right\}\\
&\quad\geq \Pr\left\{\frac{\sqrt{\mathbf{n}}\left|\overline{X}_{\mathbf{n}}-p\right|}{\sigma}<a-\eta,\ \frac{\epsilon\sqrt{\mathbf{n}}}{\sigma}\in[a-\eta,a+\eta]\right\}\\
&\quad\geq \Pr\left\{\frac{\sqrt{\mathbf{n}}\left|\overline{X}_{\mathbf{n}}-p\right|}{\sigma}<a-\eta\right\}\\
&\qquad-\Pr\left\{\frac{\epsilon\sqrt{\mathbf{n}}}{\sigma}\notin[a-\eta,a+\eta]\right\}.
\end{aligned}\tag{C.7}$$

Recall that we have established that $\mathbf{n}/N(p,\epsilon,\delta,\zeta)\to 1$ almost surely as $\epsilon\to 0$. This implies that $(\epsilon\sqrt{\mathbf{n}}/\sigma)\to a$ and $\mathbf{n}/N(p,\epsilon,\delta,\zeta)\to 1$ in probability as ϵ tends to zero. It follows from Anscombe's random central limit theorem [35] that as ϵ tends to zero, $\sqrt{\mathbf{n}}(\overline{X}_{\mathbf{n}}-p)/\sigma$ converges in distribution to a Gaussian random variable with zero mean and unit variance. Hence, from (C.6),

$$\begin{aligned}
&\limsup_{\epsilon\to 0}\Pr\left\{\frac{\sqrt{\mathbf{n}}\left|\overline{X}_{\mathbf{n}}-p\right|}{\sigma}<\frac{\epsilon\sqrt{\mathbf{n}}}{\sigma}\right\}\\
&\quad\leq \lim_{\epsilon\to 0}\Pr\left\{\frac{\sqrt{\mathbf{n}}\left|\overline{X}_{\mathbf{n}}-p\right|}{\sigma}<a+\eta\right\}\\
&\qquad+\lim_{\epsilon\to 0}\Pr\left\{\frac{\epsilon\sqrt{\mathbf{n}}}{\sigma}\notin[a-\eta,a+\eta]\right\}\\
&\quad=2\Phi(a+\eta)-1
\end{aligned}\tag{C.8}$$

and from (C.7),

$$\begin{aligned}
&\liminf_{\epsilon\to 0}\Pr\left\{\frac{\sqrt{\mathbf{n}}\left|\overline{X}_{\mathbf{n}}-p\right|}{\sigma}<\frac{\epsilon\sqrt{\mathbf{n}}}{\sigma}\right\}\\
&\quad\geq \lim_{\epsilon\to 0}\Pr\left\{\frac{\sqrt{\mathbf{n}}\left|\overline{X}_{\mathbf{n}}-p\right|}{\sigma}<a-\eta\right\}\\
&\qquad-\lim_{\epsilon\to 0}\Pr\left\{\frac{\epsilon\sqrt{\mathbf{n}}}{\sigma}\notin[a-\eta,a+\eta]\right\}\\
&\quad=2\Phi(a-\eta)-1.
\end{aligned}\tag{C.9}$$

Since this argument holds for arbitrarily small $\eta\in(0,a)$, it must be true that

$$\begin{aligned}
&\liminf_{\epsilon\to 0}\Pr\left\{\frac{\sqrt{\mathbf{n}}\left|\overline{X}_{\mathbf{n}}-p\right|}{\sigma}<\frac{\epsilon\sqrt{\mathbf{n}}}{\sigma}\right\}\\
&\quad=\limsup_{\epsilon\to 0}\Pr\left\{\frac{\sqrt{\mathbf{n}}\left|\overline{X}_{\mathbf{n}}-p\right|}{\sigma}<\frac{\epsilon\sqrt{\mathbf{n}}}{\sigma}\right\}\\
&\quad=2\Phi(a)-1.
\end{aligned}\tag{C.10}$$

So, $\lim_{\epsilon\to 0}\Pr\{|\widehat{\mathbf{p}}-p|<\epsilon\mid p\}=\lim_{\epsilon\to 0}\Pr\{\sqrt{\mathbf{n}}|\overline{X}_{\mathbf{n}}-p|/\sigma<\epsilon\sqrt{\mathbf{n}}/\sigma\}=2\Phi(a)-1=2\Phi(\sqrt{2\ln(1/\zeta\delta)})-1$ for any $p\in(0,1)$.

Now, we focus our attention to show that $\lim_{\epsilon\to 0}(\mathbb{E}[\mathbf{n}]/N(p,\epsilon,\delta,\zeta))=1$ for any $p\in(0,1)$. For this purpose, it suffices to show that

$$\begin{aligned}
1-\eta&\leq\liminf_{\epsilon\to 0}\frac{\mathbb{E}[\mathbf{n}]}{N(p,\epsilon,\delta,\zeta)}\\
&\leq\limsup_{\epsilon\to 0}\frac{\mathbb{E}[\mathbf{n}]}{N(p,\epsilon,\delta,\zeta)}\\
&\leq 1+\eta,\quad\forall p\in(0,1),
\end{aligned}\tag{C.11}$$

for any $\eta\in(0,1)$. For simplicity of notations, we abbreviate $N(p,\epsilon,\delta,\zeta)$ as N in the sequel. Since we have established $\Pr\{\lim_{\epsilon\to 0}(\mathbf{n}/N(p,\epsilon,\delta,\zeta))=1\}=1$, we can conclude that

$$\lim_{\epsilon\to 0}\Pr\{(1-\eta)N\leq\mathbf{n}\leq(1+\eta)N\}=1.\tag{C.12}$$

Noting that

$$\begin{aligned}
\mathbb{E}[\mathbf{n}]&=\sum_{m=0}^{\infty}m\Pr\{\mathbf{n}=m\}\\
&\geq\sum_{(1-\eta)N\leq m\leq(1+\eta)N}m\Pr\{\mathbf{n}=m\}\\
&\geq(1-\eta)N\sum_{(1-\eta)N\leq m\leq(1+\eta)N}\Pr\{\mathbf{n}=m\},
\end{aligned}\tag{C.13}$$

we have

$$\mathbb{E}[\mathbf{n}]\geq(1-\eta)N\Pr\{(1-\eta)N\leq\mathbf{n}\leq(1+\eta)N\}.\tag{C.14}$$

Combining (C.12) and (C.14) yields

$$\begin{aligned}
&\liminf_{\epsilon\to 0}\frac{\mathbb{E}[\mathbf{n}]}{N(p,\epsilon,\delta,\zeta)}\\
&\quad\geq(1-\eta)\lim_{\epsilon\to 0}\Pr\{(1-\eta)N\leq\mathbf{n}\leq(1+\eta)N\}\\
&\quad=1-\eta.
\end{aligned}\tag{C.15}$$

On the other hand, using $\mathbb{E}[\mathbf{n}]=\sum_{m=0}^{\infty}\Pr\{\mathbf{n}>m\}$, we can write

$$\begin{aligned}
\mathbb{E}[\mathbf{n}]&=\sum_{0\leq m<(1+\eta)N}\Pr\{\mathbf{n}>m\}+\sum_{m\geq(1+\eta)N}\Pr\{\mathbf{n}>m\}\\
&\leq\lceil(1+\eta)N\rceil+\sum_{m\geq(1+\eta)N}\Pr\{\mathbf{n}>m\}.
\end{aligned}\tag{C.16}$$

Since $\limsup_{\epsilon\to 0}(\lceil(1+\eta)N\rceil/N(p,\epsilon,\delta,\zeta))=1+\eta$, for the purpose of establishing $\limsup_{\epsilon\to 0}(\mathbb{E}[\mathbf{n}]/N(p,\epsilon,\delta,\zeta))\leq 1+\eta$, it remains to show that

$$\limsup_{\epsilon\to 0}\frac{\sum_{m\geq(1+\eta)N}\Pr\{\mathbf{n}>m\}}{N(p,\epsilon,\delta,\zeta)}=0.\tag{C.17}$$

Consider functions $f(x) = (1/4) - (|x - 1/2| - \rho\epsilon)^2$ and $g(x) = x(1-x)$ for $x \in [0,1]$. Note that

$$\begin{aligned} |f(x) - g(x)| &= \left|(x-1/2)^2 - (|x-1/2| - \rho\epsilon)^2\right| \\ &= \rho\epsilon\left||2x-1| - \rho\epsilon\right| \le \rho\epsilon(1+\rho\epsilon), \end{aligned} \tag{C.18}$$

for all $x \in [0,1]$. For $p \in (0,1)$, there exists a positive number $\gamma < \min\{p, 1-p\}$ such that $|g(x) - g(p)| < (\eta/2)p(1-p)$ for any $x \in (p-\gamma, p+\gamma)$, since $g(x)$ is a continuous function of x. From now on, let $\epsilon > 0$ be sufficiently small such that $\rho\epsilon(1+\rho\epsilon) < (\eta/2)p(1-p)$. Then,

$$\begin{aligned} f(x) &\le g(x) + \rho\epsilon(1+\rho\epsilon) \\ &< g(p) + \frac{\eta}{2}p(1-p) + \rho\epsilon(1+\rho\epsilon) \\ &< (1+\eta)\,p(1-p), \end{aligned} \tag{C.19}$$

for all $x \in (p-\gamma, p+\gamma)$. This implies that

$$\begin{aligned} &\left\{\overline{X}_m \in (p-\gamma, p+\gamma)\right\} \\ &\quad\subseteq \left\{(1+\eta)\,p(1-p) \ge \frac{1}{4} - \left(\left|\overline{X}_m - \frac{1}{2}\right| - \rho\epsilon\right)^2\right\} \end{aligned} \tag{C.20}$$

for all $m > 0$. Taking complementary events on both sides of (C.20) leads to

$$\begin{aligned} &\left\{(1+\eta)\,p(1-p) < \frac{1}{4} - \left(\left|\overline{X}_m - \frac{1}{2}\right| - \rho\epsilon\right)^2\right\} \\ &\quad\subseteq \left\{\overline{X}_m \notin (p-\gamma, p+\gamma)\right\}, \end{aligned} \tag{C.21}$$

for all $m > 0$. Since $(1+\eta)p(1-p) = ((1+\eta)N\epsilon^2/2\ln(1/\zeta\delta)) \le (m\epsilon^2/2\ln(1/\zeta\delta))$ for all $m \ge (1+\eta)N$, it follows that

$$\begin{aligned} &\left\{\frac{m\epsilon^2}{2\ln(1/\zeta\delta)} < \frac{1}{4} - \left(\left|\overline{X}_m - \frac{1}{2}\right| - \rho\epsilon\right)^2\right\} \\ &\quad\subseteq \left\{\overline{X}_m \notin (p-\gamma, p+\gamma)\right\}, \end{aligned} \tag{C.22}$$

for all $m \ge (1+\eta)N$. Therefore, we have shown that if ϵ is sufficiently small, then there exists a number $\gamma > 0$ such that

$$\begin{aligned} \{\mathbf{n} > m\} &\subseteq \left\{\left(\left|\overline{X}_m - \frac{1}{2}\right| - \rho\epsilon\right)^2 < \frac{1}{4} + \frac{m\epsilon^2}{2\ln(\zeta\delta)}\right\} \\ &\subseteq \left\{\overline{X}_m \notin (p-\gamma, p+\gamma)\right\}, \end{aligned} \tag{C.23}$$

for all $m \ge (1+\eta)N$. Using this inclusion relationship and the Chernoff-Hoeffding bound [32, 33], we have

$$\Pr\{\mathbf{n} > m\} \le \Pr\left\{\overline{X}_m \notin (p-\gamma, p+\gamma)\right\} \le 2\exp\left(-2m\gamma^2\right), \tag{C.24}$$

for all $m \ge (1+\eta)N$ provided that $\epsilon > 0$ is sufficiently small. Letting $k = \lceil (1+\eta)N \rceil$ and using (C.24), we have

$$\begin{aligned} \sum_{m \ge (1+\eta)N} \Pr\{\mathbf{n} > m\} &= \sum_{m \ge k} \Pr\{\mathbf{n} > m\} \\ &\le \sum_{m \ge k} 2\exp\left(-2m\gamma^2\right) \\ &= \frac{2\exp\left(-2k\gamma^2\right)}{1 - \exp\left(-2\gamma^2\right)}, \end{aligned} \tag{C.25}$$

provided that ϵ is sufficiently small. Consequently,

$$\begin{aligned} &\limsup_{\epsilon \to 0} \frac{\sum_{m \ge (1+\eta)N} \Pr\{\mathbf{n} > m\}}{N(p, \epsilon, \delta, \zeta)} \\ &\quad\le \limsup_{\epsilon \to 0} \frac{2}{N} \frac{\exp\left(-2k\gamma^2\right)}{1 - \exp\left(-2\gamma^2\right)} = 0, \end{aligned} \tag{C.26}$$

since $k \to \infty$ and $N \to \infty$ as $\epsilon \to 0$. So, we have established (C.11). Since the argument holds for arbitrarily small $\eta > 0$, it must be true that $\lim_{\epsilon \to 0}(\mathbb{E}[\mathbf{n}]/N(p, \epsilon, \delta, \zeta)) = 1$ for any $p \in (0,1)$. This completes the proof of the theorem.

D. Proof of Theorem 4

Recall that $\mathbf{l}$ denotes the index of stage at the termination of the sampling process. Observing that

$$\begin{aligned} n_s - n_1 \ \Pr\{\mathbf{l} = 1\} &= n_s \Pr\{\mathbf{l} \le s\} - n_1 \Pr\{\mathbf{l} \le 1\} \\ &= \sum_{\ell=2}^{s} \left(n_\ell \Pr\{\mathbf{l} \le \ell\} - n_{\ell-1} \Pr\{\mathbf{l} < \ell\}\right) \\ &= \sum_{\ell=2}^{s} n_\ell \ \left(\Pr\{\mathbf{l} \le \ell\} - \Pr\{\mathbf{l} < \ell\}\right) \\ &\quad + \sum_{\ell=2}^{s} \left(n_\ell - n_{\ell-1}\right) \Pr\{\mathbf{l} < \ell\} \\ &= \sum_{\ell=2}^{s} n_\ell \Pr\{\mathbf{l} = \ell\} \\ &\quad + \sum_{\ell=1}^{s-1} \left(n_{\ell+1} - n_\ell\right) \Pr\{\mathbf{l} \le \ell\}, \end{aligned} \tag{D.1}$$

we have $n_s - \sum_{\ell=1}^{s} n_\ell \Pr\{\mathbf{l} = \ell\} = \sum_{\ell=1}^{s-1} (n_{\ell+1} - n_\ell)\Pr\{\mathbf{l} \le \ell\}$. Making use of this result and the fact $n_s = n_1 + \sum_{\ell=1}^{s-1}(n_{\ell+1} - n_\ell)$, we have

$$\begin{aligned}
\mathbb{E}[\mathbf{n}] &= \sum_{\ell=1}^{s} n_\ell \Pr\{\mathbf{l} = \ell\} = n_s - \left(n_s - \sum_{\ell=1}^{s} n_\ell \Pr\{\mathbf{l} = \ell\} \right) \\
&= n_1 + \sum_{\ell=1}^{s-1} (n_{\ell+1} - n_\ell) - \sum_{\ell=1}^{s-1} (n_{\ell+1} - n_\ell) \Pr\{\mathbf{l} \le \ell\} \\
&= n_1 + \sum_{\ell=1}^{\tau-1} (n_{\ell+1} - n_\ell) \Pr\{\mathbf{l} > \ell\} \\
&\quad + \sum_{\ell=\tau}^{s-1} (n_{\ell+1} - n_\ell) \Pr\{\mathbf{l} > \ell\}.
\end{aligned} \tag{D.2}$$

By the definition of the stopping rule, we have

$$\begin{aligned}
\{\mathbf{l} > \ell\} &\subseteq \left\{ \left(\left| \widehat{\mathbf{p}}_\ell - \frac{1}{2} \right| - \rho\epsilon \right)^2 < \frac{1}{4} + \frac{\epsilon^2 n_\ell}{2\ln(\zeta\delta)} \right\} \\
&= \left\{ \rho\epsilon - \sqrt{\frac{1}{4} + \frac{\epsilon^2 n_\ell}{2\ln(\zeta\delta)}} < \left| \widehat{\mathbf{p}}_\ell - \frac{1}{2} \right| \right. \\
&\qquad \left. < \rho\epsilon + \sqrt{\frac{1}{4} + \frac{\epsilon^2 n_\ell}{2\ln(\zeta\delta)}} \right\} \\
&= \left\{ \rho\epsilon - \sqrt{\frac{1}{4} + \frac{\epsilon^2 n_\ell}{2\ln(\zeta\delta)}} < \frac{1}{2} - \widehat{\mathbf{p}}_\ell \right. \\
&\qquad \left. < \rho\epsilon + \sqrt{\frac{1}{4} + \frac{\epsilon^2 n_\ell}{2\ln(\zeta\delta)}},\ \widehat{\mathbf{p}}_\ell \le \frac{1}{2} \right\} \\
&\quad \cup \left\{ \rho\epsilon - \sqrt{\frac{1}{4} + \frac{\epsilon^2 n_\ell}{2\ln(\zeta\delta)}} < \widehat{\mathbf{p}}_\ell - \frac{1}{2} \right. \\
&\qquad \left. < \rho\epsilon + \sqrt{\frac{1}{4} + \frac{\epsilon^2 n_\ell}{2\ln(\zeta\delta)}},\ \widehat{\mathbf{p}}_\ell > \frac{1}{2} \right\} \\
&\subseteq \left\{ a_\ell < \widehat{\mathbf{p}}_\ell < b_\ell \right\} \cup \left\{ 1 - b_\ell < \widehat{\mathbf{p}}_\ell < 1 - a_\ell \right\},
\end{aligned} \tag{D.3}$$

for $1 \le \ell < s$, where $b_\ell = (1/2) - \rho\epsilon + \sqrt{(1/4) + (\epsilon^2 n_\ell / 2\ln(\zeta\delta))}$ for $\ell = 1, \ldots, s-1$. By the assumption that ϵ and ρ are nonnegative, we have $1 - b_\ell - a_\ell = 2\rho\epsilon \ge 0$ for $\ell = 1, \ldots, s-1$. It follows from (D.3) that $\{\mathbf{l} > \ell\} \subseteq \{\widehat{\mathbf{p}}_\ell > a_\ell\}$ for $\ell = 1, \ldots, s-1$. By the definition of τ, we have $p < a_\ell$ for $\tau \le \ell < s$. Making use of this fact, the inclusion relationship $\{\mathbf{l} > \ell\} \subseteq \{\widehat{\mathbf{p}}_\ell > a_\ell\}$, $\ell = 1, \ldots, s-1$, and Chernoff-Hoeffding bound [32, 33], we have

$$\begin{aligned}
\Pr\{\mathbf{n} > n_\ell \mid p\} &= \Pr\{\mathbf{l} > \ell \mid p\} \\
&\le \Pr\{\widehat{\mathbf{p}}_\ell > a_\ell \mid p\} \le \exp\left(n_\ell \mathscr{M}(a_\ell, p)\right)
\end{aligned} \tag{D.4}$$

for $\tau \le \ell < s$. It follows from (D.2) and (D.4) that

$$\begin{aligned}
\mathbb{E}[\mathbf{n}] &\le n_1 + \sum_{\ell=1}^{\tau-1} (n_{\ell+1} - n_\ell) + \sum_{\ell=\tau}^{s-1} (n_{\ell+1} - n_\ell) \Pr\{\mathbf{l} > \ell\} \\
&= n_\tau + \sum_{\ell=\tau}^{s-1} (n_{\ell+1} - n_\ell) \Pr\{\mathbf{l} > \ell\} \\
&\le n_\tau + \sum_{\ell=\tau}^{s-1} (n_{\ell+1} - n_\ell) \exp\left(n_\ell \mathscr{M}(a_\ell, p)\right).
\end{aligned} \tag{D.5}$$

This completes the proof of the theorem.

Acknowledgment

This paper is supported in part by NIH/NCI Grants no. 1 P01 CA116676, P30 CA138292-01 and 5 P50 CA128613.

References

[1] S.-C. Chow, J. Shao, and H. Wang, *Sample Size Calculations in Clinical Research*, vol. 20 of *Chapman & Hall/CRC Biostatistics Series*, Chapman & Hall/CRC, Boca Raton, Fla, USA, 2nd edition, 2008.

[2] C. Jennison and B. W. Turnbull, *Group Sequential Methods with Applications to Clinical Trials*, Chapman & Hall/CRC, Boca Raton, Fla, USA, 2000.

[3] Y. S. Chow and H. Robbins, "On the asymptotic theory of fixed-width sequential confidence intervals for the mean," *Annals of Mathematical Statistics*, vol. 36, pp. 457–462, 1965.

[4] B. K. Ghosh and P. K. Sen, *Handbook of Sequential Analysis*, vol. 118 of *Statistics: Textbooks and Monographs*, Marcel Dekker, New York, NY, USA, 1991.

[5] M. Ghosh, N. Mukhopadhyay, and P. K. Sen, *Sequential Estimation*, Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons, New York, NY, USA, 1997.

[6] T. L. Lai, "Sequential analysis: some classical problems and new challenges," *Statistica Sinica*, vol. 11, no. 2, pp. 303–408, 2001.

[7] D. Siegmund, *Sequential Analysis*, Springer Series in Statistics, Springer, New York, NY, USA, 1985.

[8] L. Mendo and J. M. Hernando, "Improved sequential stopping rule for Monte Carlo simulation," *IEEE Transactions on Communications*, vol. 56, no. 11, pp. 1761–1764, 2008.

[9] M. Tanaka, "On a confidence interval of given length for the parameter of the binomial and the Poisson distributions," *Annals of the Institute of Statistical Mathematics*, vol. 13, pp. 201–215, 1961.

[10] S. Franzén, "Fixed length sequential confidence intervals for the probability of response," *Sequential Analysis*, vol. 20, no. 1-2, pp. 45–54, 2001.

[11] S. Franzén, "SPRT fixed length confidence intervals," *Communications in Statistics. Theory and Methods*, vol. 33, no. 2, pp. 305–319, 2004.

[12] A. Wald, *Sequential Analysis*, John Wiley & Sons, New York, NY, USA, 1947.

[13] J. Frey, "Fixed-width sequential confidence intervals for a proportion," *The American Statistician*, vol. 64, no. 3, pp. 242–249, 2010.

[14] X. Chen, "A new framework of multistage estimation," http://arxiv.org/abs/0809.1241v1, September 2008.

[15] X. Chen, "A new framework of multistage estimation," http://arxiv.org/abs/0809.1241v12, April 2009.

[16] X. Chen, "Multistage estimation of bounded-variable means," http://arxiv.org/abs/0809.4679v1, September 2008.

[17] X. Chen, "Estimating the parameters of binomial and Poisson distributions via multistage sampling," http://arxiv.org/abs/0810.0430v1, October 2008.

[18] X. Chen, "Confidence interval for the mean of a bounded random variable and its applications in point estimation," http://arxiv.org/abs/0802.3458v2, April 2009.

[19] X. Chen, "A new framework of multistage parametric inference," in *Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense IX*, usa, April 2010.

[20] A. H. Land and A. G. Doig, "An automatic method of solving discrete programming problems," *Econometrica*, vol. 28, pp. 497–520, 1960.

[21] X. Chen, "A new framework of multistage estimation," http://arxiv.org/abs/0809.1241v16, November 2009.

[22] X. Chen, "A new framework of multistage estimation," http://arxiv.org/abs/0809.1241v4, December 2008.

[23] J. R. Schultz, F. R. Nichol, G. L. Elfring, and S. D. Weed, "Multiple stage procedures for drug screening," *Biometrics*, vol. 29, no. 2, pp. 293–300, 1973.

[24] H. Chen, "The accuracy of approximate intervals for a binomial parameter," *Journal of the American Statistical Association*, vol. 85, no. 410, pp. 514–518, 1990.

[25] E. B. Wilson, "Probable inference, the law of succession, and statistical inference," *Journal of the American Statistical Association*, vol. 22, pp. 209–212, 1927.

[26] C. J. Clopper and E. S. Pearson, "The use of confidence or fiducial limits illustrated in the case of the binomial," *Biometrika*, vol. 26, pp. 404–413, 1934.

[27] W. Feller, *An Introduction to Probability Theory and Its Applications. Vol. I*, John Wiley & Sons, New York, NY, USA, 3rd edition, 1968.

[28] G. S. Fishman, "Confidence intervals for the mean in the bounded case," *Statistics & Probability Letters*, vol. 12, no. 3, pp. 223–227, 1991.

[29] X. Chen, K. Zhou, and J. L. Aravena, "Explicit formula for constructing binomial confidence interval with guaranteed coverage probability," *Communications in Statistics. Theory and Methods*, vol. 37, no. 8–10, pp. 1173–1180, 2008.

[30] X. Chen, "Multistage estimation of bounded-variable means," http://arxiv.org/abs/0809.4679v2, October 2008.

[31] P. Massart, "The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality," *The Annals of Probability*, vol. 18, no. 3, pp. 1269–1283, 1990.

[32] H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *Annals of Mathematical Statistics*, vol. 23, pp. 493–507, 1952.

[33] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *Journal of the American Statistical Association*, vol. 58, pp. 13–30, 1963.

[34] X. Chen, "Exact computation of minimum sample size for estimation of binomial parameters," *Journal of Statistical Planning and Inference*, vol. 141, no. 8, pp. 2622–2632, 2011.

[35] F. J. Anscombe, "Sequential estimation," *Journal of the Royal Statistical Society. Series B*, vol. 15, pp. 1–21, 1953.

Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2012, Article ID 527351, 20 pages
doi:10.1155/2012/527351

Review Article

Methodology and Application of Adaptive and Sequential Approaches in Contemporary Clinical Trials

Zhengjia Chen,[1,2] Yichuan Zhao,[3] Ye Cui,[3] and Jeanne Kowalski[1]

[1] *Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA 30322, USA*
[2] *Winship Cancer Institute, Emory University, 1365-B Clifton Road, Room B4109, Atlanta, GA 30322, USA*
[3] *Department of Mathematics and Statistics, Georgia State University, Atlanta, GA 30303, USA*

Correspondence should be addressed to Zhengjia Chen, zchen38@emory.edu

Received 29 June 2012; Revised 8 October 2012; Accepted 9 October 2012

Academic Editor: Xuelin Huang

The clinical trial, a prospective study to evaluate the effect of interventions in humans under prespecified conditions, is a standard and integral part of modern medicine. Many adaptive and sequential approaches have been proposed for use in clinical trials to allow adaptations or modifications to aspects of a trial after its initiation without undermining the validity and integrity of the trial. The application of adaptive and sequential methods in clinical trials has significantly improved the flexibility, efficiency, therapeutic effect, and validity of trials. To further advance the performance of clinical trials and convey the progress of research on adaptive and sequential methods in clinical trial design, we review significant research that has explored novel adaptive and sequential approaches and their applications in Phase I, II, and III clinical trials and discuss future directions in this field of research.

1. Clinical Trials

Medicine is of paramount importance for human healthcare. Development of novel successful medicines is a lengthy, difficult, and expensive process which consists of laboratory experimentation, animal studies, clinical trials (Phase I, II, and III), and postmarket followup (Phase IV). Clinical trials are FDA-approved studies conducted in human beings to demonstrate the safety and efficacy of new drugs for health interventions under prespecified conditions. A clinical trial is conducted in a sampled small population and the conclusions reached will be applied to a whole target population; therefore, statistics is an indispensable and critical component of clinical trial development and analysis, which

has become increasingly important in contemporary clinical trials. As the gold standard for the evaluation of a new drug, every contemporary clinical trial must be well designed according to its specific purpose and conducted properly under governmental regulations. The major roles of a statistician in a clinical trial are to design an efficient trial with minimum cost and length and maximum therapeutic effect for patients in the trial, and to draw convincing conclusions by applying appropriate cutting edge statistical knowledge. In the past several decades, numerous groundbreaking novel statistical methodologies have been developed and applied to clinical trials and have significantly improved their performance. Consequently, clinical trials have evolved from simple observation studies to hypothesis-driven and well-designed prospective studies. At present, contemporary clinical trials have become the most important part of modern medicine.

2. Adaptive and Sequential Methods

Classical clinical trials are usually designed with a fixed sample size and schedule without using the information obtained from the ongoing trial. However, it has become increasingly common to modify a trial and/or statistical procedures during the conduct of a clinical trial. Specific modifiable procedures include the patient eligibility and evaluation criteria, drug or treatment dosage and schedule, laboratory testing or clinical diagnosis, study endpoints, measurement of clinical response, formulation of study objectives into statistical hypotheses, appropriate study design according to study purpose, calculation of minimum sample size, participant randomization, study monitoring with interim/futility analysis, statistical data analysis plan, and reaching conclusions, and so forth. The purpose of the modification is to improve the performance of a trial with prompt utilization of data accumulating from within the trial as well as upcoming related information from the literature.

Recently, adaptive and sequential clinical trials have become increasingly popular. The sequential method is an approach of frequentist statistics in which data are evaluated sequentially as they are accumulated and a study is monitored sequentially for stopping whenever a conclusion is reached with enough evidence. Adaptive design refers to the modification of aspects of the trial according to data accumulating during the progress of the trial, while preserving the integrity and validity of the trial. The modifiable aspects of adaptive trials include, but are not limited to, (a) sample size, (b) addition or removal of a study arm, (c) dose modification, (d) treatment switch, and so forth [1]. There are two types of adaptive methods in clinical trials, Bayesian and frequentist approaches [2]. The frequentist approach performs the modification of trials while controlling for type I and II errors. The Bayesian approach allows adaption according to the predicted probability. Common characteristics of sequential and adaptive clinical trials are that the trial and/or statistical procedures are modified during the conduct of trial according to the data accumulating during the trial. The sequential method mainly refers to sequentially monitoring the stopping criteria for futility and efficacy, while adaptive methods include modification of many more aspects of the trial as listed above, in addition to the decision of whether to stop the ongoing trial. Considerable novel statistical research has been conducted in the development of sequential and adaptive methods, especially for Phase I and II clinical trials. However, only some of these methods have actually been applied to the daily practice of real clinical trials. In the next 3 sections, we will review significant sequential and adaptive methods that have been applied to Phase I, II, and III clinical trials and have had a high impact on the field of clinical trials.

3. Statistical Methodology of Phase I Clinical Trials

A Phase I trial is one of the most important steps in a drug's development and is the first clinical trial in human subjects after laboratory and animal studies of a therapeutic agent have shown a potential cure effect on the disease. The sample size of a Phase I clinical trial is relatively small and varies in the range of twenty to eighty. It is a widely accepted assumption that the therapeutic effect of a drug depends on its toxicity and increases monotonically with its dosage level. Higher doses are correlated with both severe toxicity and better therapeutic effect. Therefore, a balance is to be achieved between toxicity level and therapeutic benefit. To achieve the best therapeutic benefit, a patient should be treated with the maximum dosage of drug at which the patient can tolerate its associated toxicities with close monitoring. Among all toxicities patients experience, some are so severe that they limit dose escalation. These toxicities are called dose limiting toxicity (DLT). In the National Cancer Institute (NCI) Common Toxicity Criteria, DLT is defined as a group of grade 3 or higher nonhematologic toxicities and grade 4 hematologic nontransient toxicities. The grades of all toxicities are classified as below:

grade 0: no toxicity;

grade 1: mild toxicity;

grade 2: moderate toxicity;

grade 3: severe toxicity;

grade 4: life-threatening toxicity;

grade 5: death.

The main goals of a Phase I trial are to determine the dose-toxicity relationship of a new therapeutic agent and estimate the maximum tolerated dose (MTD) of the agent given the specified tolerable toxicity level. The highest acceptable DLT level is usually defined as a target toxicity level (TTL). It can be said that the TTL determines the MTD of the new therapeutic agent. A careful and thoughtful approach to the design of Phase I trials and accurate MTD estimation are essential for the fate of the new drug in subsequent clinical trials.

In a Phase I clinical trial, the well accepted assumption is that the probability of toxicity increases monotonically with increasing drug dose, although a decrease in the probability of toxicity at high dose levels could happen in some special cases which are not common and not considered here. There are nonparametric and parametric manners to describe the toxicity-dose relationship. In the non-parametric way, the only assumption is that toxicity is nondecreasing with dose. In the parametric description, a distribution with some parameters is adapted to model the toxicity-dose curve. From a biological point of view, the human body has stabilization and self-salvage systems to protect the person from mild toxicity when a drug dose is at a low level below a certain threshold level, but the probability of toxicity increases at an accelerated speed once the stabilization and self-salvage systems have been overcome, and reaches rapidly the worst condition, death, and then levels off. Therefore a sigmoid shape distribution is an appropriate model to describe the relationship between toxicity probability and dose. Many statistical designs have been proposed for Phase I clinical trials; the most commonly used are summarized and compared in Table 1. According to their algorithm, Phase I clinical trial designs can be grouped into two major categories, rule based design and model based design [3].

Table 1: Summary of main Phase I clinical trial designs.

Designs	Advantages	Disadvantages
Standard 3 + 3 design	Robust. Simple. Easy to carry out.	MTD is not a dose with any particular probability of DLT, but in the range from 20% to 25% DLT. Can not estimate MTD with target probability of DLT <20% or >33%. Not all toxicity data of all patients are used to determine the MTD. Many patients are likely to be treated at low doses.
ID isotonic design	Only assumes a monotonically increasing relationship between dose and toxicity. Semiparametric. Can estimate MTD with different TTL (0~100%). Robust and easy to use. Good for combination of multiple drugs and treatments.	The accuracy of MTD may not be as good as CRM or EWOC. The trial efficiency may not be as good as CRM or EWOC.
CRM continual reassessment method	Fit parametric model for dose toxicity relationship. Adaptive optimal design. Accurate estimation of MTD. Improved trial efficiency. Allow flexible MTD with different TTL.	High risk of patients being treated with over toxic dosages. If the parametric model is not reliable, the result could be questionable. May fail to find MTD.
EWOC escalation with overdose control	Includes all advantages of CRM. Controls the overdosing probability. Further improves MTD accuracy and trial efficiency.	If the parametric model is not reliable, the result could be questionable. May fail to find MTD.

3.1. Rule Based Phase I Designs

All rule based designs follow a sequential approach. In rule based designs, a non-decreasing dose toxicity relationship is the only well accepted assumption required. Therefore rule based designs are well suited for first in human clinical trials in which the dose toxicity relationship is not well understood. Common rule based designs include 3 + 3 design [4], isotonic design [5], accelerated titration design [6], and so forth.

The 3 + 3 designs are rule based up-and-down methods used in Phase I protocol templates of the cancer therapy evaluation program (CTEP), whose mission is to improve the lives of cancer patients by sponsoring clinical trials to evaluate new anticancer agents, with a particular emphasis on translational research to elucidate molecular targets and mechanisms of drug effects. While 3 + 3 designs have become standard practice among many Phase I clinical trialists, they are not designed with the intention of producing accurate estimates of a target quantile. Rather they are designed to screen drugs quickly and identify a dose level that does not exhibit too much toxicity in a very small group of patients. These 3 + 3 designs fall into two categories, without dose de-escalation (Figure 1) and with dose de-escalation (Figure 2). In the 3 + 3 design without dose de-escalation, three patients are assigned to the first dose level. If no DLT is observed, the trial proceeds to the next dose level and another cohort of three patients is enrolled. If at least two out of the three patients experience at least one DLT, then the previous dose level is considered as the MTD; otherwise, if only one patient experiences the DLT, then three additional patients are enrolled at the same dose level.

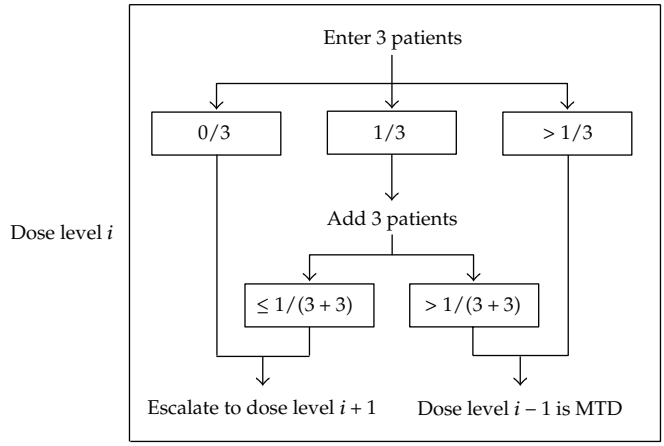


Figure 1: Escalation scheme for 3 + 3 design without dose deescalation (adapted from Lin and Shih [4]).

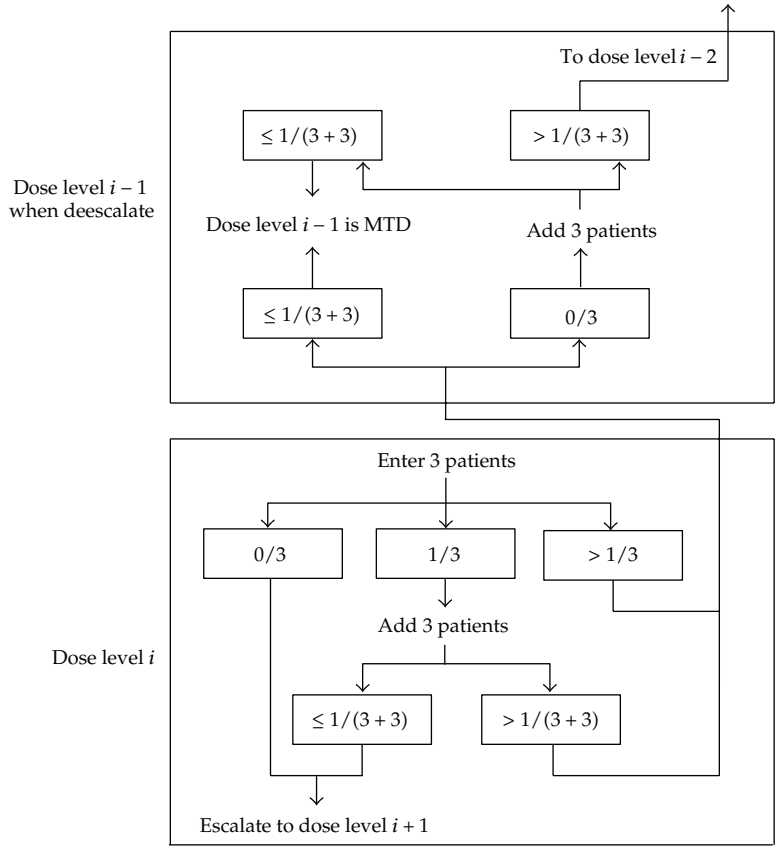


Figure 2: Escalation scheme for 3 + 3 design with dose de-escalation (adapted from Lin and Shih [4]).

If at least one of the three additional patients experiences the DLT, then the previous dose is considered as the MTD; otherwise, the dose will be escalated. The 3 + 3 design with dose de-escalation allows three new patients to be treated at a previous dose level if only three patients were treated at that level previously. Dose reduction continues until a dose level is reached at which six patients are treated and at most one DLT is observed in the six patients. The MTD is defined as the highest dose level at which at most one of six patients experiences DLT, and the immediate higher dose level has at least two patients who experience DLTs. If the first dose is not tolerable, then the MTD cannot be established within the confines of the study. Hence, the MTD is identified from the data and is a statistic rather than a parameter. Storer (1989) was probably the first to examine the characteristics of the 3 + 3 design from the standpoint of the statistician [7]. The operating characteristics of the 3 + 3 design were discussed in Lin and Shih (2001) [4]. Note that any design with sampling that is asymmetric about the MTD will yield a biased result; thus the standard design, and all other designs that approach the MTD from below, will tend to yield a low estimate of the MTD. The 3 + 3 designs are simple and can usually determine a reasonable MTD and are thus the most widely used methods for Phase I clinical trials. But they also have many shortcomings; for example, the methods are not designed around a quantile of interest; not all toxicity data are used to determine the MTD; the MTD is not a dose with any particular probability of toxicity. These disadvantages led to the exploration of extended isotonic design for Phase I clinical trials.

Leung and Wang (2001), for the first time, introduced a semiparametric Phase I design called isotonic design in which only a non-decreasing dose toxicity relationship is the required assumption [5]. In their isotonic design, the pool-adjacent-violators algorithm (PAVA) and isotonic regression are used to update the probability of DLT of each dose level after the toxicity response of each newly treated cohort has been obtained. The dose allocation rationale is to treat each new cohort at a dose level with an estimated probability of DLT closer to the pre-specified target acceptable toxicity level. The trial stops when the same dose has been tested consecutively for a certain number of cohorts or a maximum number of patients have been treated. The recommended dose level for the next cohort based on all completed data after the trial stops is the MTD. Through simulation studies, the isotonic design was demonstrated to perform substantially better than the 3 + 3 design and comparably to the continual reassessment method (CRM) [8], Storer's up-and-down designs, and escalation with overdose control (EWOC) design [9]. Moreover, the isotonic design is model-free and especially appropriate in cases where the parametric dosetoxicity relationship is not well understood.

There are many other rule based designs. All rule based designs can estimate a reasonable MTD using a stopping rule based either on observed DLTs or on convergence criteria. Ad hoc additional dose levels can also be added when needed without any impact on their robustness. Most rule-based designs are practically simple and easy to implement. At present, 3 + 3 designs are still the most popular in Phase I clinical trials.

3.2. Model Based Designs

In model based designs, three parametric dose-toxicity functions (logistic model, hyperbolic model, and power function) are usually employed to depict the relationship between dose and toxicity. Model based designs often fail to find an MTD in first in human studies that are based on observed DLTs. The most common model based designs are CRM and EWOC. Their algorithms are illustrated in Figure 3.

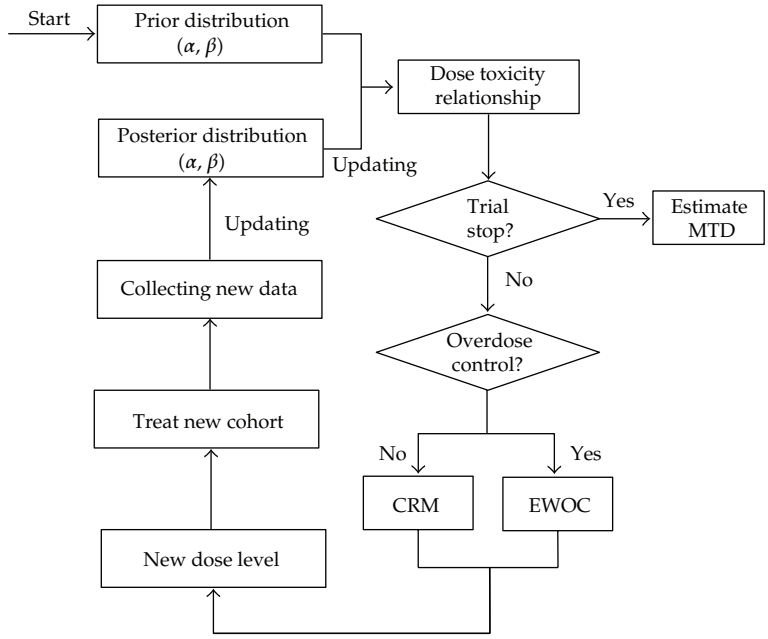


Figure 3: Diagram of model based phase I designs: continuous reassessment design (CRM) and escalation with overdose control (EWOC).

O'Quigley et al. (1990) originally introduced the CRM, a Bayesian approach to fully and efficiently use all data and prior information available in a Phase I study [8]. As in rule based designs, a TTL is specified and the goal is to estimate the dose associated with the TTL, Γ. A parametric model depicting the dose toxicity relationship and a *prior* distribution for each unknown parameter of the model are required to implement CRM. The posterior mean of each parameter is computed using the prior for the parameter and all available toxicity data for the probability of toxicity, P_{DLT}, of each dose level. The computation is conducted and P_{DLT} of each dose level is updated with accumulative toxicity data available when a new patient is recruited. The main idea of CRM is to treat each patient at the dose level with P_{DLT} closest to Γ. The MTD is defined as the dose level of the last patient treated in the trial. In the originally proposed CRM, a one parameter model of dose toxicity function and a single patient cohort are used. Furthermore, the first patient is proposed to be treated at a dose level determined purely by a guess in the original CRM, which makes the method impractical. Therefore, Korn et al. (1994) proposed a modified CRM in which the trial starts at the lowest dose level, no dose level can be skipped during the dose escalation, and the trial stops when the same dose has been recommended for a new patient consecutively for a fixed number of times [10]. However, patients still may be treated at excessively toxic doses in the modified CRM because of its single patient per cohort and the length of study is still very long because of the restriction that the toxicity of all treated patients must be obtained to calculate the new dose level for the next patient. In addition to the modification of Korn at al. (1994) [10], Faries (1994) [11], in his modified CRM, added another rule that no dose escalation is allowed for the next patient when the last patient has DLT. This rule can avoid treating patients at overly toxic doses compared with the traditional 3 + 3 design. In order to address the ethical requirement that the probability of a patient being treated at overdose

is under a pre-specified value, Babb et al. (1998) introduced an adaptive dose escalation scheme called EWOC [9]. The constraint on overdosing of EWOC is a superior feature over the CRM and its theoretical foundation was further elaborated by Zacks et al. (1998) [12]. A two-parameter model logit $(P_{\mathrm{DLT}}(x_i)) = \alpha + \beta x_i$ was first used to depict the dose, x_i, and DLT relationship and then the joint posterior for α and β was transformed to a joint posterior for the MTD and the probability of DLT at the lowest dose level, ρ_0. EWOC is also designed to rapidly approach the MTD in addition to the overdose constraint so that it starts from the lowest dose level and a single patient per cohort is used. After the toxicity response of the last enrolled patient has been obtained, the joint posterior for the MTD and ρ_0 is updated using all the available information and the next coming patient is treated at the 25th percentile of the marginal posterior for the MTD. The trial stops after a fixed number of patients have been treated and then the MTD is computed as its posterior mean or estimated by minimizing the posterior expected loss in a loss function. In order to be safe and shorten the length of the trial, no dose level can be skipped during the dose escalation procedure and multiple patient cohorts can be used instead in EWOC. Through simulation studies, EWOC has been shown to be effective in overdose control and have comparable accuracy of estimated MTD as CRM. Fewer patients are treated at nonoptimal dose levels, resulting in less DLT, and the estimated MTD has smaller average bias and mean squared error in EWOC than in some other nonparametric designs, such as four up-and-down designs and two stochastic approximation methods [9]. It seems that EWOC is a promising alternative design for Phase I clinical trials, especially when the ethical and safety requirement of overdose control is a particular concern. Both CRM and EWOC belong to adaptive dose finding designs in which a Bayesian approach is usually employed and the dose level for the new incoming cohort is adaptive based on the toxicity responses of the previously treated patients in the ongoing trial. Another adaptive dose design is the nonparametric adaptive urn design approach for estimating a dose-response curve [13].

All ruled based designs are robust and simple to implement and usually give a reasonable MTD under certain rules. Applying some sort of models, such as isotonic regression, to data can improve the accuracy of the MTD. Model based designs require a parametric model of dose toxicity relationship and may greatly improve the probability of estimating the correct MTD compared with rule based designs when certain assumptions are satisfied. However, model based designs are not robust and should not be used unless their underlying assumptions can be met with confidence. The accuracy of the estimated MTD depends substantially on the number of observed DLTs, and the sample size is also an important factor. Overall, different designs, whether rule based or model based, usually perform similarly when they are similar in sample size and aggressiveness. Thus, simple designs, especially standard designs, are still very popular in Phase I clinical trial practices.

The design of Phase I clinical trials can involve one or two stages. Rule based or model based designs can be implemented in each stage of two stage designs. There are other critical issues in Phase I clinical trial designs, such as the operating characteristics of 3 + 3 design in terms of expected toxicity level [14], two or multiple stage Phase I design, within-patient dose escalation, late toxicity, combination of multiple agents, balance between toxicity and efficacy, individual MTD, fully utilization of all toxicities [15, 16], and so forth. Some outstanding research studies have been conducted on these topics, which will not be elaborated on herein due to space constraints but have been described in several comprehensive review articles [3, 17–19].

4. Statistical Methodology of Phase II Clinical Trials

After the safety and MTD of an experimental drug have been established in a Phase I clinical trial, the drug will enter Phase II clinical trials, which initially evaluate the drug's therapeutic effects at the recommended MTD. Phase II trials are sometimes further classified as Phase IIa and IIb studies. Phase IIa trials screen the promising novel experimental agent for significant antidisease activity and Phase IIb trials focus on the drug's improved therapeutic effectiveness over the standard treatment. Phase II studies provide critical information to decide whether further testing of the experimental drug in a large confirmatory Phase III trial is warranted. The surrogate endpoint used in Phase II clinical trials needs to be obtained in a short time and should be able to assess the treatment's primary benefit. For cancer trials, the experimental drug's antitumor activity and progression-free survival (PFS) of treated patients are often used as surrogates of the drug's efficacy. The drug's anti-tumor activity is measured as clinical response within a short period of time following the treatment and is classified as complete response (CR), partial response (PR), progressive disease (PD), or stable disease (SD). PFS, which is estimated as the time elapsed from the date of treatment to the date of adverse event, resembles the outcome (overall survival) of the following Phase III clinical trial and is also widely used when it can be measured in a short time.

4.1. Single Arm Phase II Designs

The most commonly used Phase II clinical trial designs are summarized in Table 2. Phase II trials can involve either a single arm, which compares the new treatment with the standard response rate reported by historical data, or two or more arms with patients randomized among different treatments. In a single arm Phase II trial, two or multistage designs may be used to improve the trial efficiency and save resources with early termination of a futile trial. The interim analysis between the consecutive stages examines the accumulated data and decides whether the trial should stop as suggested by the early evidence of futility or should continue to next stage. The earliest two stage Phase II design was proposed by Gehan et al. in 1961 [20], in which a trial is terminated for futility when no patients enrolled in the first stage show any response or continues with the second stage, enrolling an additional number of patients to estimate a more accurate response rate with additional patient data. This design provides interim monitoring and can rule out ineffective drug with minimized sample size. This design is only appropriate for binary outcomes, which differ from the overall survival endpoint used in the following Phase III trial. Moreover, this design has no statistical testing on agents showing some promise and is not optimized. Therefore, Simon (1989) proposed an optimized two stage Phase II design by controlling both type I and type II errors as well as optimizing the sample sizes in both stages [21]. This design can quickly screen out agents without effectiveness while testing further agents with some promise. The design has two subtypes, optimal and minimax. The optimal subtype minimizes the expected overall sample size with the probability of the trial stopping after only the first stage so that it is appropriate for experimental drugs with a high probability of failure after the first stage. The minimax subtype minimizes the maximum possible sample size when the trial stops after completion of two stages so that it is better for highly promising experimental drugs. As with Gehan's design, Simon's two stage designs are only appropriate for binary outcomes. Other investigators have further proposed to conduct multiple interim analyses in Phase II clinical trials by using multistages. For example, Fleming (1982) [22] and Chang et al. (1987)

Table 2: Summary of main Phase II clinical trial designs.

Designs	Advantages	Disadvantages
One stage one arm design	Compare with historical control. Smallest sample size. Simple.	Delay the evaluation of effectiveness. Historical control may not be valid. Subject to population differences, time trends, evaluation bias, and so forth.
Gehan's two stage design	With interim monitoring. Rule out ineffective drug with minimized sample size.	No testing on agents showing some promise. Only suitable for binary outcome. The endpoint is different from that in following Phase III trial.
Simon's two stage design	The samples in two stages are optimized. Quickly screen out agents without effectiveness while testing further agents with some promise. Two choices: optimal versus minimax	Only suitable for binary outcome. The endpoint is different from that in the following Phase III trial.
Bayesian Phase II design	Flexible monitoring schedule. More efficient and robust.	Intensive computation. Relies heavily on statistician during trial.
Randomized Phase II design	Use of randomization. Reliable control and less bias. More similar to Phase III trial.	Sample size increases. Length of trial increases. Cost increases.
Phase II pick the winner design	Efficient and effective way of comparing two or multiple experimental regimens. Each experimental regimen compared with historical controls.	Not appropriate for comparison of adding an experimental agent to standard regimen.
Phase II screening design	Limits the sample size required for a randomized Phase II comparison. Good for comparison of the addition of an experimental agent to standard regimen.	No statistical comparison between the selected arms.
Phase II randomized discontinuation design	Good when significant continued benefit after initial benefit implies significant benefit overall, and vice versa, or when benefit is restricted to a nonidentifiable subgroup of patients.	May need a large number of patients treated at a treatment not effective for them.
Phase II/III design	Use of Phase II data in Phase III trial. Minimize delay in starting up Phase III study. Use of concurrent control. Useful for new drugs showing efficacy.	Large sample sizes. Needs Phase III infrastructure developed even if it stops early.

[23] studied multiple testing and group sequential methods for Phase II trial designs. But the issue of inflating overall type I error needs to be considered in these kinds of Phase II designs.

Among the single arm Phase II designs, another major group is Bayesian Phase II design. For example, Thall and Simon (1994) [24] proposed a Bayesian Phase II design which continuously examines the results after each new enrolled patient and determines whether the trial can stop with a solid decision on the efficacy of the experimental drug or should continue to enroll more patients and obtain enough data for making a decision. Lee and Liu (2008) [25] proposed a Bayesian approach called predictive probability Phase II design. This novel Bayesian design provides a flexible monitoring schedule for Phase II clinical trials which becomes more efficient and robust, but at the cost of intensive computation, and relies heavily on the statistician during the trial. Yin et al. (2011) further coupled the methods of

predictive probability monitoring and adaptive randomization in a randomized Phase II trial and extensively compared this hybrid Bayesian approach with group sequential methods [26].

4.2. Two or More Arm Phase II Designs

Some Phase II clinical trials may have two arms and randomization is frequently used to generate a reliable concurrent control arm and reduce biases. This kind of randomized Phase II trial is more similar to a Phase III trial. Randomized Phase II trials may reduce the so-called trial effect which often arises due to different patient populations, physician preferences, and medical environments between current and previous studies. But the sample size, trial length, and cost increase about 4-fold.

There are several multiple arm Phase II designs [27]. The Phase II "pick the winner" design is one in which each experimental regimen is compared with a historical control. No formal statistical comparison between groups is conducted and the simple winner of the all arms is the winner of the trial. This design provides an efficient and effective way of comparing two or multiple experimental regimens but is not appropriate for the comparison of adding an experimental agent to a standard regimen.

Phase II screening design is another Phase II design with multiple arms in which all experimental arms are compared with the standard treatment arm and all the experimental arms beating the standard treatment arm are winners. Therefore this design limits the sample size required for a randomized Phase II comparison and it is appropriate for testing the effect of adding an experimental agent to a standard regimen. However, it provides no statistical comparison between the selected (winning) arms.

Some investigators have proposed a novel Phase II randomized discontinuation design in which all patients receive the same treatment for a period of time and those with stable disease are randomized to continue or discontinue. This design is particularly appropriate when the treatment is known to have better therapeutic effects and it is ethical for all participants to benefit from it, or when the potential subgroup of patients who can benefit from the treatment is unknown before receiving it. However, this design requires a large number of patients to be treated with a treatment not effective for them. Therefore this design has specific applications but is not widely used.

Conventionally, Phase II and III trials are conducted separately in a sequential order and only an experimental drug that has successfully passed a Phase II trial can enter a Phase III trial. The resulting gap between trials and time lag may be unnecessary under certain circumstances. Therefore, a seamless Phase II/III design has been proposed, which uses Phase II data in a Phase III trial and minimizes delay in starting up the Phase III study [28, 29]. Usually the Phase II part is a randomized Phase II trial which uses a concurrent control. This nonstop Phase II/III design is particularly useful for new drugs showing efficacy. It usually requires large sample sizes and requires a Phase III infrastructure to be developed even if it stops early.

4.3. Other Advanced Topics in Phase II Designs

Categorical tumor response has been the most common endpoint in the Phase II clinical trial designs. However, from a statistical standpoint, categorizing a continuous tumor change percentage into a categorical tumor response with 4 levels results in a loss of study power by

not fully utilizing all available data. Several publications have studied extensively the direct utilization of continuous tumor shrinkage as the primary endpoint for the measurement of drug efficacy in Phase II clinical trials [30–32]. The success rate of Phase III oncology trials remains very low (e.g., 50–60%) despite the success demonstrated in the preceding Phase II trials [30]. The relationship between tumor response/tumor shrinkage percentage and overall survival as the gold standard for drug efficacy has been revisited [33]. PFS has the advantage of short follow-up time [34] and has been confirmed as the best estimate of overall survival [35] so that PFS is recommended as the primary endpoint over categorical tumor response in Phase II clinical trials when feasible.

5. Statistical Methodology of Phase III Clinical Trials

If an experimental agent exhibits adequate short term therapeutic effects in a Phase II trial, the drug will be moved forward to a Phase III study for confirmative testing of its long term effectiveness. The typical endpoint in a Phase III trial is a time to event measurement, such as progression free survival or overall survival. Phase III trials are large scale in terms of sample size, resources, efforts, and costs. This Phase collects a large amount of data over a long period of followup to evaluate the ultimate therapeutic effect of a new drug. The design of Phase III clinical trials has become a very important research field in order to improve the performance of these critical clinical trials. The most commonly used Phase III clinical trial designs are summarized in Table 3.

5.1. Randomization

The earliest design of Phase III clinical trials is a single arm study design using historical controls from the literature, existing databases, or medical charts. This kind of Phase III design allows ethical consideration and can increase enrollment as patients are assured of receiving new therapy. In addition, trials will have shorter time and lower cost, making this type of trial a good choice for the initial testing of new treatments, or when disease diagnosis is clearly established, prognosis is well known, or the disease is highly fatal. This Phase III design, however, provides no comparison to control group data and is vulnerable to biases because disease and mortality rates have changed over time and literature controls are particularly poor. Phase III trials conducted using this design tend to exaggerate the value of a new treatment. In order to avoid bias and eliminate time trends, a concurrent control but nonrandomized design for Phase III clinical trials was then proposed and implemented. In this design, randomization does not interfere with treatment selection. It is easier to select a group to receive the intervention and select the controls matching key characteristics. Therefore, this design can reduce costs and is relatively simple and easily acceptable to both the investigator and participant. But in this Phase III design, intervention and control groups may not be comparable because of selection bias and incomparable different group populations. It is difficult to prove comparability because it is impractical to have information on all important prognostic factors and to match several factors. The existence of unknown or unmeasured factors in large studies is also uncertain. The afterward covariance analysis is not adequate for offsetting the imbalance between groups.

To eliminate the bias, facilitate masking treatments, and permit the use of statistical theory, randomization has been employed widely in the Phase III clinical trials [36]. There are two major types of randomization approaches, non adaptive versus adaptive.

Table 3: Summary of main Phase III clinical trial designs.

Designs	Advantages	Disadvantages
Historical control (literature and existing databases or medical charts)	Allows ethical consideration. Increase enrollment as patients are assured of receiving new therapy. Shorter time and less cost. Good for initial testing of new treatments, when disease diagnosis is clearly established, prognosis is well known, or disease is highly fatal.	Vulnerable to bias. Disease rate and mortality rate have changed over time. No comparison to control group data. Literature controls particularly poor. Tends to exaggerate the value of a new treatment.
Concurrent control, not randomized	Eliminates time trends. Data of comparable quality. Randomization does not interfere with treatment selection. Easier to select a group to receive the intervention and select the controls matching key characteristics. Reduced cost, relative simplicity, investigator and participant acceptance.	Intervention and control groups may not be comparable because of selection bias and different treatment groups are not comparable. Difficult to prove comparability because of the need for information on all important prognostic factors and matching several factors is impractical. Uncertainty about unknown or unmeasured factors exists even for large studies. Covariance analysis not adequate.
Randomized clinical trials (RCT)	Considered to be "gold standard". Removes potential bias in group allocation. Randomization and concurrent control produce comparable groups. Guarantees the validity of statistical tests and valid comparison. General use.	Subjects may not represent general patient population. Increased sample size and cost. Acceptability of randomization process. Administrative complexity.
Sequential RCT design	Continues to randomize subjects until null hypothesis is either rejected or "accepted." Good for acute response, paired subjects, and continuous testing. Good for one-time dichotomous decisions such as regulatory approval, and so forth.	Multiple testing inflates type I error. Inhibits adaptation due to the requirement of prespecifying all possible study outcomes.
Bayesian RCT design	Dynamic learning adaptive feature. Incorporates external evidence. Add new interventions and drop less effective ones without restarting trial. Improves timeliness and clinical relevance of trial results. Lowest sample size and cost.	May be criticized as too subjective, not well planned, or too complicated.

Simple randomization, block randomization, and stratified randomization belong to the nonadaptive randomization type. The simple randomization is robust against both selection and accidental biases and appropriate for RCTs with over 200 subjects because of the possibility of imbalanced group sizes in small RCTs [37]. Block randomization can guarantee balanced group sizes by pre-specifying the block size and allocation ratio and allocating subjects randomly within each block [33]. Block randomization is often used with "stratified randomization" in small RCTs. There are several adaptive randomization approaches:

adaptive biased coin, covariate adaptive, and response adaptive [33]. The adaptive biased-coin randomization method can reduce the imbalance of group size and is less affected by selection bias than permuted-block randomization by decreasing and increasing the probability of being assigned to an overrepresented group and underrepresented group, respectively. Randomization can be adaptive to covariate in order to produce balanced groups in terms of the sample size of several covariates. The most common covariate adaptive randomization approaches are the Taves's method [38], Pocock and Simon method [39], and Frane's method [40] for both continuous and categorical types. Overall, covariate adaptive randomization can reduce the imbalance further and handle more covariates simultaneously than using the combination of block and stratified randomization [41]. Randomization can be adaptive to response or outcome in order to increase the trial therapeutic effect, taking into account ethical considerations. Response-adaptive randomization can assign more patients to receive better treatment by skewing the probability of assigning new patients to the group showing favorable response as the data of the trial are accumulating while maintaining a certain study power [41]. The most common approaches used for response-adaptive randomization are the urn model, biased coin design, and Bayesian's approach [34]. Each randomization approach has its own merits and limitations. The selection of randomization method depends on the specific study purpose.

5.2. Randomized Controlled Phase III Trials

The statistical approach of randomization removes any potential bias in group allocation. The use of randomization and a concurrent control together produce comparable groups and make conclusions more convincing. The use of feasible blinding minimizes the bias after randomization. At present, the standard form of a Phase III trial is a randomized and placebo-controlled clinical trial (RCT) with double blinds. The control arm may be a placebo or the standard of care. The use of placebo is only acceptable if there is no other better or standard therapy available. Interim monitoring is also often considered for a long term confirmatory RCT. The RCT which guarantees the validity of statistical tests and valid comparisons has been generally used as the "gold standard" for verifying the efficacy of new drugs. However, there are still some limitations in RCTs; for example, subjects may not represent the general patient population; sample size and cost increase substantially; the randomization process may not be widely accepted; the administrative process may be complex; and so forth. According to their statistical algorithm and characteristics, besides the conventional fixed sample Phase III clinical trial in which only one final data analysis is conducted at the end of the study, other RCT designs with additional analyses before final analysis can be divided into two distinct categories: sequential RCT design and Bayesian adaptive RCT design.

5.2.1. Group Sequential RCT Design

The scheme of the group sequential design is summarized in Figure 4. In this design, type I and II errors are explicitly controlled while testing the study hypotheses, and patients continue to be enrolled and randomized until the primary hypothesis has been proved or disproved. To design a Phase III clinical trial with the group sequential method, the total number of stages, the sample size, and stopping criterion at each stage for the null hypothesis testing as well as the usual specifications in a conventional Phase III clinical trial must be pre-specified before the trial starts. At each interim stage, all accumulated data up to the point are

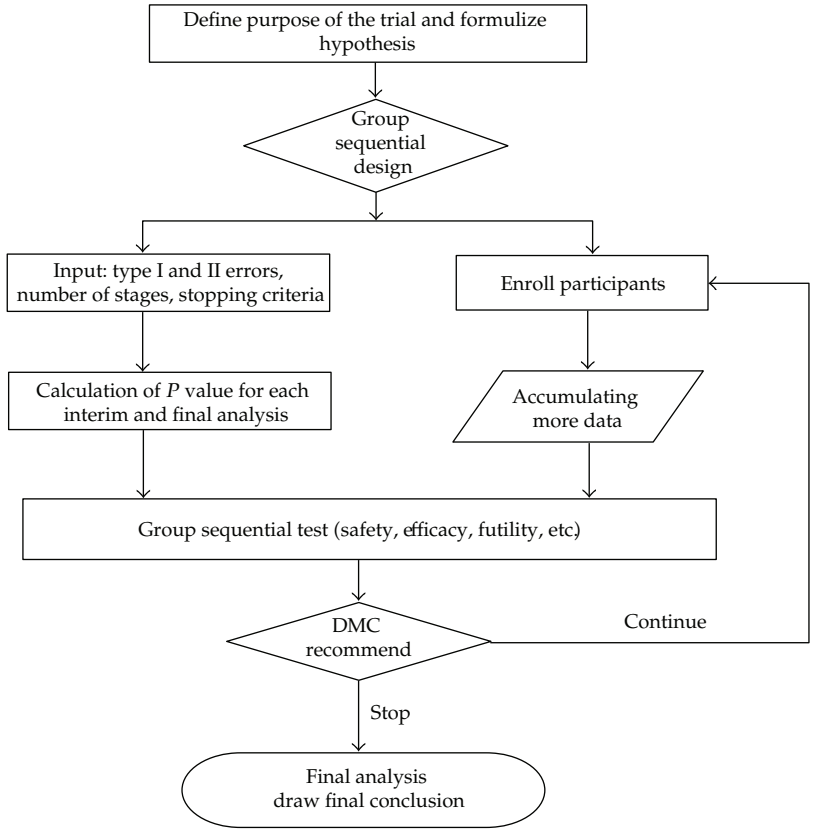


Figure 4: Diagram of group sequential design.

analyzed and the test statistics is compared with critical values generated from the sequential design to determine whether the trial should stop or continue. A conclusion on the primary hypothesis must be reached at the final stage when the sequential trial passes all interim analyses and completes with the final stage.

Multiple testing during the sequential trial may inflate type I error which can be controlled using the Pocock approach [42], O'Brien-Fleming approach [43], and alpha spending function [44]. The Pocock approach was the first method for group sequential testing with given overall type 1 error and power by dividing type I error evenly across the number of interim and final analyses. For example, in a clinical trial with 2 interim analyses and 1 final analysis, the Pocock procedure uses the same cut-off for both the interim and final analyses and the clinical trial can stop and claim a positive outcome if the P value is less than 0.022 at any of the analysis times. One obvious problem with the Pocock approach is its too high probability of stopping the trial early. In order to prevent early stopping and to keep the final P value close to the overall significance level, such as 0.05, O'Brien and Fleming's approach [43] uses a very strict cut-off P value at the beginning, then relaxes the cut-off P value over time. As in the above clinical trial, the P values for the first and second interim analyses are 0.005 and 0.014, respectively. The P value for the final analysis is 0.045 which is close to 0.05. Both the Pocock and O'Brien-Fleming approaches maintain the overall type I error by paying a penalty at the final analysis, but the O'Brien-Fleming

method involves much less of a penalty at the planned conclusion of the study because it requires stricter standards earlier. Both methods have some limitations; both require a pre-specified maximum number of patients, the number of interim analysis, and equal increments of information between interim stages. Therefore, DeMets and Lan [44] (1994) introduced a spending function approach to relax the requirement of the equal increments of information. The approach spends the allowable type I error rate over time according to a chosen spending principle and the amount of information accrued and allows dropping or adding an interim analysis during conduct of the trial. There are several types of spending functions proposed in the literature. Besides the Pocock-type and O'Brien-Fleming-type error spending functions proposed by Lan and DeMets, the gamma error spending function [45] proposed by Hwang, Shih, and DeCani and the power error spending function [46] proposed by Jennison and Turnbull are also commonly used in clinical trials. The conclusions drawn at the interim and final analyses are affected heavily by the pre-specified boundaries so that the choices of the type of spending function are very important and depend on the specific purpose of the trial and its associated clinical program. In addition to efficacy, the safety profile of drug is also an important factor when considering the early stopping of a trial.

The major advantages of the group sequential RCT design are its abilities to prevent unnecessary exposure of patients to an unsafe or ineffective new drug or to a placebo treatment, and to save time and resources by stopping the trial early for efficacy, futility, and safety. The sequential RCT design is suitable for acute response, paired subjects, and continuous testing. It is especially appropriate for dichotomized decisions (yes/no) because the result of the RCT trial is determined to be significant or not according to a pre-specified significance level (type I error). Although sequential RCT is the most widely used design in Phase III clinical trials, it has some limitations. Sequential RCT may require larger sample sizes than Bayesian adaptive RCT as a result of additional variability and comparison of multiple treatments with similar efficacies. Sequential RCT is somewhat adaptive by using interim monitoring and stopping rules, but it requires prespecification of all possible study outcomes, thus inhibiting the full adaptation and utilization of newly accumulated data from the ongoing trial.

5.2.2. Bayesian RCT Design

Bayesian randomized clinical trials refer to trials in which Bayesian approaches are applied extensively to some or all of the processes of a trial including randomization, monitoring, interim and futility analysis, final analysis, and adaptive decisions. Berry and Kadane [47] proposed optimal Bayesian randomization in 1997 and the practical uses of Bayesian adaptive randomization in clinical trials have been reviewed by Thall and Wathen [48]. Bayesian monitoring has been frequently used in some Phase III clinical trials, especially in those with failure time endpoints [49]. Bayesian analysis in clinical trials has become increasingly common recently as it can borrow strength from outside the study [50]. Bayesian adaptive decisions in clinical trials can be made according to a posterior probability or predictive probability of trial success or from the result of Bayesian final analysis. Bayesian adaptive decisions have been compared to frequentist sequential approaches [51] and some studies [52–54] proposed to use Bayesian decision theoretical approaches in the optimization of designs under various settings.

Bayesian RCT design is dynamic learning adaptive in nature as it prespecifies the approaches to combine all available data accumulated during the process of the study,

calculate probabilistic estimation of uncertainty, control the probability of false-positive and false-negative conclusions, and change the study design correspondingly [55]. Bayesian and adaptive RCT design cannot only compare multiple active treatments but can also allow the ongoing trial to add new emerging effective interventions, discontinue less effective ones proved by accumulated within-trial data, or focus on patient subgroups identified by certain biomarkers for whom interventions are more (or less) effective so that the trial tests the most current interventions, improves the clinical relevance, and targets biomarkers that predict response to alternative intervention. Using external existing data from previous studies during the design stage and the accumulated within-trial data to update the design results in smaller sample size, shorter time, and reduced cost of Bayesian and adaptive RCT [56]. But Bayesian RCT may be criticized as being too subjective, not well planned, or too complicated.

Both Bayesian and sequential RCT designs have their advantages and disadvantages. Instead of biasing toward either Bayesian or sequential methods, statisticians and investigators should choose the design of Phase III clinical trial that best fits the goals of the trial and is most likely to provide the best performance.

5.2.3. Adaptive Sample Size Calculation and Adaptive Stopping

In the planning stage of a Phase III clinical trial, sample size is one of the most important factors to be considered because the budget for the trial depends on the minimum required sample size. Usually sample size is fixed in a trial, but an adaptive sample size calculation is often used in adaptive clinical trials and the sample size is adjusted based on the observed data at the interim analysis [1]. Sample size determination depends on the expected treatment difference and its standard deviation; however, their initial estimations often turn out to be too large or small as suggested by the accumulating data from the ongoing trial or other newly completed studies. In this case, keeping the original sample size will lead to an underpowered or overpowered trial, and so the sample size should be adjusted according to the updated effect size for the ongoing trial. There are several approaches for sample size adjustment based on the criteria of treatment effect size, conditional power, and/or reproducibility probability [57–61]. The observed treatment effect and estimated standard deviation from a limited number of subjects at the interim analysis may not be of statistical significance. Therefore, these factors should not be weighed too heavily and the targeted clinically meaningful difference in the ongoing clinical trial should always be considered fully in the adaptive sample size calculation.

The fate of an ongoing Phase III trial is determined at its data monitoring committee (DMC) meeting, which makes recommendations based on the available data according to stopping rules in the statistical guidelines. The common factors considered in stopping rules are safety, efficacy, futility, benefit-risk ratio, weight between the short term and long term treatment effects, and conditional power or predictive power [1]. Current tools for monitoring Phase III trials are stopping boundaries, conditional and predictive powers, futility index, repeated confidence interval, and Bayesian monitoring tools. Even though the stopping rules are usually stipulated in the design stage, adaptive stopping is becoming more and more common due to unpredicted events during the conduct of the trial, such as a change in the DMC meeting date because of unavailability of committee members, different patient accrual progress, and deviation in the analysis schedule. Moreover, the true variability in the parameters to construct these boundaries of stopping rules is never known and it is very common that the initial estimates of the variability and treatment effect in

the design phase are inaccurate as shown by the preliminary results of the ongoing trials. These deviations could affect substantially the stopping boundaries so that adaptive stopping becomes especially desirable in these cases. To stop a trial prematurely under adaptive stopping algorithm, thresholds for the number of subjects randomized and some rules (such as utility rules, futility rules, etc.) in terms of boundaries must pass.

6. Concluding Remarks

Clinical trials remain an indispensable component of new drug development. Novel statistical approaches have been applied to clinical trials and have significantly improved their performance in every step from design, conduct, and monitoring to data analysis and drawing final conclusions. As modern medicine progresses, increasingly complex requirements and factors need to be considered in clinical trials, which in turn create new challenges for statisticians. In the future, more novel statistical approaches, frequentist and Bayesian, should be developed to enhance the performance of clinical trials in terms of therapeutic effect, safety, accuracy, efficiency, simplicity, and validity of conclusions and to expedite the development of effective new drugs to improve human healthcare.

Acknowledgments

This work is supported in part by NIH/NCI Grants no. 1 P01 CA116676 (Z. Chen.), P30 CA138292-01 (Z. Chen. and J. Kowalski.), and 5 P50 CA128613 (Z. Chen); NSA Grant H98230-12-1-0209 (Y. Zhao).

References

[1] S. C. Chow and M. Chang, "Adaptive design methods in clinical trials—a review," *Orphanet Journal of Rare Diseases*, vol. 3, no. 1, article 11, 2008.
[2] P. Gallo, C. Chuang-Stein, V. Dragalin, B. Gaydos, M. Krams, and J. Pinheiro, "Adaptive designs in clinical drug development—an executive summary of the PhRMA working group," *Journal of Biopharmaceutical Statistics*, vol. 16, no. 3, pp. 275–283, 2006.
[3] D. M. Potter, "Phase I studies of chemotherapeutic agents in cancer patients: a review of the designs," *Journal of Biopharmaceutical Statistics*, vol. 16, no. 5, pp. 579–604, 2006.
[4] Y. Lin and W. J. Shih, "Statistical properties of the traditional algorithm-based designs for phase I cancer clinical trials," *Biostatistics*, vol. 2, no. 2, pp. 203–215, 2001.
[5] D. H. Y. Leung and Y. G. Wang, "Isotonic designs for phase I trials," *Controlled Clinical Trials*, vol. 22, no. 2, pp. 126–138, 2001.
[6] R. Simon, B. Freidlin, L. Rubinstein, S. G. Arbuck, J. Collins, and M. C. Christian, "Accelerated titration designs for phase I clinical trials in oncology," *Journal of the National Cancer Institute*, vol. 89, no. 15, pp. 1138–1147, 1997.
[7] B. E. Storer, "Design and analysis of phase I clinical trials," *Biometrics*, vol. 45, no. 3, pp. 925–937, 1989.
[8] J. O'Quigley, M. Pepe, and L. Fisher, "Continual reassessment method: a practical design for phase 1 clinical trials in cancer," *Biometrics*, vol. 46, no. 1, pp. 33–48, 1990.
[9] J. Babb, A. Rogatko, and S. Zacks, "Cancer phase I clinical trials: efficient dose escalation with overdose control," *Statistics in Medicine*, vol. 17, no. 10, pp. 1103–1120, 1998.
[10] E. L. Korn, D. Midthune, T. Timothy Chen et al., "A comparison of two phase I trial designs," *Statistics in Medicine*, vol. 13, no. 18, pp. 1799–1806, 1994.
[11] D. Faries, "Practical modifications of the continual reassessment method for phase I cancer clinical trials," *Journal of Biopharmaceutical Statistics*, vol. 4, no. 2, pp. 147–164, 1994.
[12] S. Zacks, A. Rogatko, and J. Babb, "Optimal Bayesian-feasible dose escalation for cancer phase I trials," *Statistics & Probability Letters*, vol. 38, no. 3, pp. 215–220, 1998.

[13] R. Mugno, W. Zhus, and W. F. Rosenberger, "Adaptive urn designs for estimating several percentiles of a dose-response curve," *Statistics in Medicine*, vol. 23, no. 13, pp. 2137–2150, 2004.
[14] Z. Chen, M. D. Krailo, J. Sun, and S. P. Azen, "Range and trend of expected toxicity level (ETL) in standard A + B designs: a report from the children's oncology group," *Contemporary Clinical Trials*, vol. 30, no. 2, pp. 123–128, 2009.
[15] Z. Chen, M. D. Krailo, S. P. Azen, and M. Tighiouart, "A novel toxicity scoring system treating toxicity response as a quasi-continuous variable in phase I clinical trials," *Contemporary Clinical Trials*, vol. 31, no. 5, pp. 473–482, 2010.
[16] Z. Chen, M. Tighiouart, and J. Kowalski, "Dose escalation with overdose control using a quasi-continuous toxicity score in cancer phase I clinical trials," *Contemporary Clinical Trials*, vol. 33, no. 5, pp. 949–958, 2012.
[17] M. Tighiouart and A. Rogatko, "Dose finding with escalation with overdose control (EWOC) in cancer clinical trials," *Statistical Science*, vol. 25, no. 2, pp. 217–226, 2010.
[18] W. F. Rosenberger and L. M. Haines, "Competing designs for phase I clinical trials: a review," *Statistics in Medicine*, vol. 21, no. 18, pp. 2757–2770, 2002.
[19] C. Le Tourneau, J. J. Lee, and L. L. Siu, "Dose escalation methods in phase I cancer clinical trials," *Journal of the National Cancer Institute*, vol. 101, no. 10, pp. 708–720, 2009.
[20] E. A. Gehan, "The determination of the number of patients required in a preliminary and a follow-up trial of a new chemotherapeutic agent," *Journal of Chronic Diseases*, vol. 13, no. 4, pp. 346–353, 1961.
[21] R. Simon, "Optimal two-stage designs for phase II clinical trials," *Controlled Clinical Trials*, vol. 10, no. 1, pp. 1–10, 1989.
[22] T. R. Fleming, "One-sample multiple testing procedure for phase II clinical trials," *Biometrics*, vol. 38, no. 1, pp. 143–151, 1982.
[23] M. N. Chang, T. M. Therneau, H. S. Wieand, and S. S. Cha, "Design for group sequential phase II clinical trials," *Biometrics*, vol. 43, no. 4, pp. 865–874, 1987.
[24] P. F. Thall and R. Simon, "A Bayesian approach to establishing sample size and monitoring criteria for phase II clinical trials," *Controlled Clinical Trials*, vol. 15, no. 6, pp. 463–481, 1994.
[25] J. J. Lee and D. D. Liu, "A predictive probability design for phase II cancer clinical trials," *Clinical Trials*, vol. 5, no. 2, pp. 93–106, 2008.
[26] G. Yin, N. Chen, and J. J. Lee, "Phase II trial design with Bayesian adaptive randomization and predictive probability," *Journal of the Royal Statistical Society C*, vol. 61, no. 2, pp. 219–235, 2012.
[27] L. Rubinstein, J. Crowley, P. Ivy, M. Leblanc, and D. Sargent, "Randomized phase II designs," *Clinical Cancer Research*, vol. 15, no. 6, pp. 1883–1890, 2009.
[28] P. F. Thall, "A review of phase 2-3 clinical trial designs," *Lifetime Data Analysis*, vol. 14, no. 1, pp. 37–53, 2008.
[29] L. Y. T. Inoue, P. F. Thall, and D. A. Berry, "Seamlessly expanding a randomized phase II trial to phase III," *Biometrics*, vol. 58, no. 4, pp. 823–831, 2002.
[30] P. T. Lavin, "An alternative model for the evaluation of antitumor activity," *Cancer Clinical Trials*, vol. 4, no. 4, pp. 451–457, 1981.
[31] Y. Wang, C. Sung, C. Dartois et al., "Elucidation of relationship between tumor size and survival in non-small-cell lung cancer patients can aid early decision making in clinical drug development," *Clinical Pharmacology and Therapeutics*, vol. 86, no. 2, pp. 167–174, 2009.
[32] T. G. Karrison, M. L. Maitland, W. M. Stadler et al., "Design of phase II cancer trials using a continuous endpoint of change in tumor size: application to a study of sorafenib and erlotinib in non-small-cell lung cancer," *Journal of the National Cancer Institute*, vol. 99, no. 19, pp. 1455–1461, 2007.
[33] M. W. An, S. J. Mandrekar, M. E. Branda et al., "Comparison of continuous versus categorical tumor measurement–based metrics to predict overall survival in cancer treatment trials," *Clinical Cancer Research*, vol. 17, no. 20, pp. 6592–6599, 2011.
[34] G. Yothers, "Toward progression-free survival as a primary end point in advanced colorectal cancer," *Journal of Clinical Oncology*, vol. 25, no. 33, pp. 5153–5154, 2007.
[35] M. Buyse, P. Thirion, R. W. Carlson, T. Burzykowski, G. Molenberghs, and P. Piedbois, "Relation between tumour response to first-line chemotherapy and survival in advanced colorectal cancer: a meta-analysis. Meta-Analysis Group in Cancer," *The Lancet*, vol. 356, no. 9227, pp. 373–378, 2000.
[36] K. F. Schulz and D. A. Grimes, "Generation of allocation sequences in randomised trials: chance, not choice," *The Lancet*, vol. 359, no. 9305, pp. 515–519, 2002.
[37] J. M. Lachin, J. P. Matts, and L. J. Wei, "Randomization in clinical trials: conclusions and recommendations," *Controlled Clinical Trials*, vol. 9, no. 4, pp. 365–374, 1988.

[38] D. R. Taves, "Minimization: a new method of assigning patients to treatment and control groups," *Clinical Pharmacology and Therapeutics*, vol. 15, no. 5, pp. 443–453, 1974.
[39] S. J. Pocock and R. Simon, "Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial," *Biometrics*, vol. 31, no. 1, pp. 103–115, 1975.
[40] J. W. Frane, "A method of biased coin randomization, its implementation, and its validation," *Drug Information Journal*, vol. 32, no. 2, pp. 423–432, 1998.
[41] M. Kang, B. G. Ragan, and J. H. Park, "Issues in outcomes research: an overview of randomization techniques for clinical trials," *Journal of Athletic Training*, vol. 43, no. 2, pp. 215–221, 2008.
[42] S. J. Pocock, "Group sequential methods in the design and analysis of clinical trials," *Biometrika*, vol. 64, no. 2, pp. 191–199, 1977.
[43] P. C. O'Brien and T. R. Fleming, "A multiple testing procedure for clinical trials," *Biometrics*, vol. 35, no. 3, pp. 549–556, 1979.
[44] D. L. DeMets and K. K. G. Lan, "Interim analysis: the alpha spending function approach," *Statistics in Medicine*, vol. 13, no. 13-14, pp. 1341–1352, 1994.
[45] I. K. Hwang, W. J. Shih, and J. S. Decani, "Group sequential designs using a family of type I error probability spending functions," *Statistics in Medicine*, vol. 9, no. 12, pp. 1439–1445, 1990.
[46] C. Jennison and B. W. Turnbull, *Group Sequential Methods with Applications to Clinical Trials*, CRC Press, 2000.
[47] S. M. Berry and J. B. Kadane, "Optimal Bayesian randomization," *Journal of the Royal Statistical Society B*, vol. 59, no. 4, pp. 813–819, 1997.
[48] P. F. Thall and J. K. Wathen, "Practical Bayesian adaptive randomisation in clinical trials," *European Journal of Cancer*, vol. 43, no. 5, pp. 859–866, 2007.
[49] G. L. Rosner, "Bayesian monitoring of clinical trials with failure-time endpoints," *Biometrics*, vol. 61, no. 1, pp. 239–245, 2005.
[50] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, Chapman & Hall, Boca Raton, Fla, USA, 1995.
[51] D. A. Berry and C. H. Ho, "One-sided sequential stopping boundaries for clinical trials: a decision-theoretic approach," *Biometrics*, vol. 44, no. 1, pp. 219–227, 1988.
[52] J. D. Eales and C. Jennison, "An improved method for deriving optimal one-sided group sequential tests," *Biometrika*, vol. 79, no. 1, pp. 13–24, 1992.
[53] N. Cressie and J. Biele, "A sample-size-optimal Bayesian procedure for sequential pharmaceutical trials," *Biometrics*, vol. 50, no. 3, pp. 700–711, 1994.
[54] S. Barber and C. Jennison, "Optimal asymmetric one-sided group sequential tests," *Biometrika*, vol. 89, no. 1, pp. 49–60, 2002.
[55] M. Krams, K. R. Lees, W. Hacke, A. P. Grieve, J. M. Orgogozo, and G. A. Ford, "Acute stroke therapy by inhibition of neutrophils (ASTIN): an adaptive dose-response study of UK-279,276 in acute ischemic stroke," *Stroke*, vol. 34, no. 11, pp. 2543–2548, 2003.
[56] B. R. Luce, J. M. Kramer, S. N. Goodman et al., "Rethinking randomized clinical trials for comparative effectiveness research: the need for transformational change," *Annals of Internal Medicine*, vol. 151, no. 3, pp. 206–209, 2009.
[57] M. Posch and P. Bauer, "Adaptive two stage designs and the conditional error function," *Biometrical Journal*, vol. 41, no. 6, pp. 689–696, 1999.
[58] L. Cui, H. M. J. Hung, and S. J. Wang, "Modification of sample size in group sequential clinical trials," *Biometrics*, vol. 55, no. 3, pp. 853–857, 1999.
[59] W. J. Shih, "Group sequential, sample size re-estimation and two-stage adaptive designs in clinical trials: a comparison," *Statistics in Medicine*, vol. 25, no. 6, pp. 933–941, 2006.
[60] M. A. Proschan and S. A. Hunsberger, "Designed extension of studies based on conditional power," *Biometrics*, vol. 51, no. 4, pp. 1315–1324, 1995.
[61] M. A. Proschan, "Two-stage sample size re-estimation based on a nuisance parameter: a review," *Journal of Biopharmaceutical Statistics*, vol. 15, no. 4, pp. 559–574, 2005.

Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2012, Article ID 317634, 18 pages
doi:10.1155/2012/317634

Research Article

Escalation with Overdose Control Using Ordinal Toxicity Grades for Cancer Phase I Clinical Trials

Mourad Tighiouart, Galen Cook-Wiens, and André Rogatko

Samuel Oschin Comprehensive Cancer Institute, 8700 Beverly Boulevard, Los Angeles, CA 90048, USA

Correspondence should be addressed to Mourad Tighiouart, mourad.tighiouart@cshs.org

Received 29 June 2012; Accepted 10 September 2012

Academic Editor: Zhengjia Chen

We extend a Bayesian adaptive phase I clinical trial design known as escalation with overdose control (EWOC) by introducing an intermediate grade 2 toxicity when assessing dose-limiting toxicity (DLT). Under the proportional odds model assumption of dose-toxicity relationship, we prove that in the absence of DLT, the dose allocated to the next patient given that the previously treated patient had a maximum of grade 2 toxicity is lower than the dose given to the next patient had the previously treated patient exhibited a grade 0 or 1 toxicity at the most. Further, we prove that the coherence properties of EWOC are preserved. Simulation results show that the safety of the trial is not compromised and the efficiency of the estimate of the maximum tolerated dose (MTD) is maintained relative to EWOC treating DLT as a binary outcome and that fewer patients are overdosed using this design when the true MTD is close to the minimum dose.

1. Introduction

Cancer phase I clinical trials are sequential designs enrolling late stage cancer patients who have exhausted standard treatment therapies [1]. For cytotoxic agents or combinations of biologic with cytotoxic drugs, the main objectives of these trials are to characterize treatment-related toxicities and estimate a dose level that is associated with a predetermined level of dose limiting toxicity (DLT). Such a dose is called maximum tolerated dose (MTD) or phase II dose. Specifically, the MTD, γ, is defined as the dose that is expected to produce DLT after one cycle of therapy in a specified proportion θ of patients:

$$P(\text{DLT} \mid \text{Dose} = \gamma) = \theta. \tag{1.1}$$

Although the definition of DLT depends on the cancer type and the agent under study, it is typically defined as a grade 3 or 4 nonhematologic and grade 4 hematologic toxicity for

cytotoxic agents, see the National Cancer Institute (NCI) common toxicity criteria [2]. The value chosen for the target probability of DLT θ depends on the nature and severity of the DLT; it is set relatively high when the DLT is a transient, reversible, and nonfatal condition and low when it is lethal or life threatening.

Model-based designs for cancer phase I clinical trials have been studied extensively in the last two decades; see O'Quigley et al. [3], Gatsonis and Greenhouse [4], Durham and Flournoy [5], Korn et al. [6], Whitehead and Brunier [7], Whitehead [8], Babb et al. [9], Gasparini and Eisele [10], Mukhopadhyay [11], and Haines et al. [12]. In particular, the continual reassessment method (CRM) first proposed by O'Quigley et al. [3] is Bayesian outcome adaptive, and has been modified and extended by many authors; see, for example, Faries [13], Moller [14], Goodman et al. [15], O'Quigley and Shen [16], Piantadosi et al. [17], Cheung and Chappell [18], Storer [19], and Leung and Wang [20]. The attribute of the CRM is that at each stage of the trial, we seek a dose for which a Bayes estimate of the probability of DLT is closest to the target probability of DLT θ. Escalation with overdose control (EWOC) originally proposed by Babb et al. [9] is another alternative Bayesian outcome adaptive design for dose finding in early phase cancer trials. Unlike CRM, its main feature is that at each stage of the trial, we seek a dose for which the posterior probability of exceeding the MTD γ is bounded by a feasibility bound α. Statistical properties of EWOC have been studied by Zacks et al. [21], Tighiouart et al. [22], and Tighiouart and Rogatko [23], and examples of phase I trials using EWOC can be found in [24–33]. Further literature review of statistical methods for dose finding in cancer phase I trials can be found in Ting [34] and Chevret [35].

The above methods allocate future doses based on a binary outcome of DLT of previously treated patients. Such designs may not be efficient in the sense that the dose recommended for the next patient is the same regardless whether the previously treated patient had no toxicity or had an intermediate grade 2 toxicity. The work we present in this manuscript is motivated by the ethical concern raised by clinical colleagues regarding dose escalation in the absence of DLT. Specifically, if the current patient experiences drug related grade 2 toxicity at the most, then the dose to be allocated to the next patient should not be as high as the one had the current patient experienced a maximum of grade 0 or 1 toxicity. We present a Bayesian outcome adaptive design which is an extension of EWOC by accommodating an intermediate grade 2 toxicity to the model. We use a proportional odds model to describe the dose-toxicities relationship and the design is termed EWOC proportional odds model, written as EWOC-POM. We show that the design satisfies the above ethical consideration without compromising the safety and efficiency of the trial. Furthermore, we show that the design maintains the coherence properties of EWOC.

Cancer phase I clinical trials designs taking into account all grades and types of toxicities have been proposed by many authors with the goal of improving the safety and efficiency of the trial, see Gordon and Willson [36], Wang et al. [37], Bekele and Thall [38], Yuan et al. [39], Potthoff and George [40], Bekele et al. [41], Van Meter et al. [42, 43], Iasonos et al. [44], Lee et al. [45], and Chen et al. [46]. Some of these methods use multivariate models for characterizing the different grades of toxicities as a function of dose while others summarize the information from all toxicity grades into a continuous score. Depending on the underlying scenarios, in general, adding multiple toxicity grades information to the model has little effect on the safety of the trial with a modest gain in the precision of the estimate of the MTD. Our contribution in this manuscript is motivated by the ethical constraint that dose escalation in the absence of DLT should be mitigated by the occurrence of an intermediate grade 2 toxicity. The model and design we propose is constructed in such a way as to maintain the main

properties of EWOC while at the same time, satisfying those ethical considerations raised by our clinical colleagues.

The manuscript is organized as follows. In Section 2, we give a detailed description of the methodology and describe the trial design. In Section 3, we state and prove the ethical considerations and coherence properties of EWOC-POM. The simulation results of the design operating characteristics and comparison with EWOC design are included in Section 4. Section 5 contains some final remarks and discussion of practical implementations.

2. Method

2.1. Model

Let $G = 0, 1, \ldots, 4$ be the maximum grade of toxicity experienced by a patient by the end of one cycle of therapy and define DLT as a maximum of grade 3 or 4 toxicity. Let

$$Y = \begin{cases} 0 & \text{if } G = 0 \text{ or } 1 \\ 1 & \text{if } G = 2 \\ 2 & \text{if } G = 3 \text{ or } 4. \end{cases} \tag{2.1}$$

We model the dose-toxicities relationship by assuming that

$$P(Y \geq j \mid x) = F(\alpha_j + \beta x), \quad j = 1, 2, \tag{2.2}$$

where $F(\cdot)$ is a known strictly increasing cdf. This implies that $\alpha_2 \leq \alpha_1$. We assume that $\beta > 0$ so that the probability of DLT is an increasing function of dose. The MTD, γ, is defined as the dose that is expected to produce DLT in a specified proportion θ of patients:

$$P(Y = 2 \mid x = \gamma) = F(\alpha_2 + \beta\gamma) = \theta. \tag{2.3}$$

The value chosen for the target probability θ depends on the nature and clinical manageability of the DLT; it is set relatively high when the DLT is a transient, correctable, or nonfatal condition and low when it is lethal or life threatening. Suppose that dose levels in the trial are selected in the interval $[X_{\min}, X_{\max}]$.

2.1.1. Likelihood

Let $D_n = \{(x_i, Y_i),\ i = 1, \ldots, n\}$ be the data after enrolling n patients to the trial. The likelihood function for the parameters α_1, α_2, and β is

$$\begin{aligned} L(\alpha_1, \alpha_2, \beta \mid D_n) = \prod_{i=1}^{n} &\left[1 - F(\alpha_1 + \beta x_i)\right]^{I(Y_i=0)} \left[F(\alpha_1 + \beta x_i) - F(\alpha_2 + \beta x_i)\right]^{I(Y_i=1)} \\ &\times \left[F(\alpha_2 + \beta x_i)\right]^{I(Y_i=2)}, \end{aligned} \tag{2.4}$$

where $I(\cdot)$ is the indicator function.

We reparameterize model (2.2) in terms of $\rho_0 = P(Y = 2 \mid x = X_{\min})$, the probability that a DLT manifests within the first cycle of therapy for a patient given dose $x = X_{\min}$, $\rho_1 = P(Y \geq 1 \mid x = X_{\min})$, the probability that a grade 2 or more toxicity manifests within the first cycle of therapy for a patient given dose $x = X_{\min}$, and the MTD γ. This reparameterization is convenient to clinicians since γ is the parameter of interest. Assuming that the dose is standardized to be in the interval $[0, 1]$, it can be shown that

$$\begin{aligned} \alpha_1 &= F^{-1}(\rho_1), \qquad \alpha_2 = F^{-1}(\rho_0), \\ \beta &= \frac{1}{\gamma}\Big(F^{-1}(\theta) - F^{-1}(\rho_0)\Big). \end{aligned} \tag{2.5}$$

The conditions $\alpha_2 \leq \alpha_1$, $\beta > 0$, and (2.2) imply that $0 \leq \rho_0 \leq \rho_1$ and $0 \leq \rho_0 \leq \theta$. Define

$$\begin{aligned} F_1(\rho_0, \rho_1, \gamma; x) &= F\left(F^{-1}(\rho_1) + \Big(F^{-1}(\theta) - F^{-1}(\rho_0)\Big)\frac{x}{\gamma}\right) \\ F_2(\rho_0, \rho_1, \gamma; x) &= F\left(F^{-1}(\rho_0) + \Big(F^{-1}(\theta) - F^{-1}(\rho_0)\Big)\frac{x}{\gamma}\right). \end{aligned} \tag{2.6}$$

Using (2.4), (2.5), and (2.6), the likelihood of the reparameterized model is

$$\begin{aligned} L(\rho_0, \rho_1, \gamma \mid D_n) = \prod_{i=1}^{n} &\big[1 - F_1(\rho_0, \rho_1, \gamma; x_i)\big]^{I(Y_i=0)} \big[F_1(\rho_0, \rho_1, \gamma; x_i) - F_2(\rho_0, \rho_1, \gamma; x_i)\big]^{I(Y_i=1)} \\ &\times \big[F_2(\rho_0, \rho_1, \gamma; x_i)\big]^{I(Y_i=2)}. \end{aligned} \tag{2.7}$$

2.1.2. Prior and Posterior Distributions

Let $g(\rho_0, \rho_1, \gamma)$ be the prior distribution on Ω, where $\Omega = \{(x, y, z) : 0 \leq x \leq \theta,\ x \leq y \leq 1,\ X_{\min} \leq z \leq X_{\max}\}$. Using Bayes rule, the posterior distribution of the model parameters is proportional to the product of the likelihood and prior distribution

$$\pi(\rho_0, \rho_1, y \mid D_n) \propto L(\rho_0, \rho_1, \gamma \mid D_n) \times g(\rho_0, \rho_1, \gamma). \tag{2.8}$$

We designed an MCMC sampler based on the Metropolis-Hastings algorithm [47, 48] to obtain model operating characteristics. We also used WinBUGS [49] to estimate features of the posterior distribution of the MTD and design a trial. The WinBUGS code is included in the Appendix section. In the absence of prior information about the MTD and probability of DLT at $X_{\min}$, we specify vague priors for the model parameters as follows:

$$\begin{aligned} \gamma &\sim \text{Unif}[X_{\min}, X_{\max}] \\ \rho_0 &\sim \text{Unif}[0, \theta] \\ \rho_1 \mid \rho_0 &\sim \text{Unif}[\rho_0, 1]. \end{aligned} \tag{2.9}$$

2.1.3. Trial Design

Dose levels in the trial are selected in the interval $[X_{\min}, X_{\max}]$. The adaptive design proceeds as follows. The first patient receives a dose $x_1 > X_{\min}$ that is deemed to be safe by the clinician. Denote by $\Pi k(\gamma) = \Pi(\gamma \mid D_k)$ the marginal posterior cdf of the MTD, $k = 1, \ldots, n-1$. The $(k+1)-st$ patient receives the dose $x_{k+1} = \Pi_k^{-1}(\alpha)$ so that the posterior probability of exceeding the MTD is equal to the feasibility bound α. This is the overdose protection property of EWOC, where at each stage of the design, we seek a dose to allocate to the next patient while controlling the posterior probability of exposing patients to toxic dose levels. The trial proceeds until a predetermined number of n patients are enrolled to the trial. At the end of the trial, we estimate the MTD as $\widehat{\gamma} = \Pi_n^{-1}(\alpha)$.

3. Properties of EWOC-POM

3.1. Characteristics of EWOC-POM

The proposed design EWOC-POM assigns dose levels to future patients by taking into account the maximum observed toxicity grade during the first cycle of therapy according to the following properties.

(i) At each stage of the design, we seek a dose to allocate to the next patient while controlling the posterior probability of exposing patients to toxic dose levels.

(ii) If the maximum grade of toxicity experienced by patient $k-1$ within one cycle of therapy is grade 2, then the dose allocated to patient k is lower than the dose that would have been given to patient k had the maximum grade of toxicity experienced by patient $k-1$ been grade 0 or 1.

Property (i) is the overdose protection defining characteristic of EWOC which is also satisfied by EWOC-POM. Property (ii) is naturally appealing because when a patient experiences grade 2 dose-related toxicity, then it is ethical not to increase the dose by the same amount had this patient exhibited grade 0 or 1 toxicity at the most. Characteristic (ii) is summarized in the following theorem.

Theorem 3.1. *Let $D_k = \{(Y_1, x_1), \ldots, (Y_k, x_k)\}$ be the data on the first k patients generated by the design described in Section 2.1.3, and, $\Pi_k(\gamma; Y_k)$ be the cdf of γ given the data D_k. Let $x_{k+1} = \Pi_k^{-1}(\alpha; Y_k)$ and $x'_{k+1} = \Pi_k^{-1}(\alpha; Y'_k)$. Suppose that for all $x \in [X_{\min}, X_{\max}]$ and all (ρ_0, ρ_1) such that $0 \leq \rho_0 \leq \rho_1 \leq 1$ and $\rho_0 \leq \theta$, $(F_1 - F_2)/(1 - F_1)$ is a monotonically decreasing function in γ. Then, $x'_{k+1} \geq x_{k+1}$ whenever $Y_k{}' = 0$ and $Y_k = 1$.*

The proof of Theorem 3.1 is given in the Appendix section. It is easy to verify that the monotonicity condition of Theorem 3.1 holds for the logistic function $F(w) = 1/(1 + e^{-w})$. Using this link function and the uniform priors given in (2.9) with $\theta = 0.33$, Figure 1 gives all possible dose assignments for patients 1 and 2 and selected situations for patients 3 and 4 using the trial design described in Section 2.1.3. The dose has been standardized so that $X_{\min} = 0$ and $X_{\max} = 1$, and the first patient is given dose 0.10. If this patient experiences grade 0 or 1 toxicity at the most, the dose recommended for patient 2 is 0.36. On the other hand, if patient 1 experiences grade 2 toxicity at the most, the dose recommended for patient 2 is 0.22, much lower than 0.36. This figure also illustrates some of the coherence properties stated in Theorem 3.2 below.

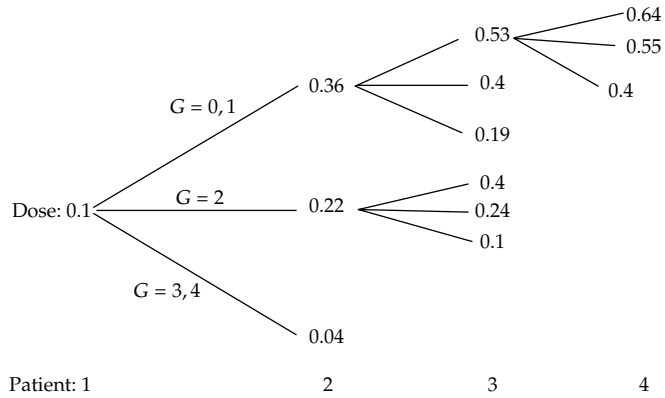


Figure 1: Tree of possible dose allocations for patients 1 and 2 and selected situations for patients 3 and 4. $G = 0, 1$ corresponds to $Y = 0$, $G = 2$ corresponds to $Y = 1$, and $G = 3, 4$ corresponds to $Y = 2$ or DLT.

3.2. Coherence of EWOC-POM

In cancer phase I clinical trials, it is ethical not to increase a dose of a cytotoxic agent for the next patient if the previously treated patient exhibited DLT when given the same dose level. Furthermore, it is desirable not to decrease the dose of an agent for the next patient if the previously treated patient did not experience DLT when given that same dose level. These two properties are known as coherence in escalation and de-escalation, respectively, see Cheung [50]. A design that satisfies both of these properties is said to be coherent. Tighiouart and Rogatko [23] show that EWOC is coherent. The next theorem states some of the coherence properties of EWOC-POM.

Theorem 3.2. *Suppose that for all $x \in [X_{\min}, X_{\max}]$ and all (ρ_0, ρ_1) such that $0 \leq \rho_0 \leq \rho_1 \leq 1$ and $\rho_0 \leq \theta$, F_1 and F_2 are monotonically decreasing in γ. Then the design EWOC-POM described in 2.1.3 is coherent in deescalation. Furthermore, if the toxicity response for patient k is $Y_k = 0$, then the dose allocated to patient $k + 1$ satisfies $x_{k+1} \geq x_k$.*

The proof of Theorem 3.2 is given in the Appendix section.

4. Simulation Studies

We compare the design operating characteristics of EWOC-POM with the original EWOC by simulating a large number of trials under several scenarios. We used the logistic function $F(w) = 1/(1+e^{-w})$ to model the dose-toxicities relationship in (2.2). EWOC was implemented as in Tighiouart et al. [22] using the same logistic function to model the dose-toxicity relationship. For all scenarios, we standardize the dose to be in the interval $[0, 1]$, $\theta = 0.33$, the feasibility bound $\alpha = 0.25$, and the trial sample size is $n = 30$. The priors in (2.9) were adopted for EWOC-POM. We also carried out design operating characteristics for $\theta = 0.20$, and the conclusions regarding the performance of EWOC-POM relative to EWOC were essentially the same.

4.1. Algorithm

For a given scenario determined by ρ_0, ρ_1, and γ, the first patient receives dose 0, and the next dose x_2 is determined according to the trial design described in 2.1.3. The second

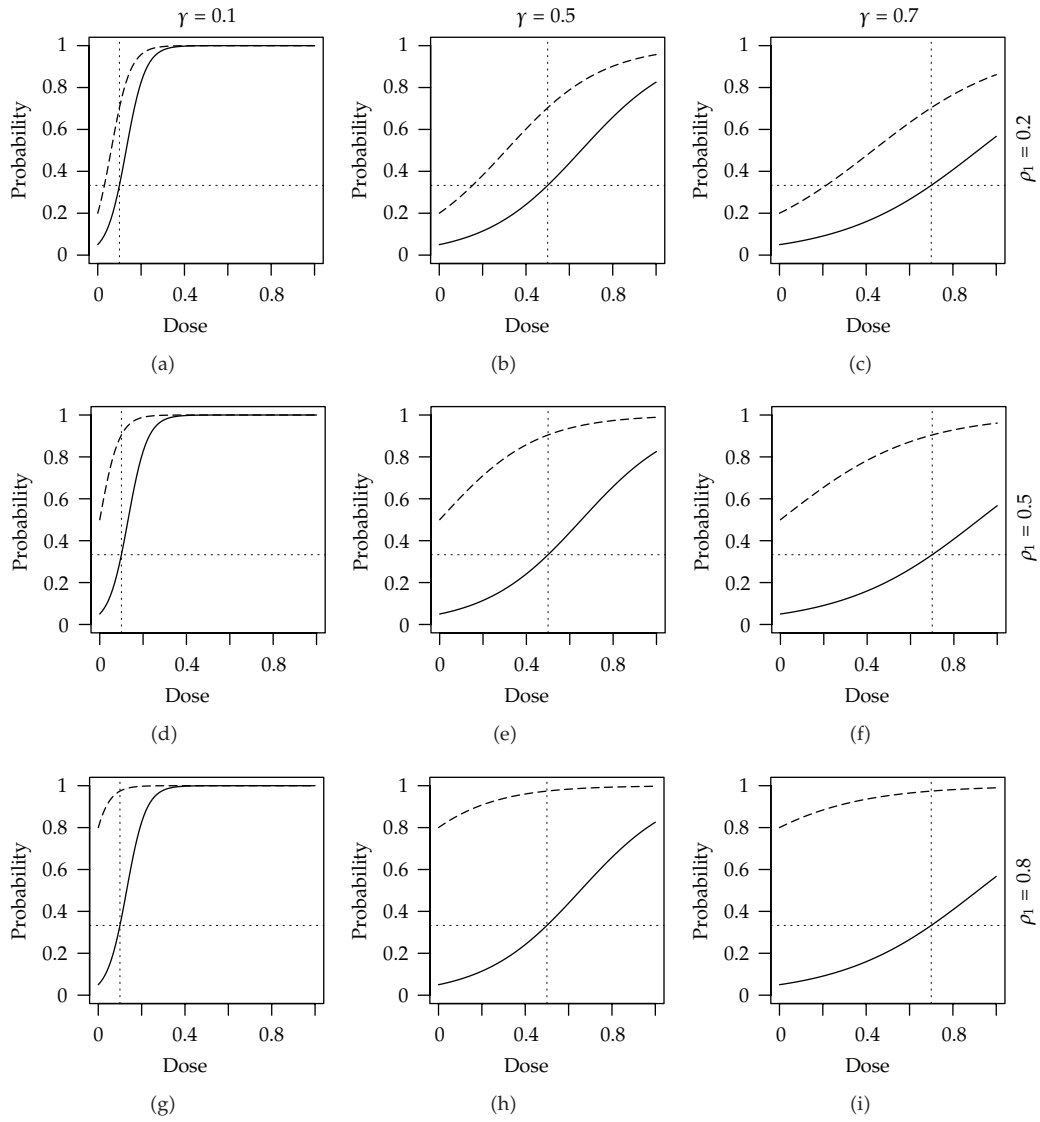


Figure 2: Dose-toxicity relationship for the nine scenarios considered in the design operations characteristics. The solid curves correspond to $P(Y = 2 \mid x) = P(DLT \mid x)$ and the dashed curves in bold correspond to $P(Y > 1 \mid x)$. The horizontal dashed lines represent the target probability of DLT $\theta = 0.33$ and the vertical lines correspond to the true values of the MTD γ.

response y_2 is then generated from model (2.2) reparameterized in terms of ρ_0, ρ_1, and γ with $x = x_2$. This process is repeated until all n patients have been enrolled to the trial. We considered 9 scenarios corresponding to a fixed value for $\rho_0 = 0.05$, three values of ρ_1, 0.2, 0.5, and 0.8, and three values of the MTD γ, 0.1, 0.5, and 0.7. The corresponding dose-toxicity relationships for these nine scenarios are illustrated in Figure 2. For each model and each scenario, we simulated $M = 1000$ trials. EWOC and EWOC-POM were compared in terms of the proportion of patients exhibiting DLT, the average bias, $\text{bias}_{\text{ave}} = M^{-1}(\sum_{i=1}^{M} \widehat{\gamma}_i - \gamma_{\text{true}})$, and the estimated mean square error $\text{MSE} = M^{-1}(\sum_{i=1}^{M} (\widehat{\gamma}_i - \gamma_{\text{true}})^2)$, where $\widehat{\gamma}_i$ is the Bayes

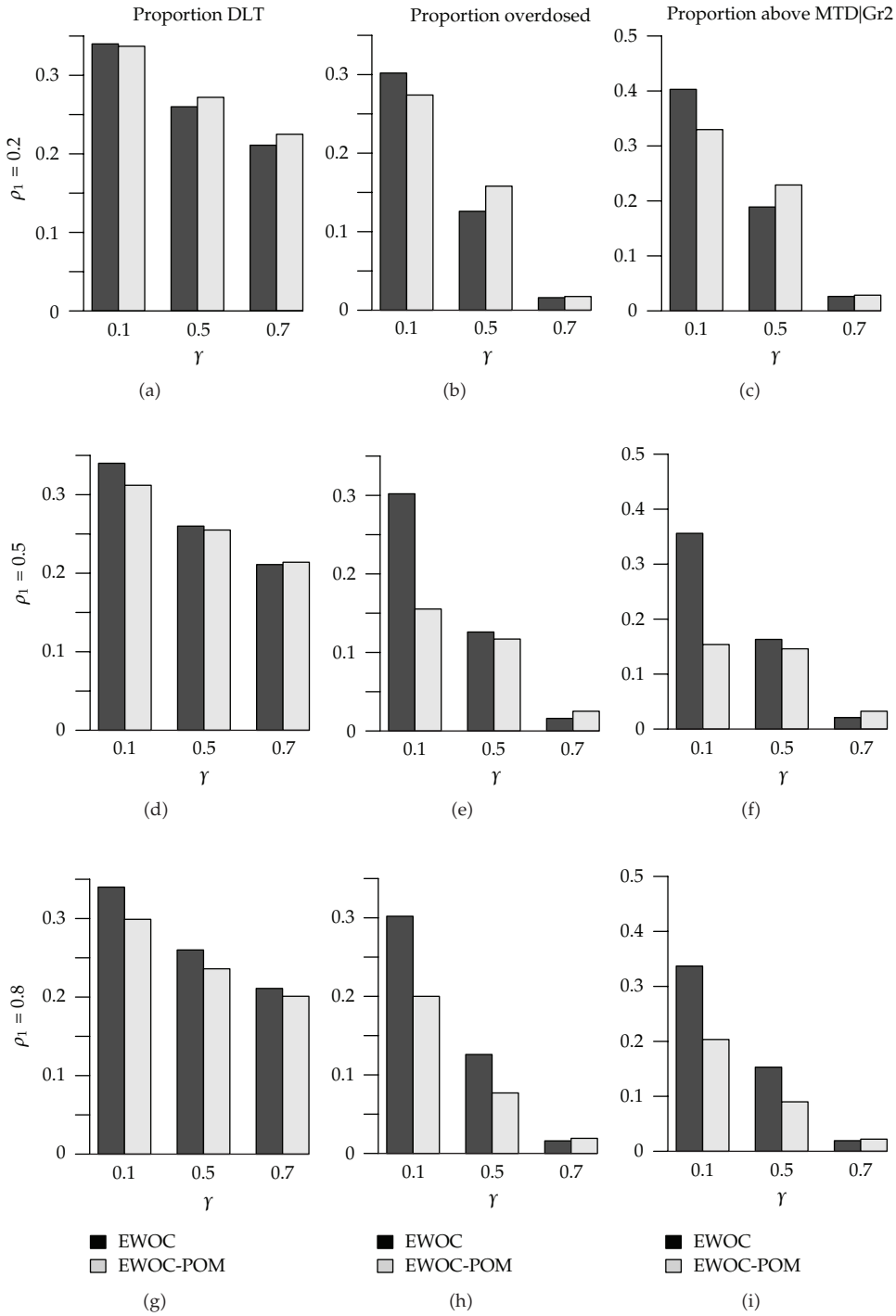


Figure 3: Summary statistics for trial safety for EWOC and EWOC-POM under all scenarios. Each graph represents mean proportion obtained from all patients from all 1000 simulated trials.

Table 1: Percent of trials with estimated MTD within 5% of the dose range and 10% of the dose range of the true MTD γ and percent of trials for which the rate of DLT exceeds 40% for EWOC and EWOC-POM under the nine scenarios.

Scenarios	Percent of trial with estimated MTD				Percent of Trial	
	Within 0.05 of γ		Within 0.10 of γ		with rate of DLT >0.4	
	EWOC	EWOC-POM	EWOC	EWOC-POM	EWOC	EWOC-POM
$\gamma = 0.1, \rho_1 = 0.2$	98.3	98.4	100	100	7.5	6.6
$\gamma = 0.1, \rho_1 = 0.5$	98.3	97.5	100	100	7.5	3.0
$\gamma = 0.1, \rho_1 = 0.8$	98.3	96.4	100	100	7.5	2.9
$\gamma = 0.5, \rho_1 = 0.2$	39.6	40.5	70.3	71.3	0.2	0.0
$\gamma = 0.5, \rho_1 = 0.5$	39.6	35.6	70.3	63.2	0.2	0.0
$\gamma = 0.5, \rho_1 = 0.8$	39.6	31.0	70.3	59.4	0.2	0.0
$\gamma = 0.7, \rho_1 = 0.2$	24.3	27.6	49.1	53.3	0.0	0.0
$\gamma = 0.7, \rho_1 = 0.5$	24.3	23.2	49.1	45.7	0.0	0.0
$\gamma = 0.7, \rho_1 = 0.8$	24.3	20.1	49.1	37.1	0.0	0.0

estimate of the posterior distribution of the MTD at the end of the ith trial with respect to the asymmetric loss function:

$$L(x,\gamma) = \begin{cases} \alpha(\gamma - x) & \text{if } x \leq \gamma \\ (1-\alpha)(x-\gamma) & \text{if } x > \gamma. \end{cases} \tag{4.1}$$

We also compared the models with respect to the proportion of patients that were overdosed. Here, a dose x is defined as an overdose if $x > x^*$, where x^* is defined as the dose for which $P(\text{DLT} \mid x^*) = \theta + 0.05$. This probability is calculated using the parameter values from the corresponding scenario. These models are further compared with respect to the proportion of patients that were overdosed given that the previously treated patient exhibited grade 2 toxicity. Finally, we compared EWOC-POM to EWOC in terms of the proportion of trials for which the probability of DLT exceeds 0.4. This gives us an estimate of the probability that a prospective trial will result in an excessively high DLT rate. As for the proportion of trials with correct MTD recommendation, we presented percent of trials with estimated MTD within 10% and 20% of the dose range of the true MTD for EWOC-POM and EWOC.

4.2. Results

Figure 3 shows that the proportion of patients exhibiting DLT is always less than 34% for both EWOC and EWOC-POM under all scenarios and 4% fewer patients experiencing DLT under EWOC-POM when the MTD is small ($\gamma = 0.1$) and $\rho_1 = 0.8$. The same figure shows that the proportion of patients that are overdosed using EWOC is uniformly higher relative to EWOC-POM when the MTD is small. The same trend is observed when $\gamma = 0.5$ except when $\rho_1 = 0.2$. The difference in the proportion of patients being overdosed when $\gamma = 0.7$ is negligible. The last panel of Figure 3 shows that the proportion of patients that are overdosed given that the previously treated patient exhibited grade 2 toxicity using EWOC is uniformly higher relative to EWOC-POM when $\gamma = 0.1, 0.5$ except when $\rho_1 = 0.2$. The difference in these proportions when $\gamma = 0.7$ is negligible. The last two columns of Table 1 show that the percent of trials with DLT rate of 0.4 or more is 7.5% at the most for EWOC and 6.6% for EWOC-POM. A more detailed comparison is shown in Figure 4, where side-by-side box plots of

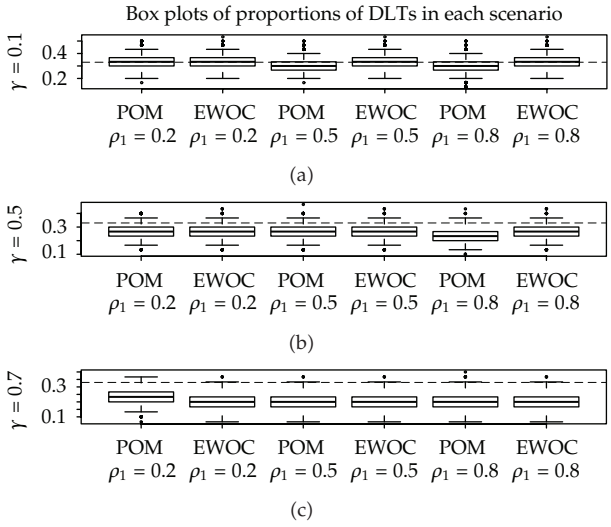


Figure 4: Box plots for the proportion of DLTs for EWOC-POM and EWOC under the nine scenarios. Each box plot was constructed from the DLT rates of the 1000 simulated trials. The dashed horizontal line corresponds to the target probability of DLT $\theta = 0.33$.

the distributions of the proportion of DLTs for EWOC-POM and EWOC under the nine scenarios are displayed. These results show that EWOC-POM maintains the safety of the trial relative to EWOC and is much safer when the true MTD is close to the minimum dose by reducing the number of patients that are exposed to toxic doses.

Figure 5 shows that the estimated MTDs using EWOC and EWOC-POM are very close in general, with the highest difference observed when $\rho_1 = 0.8$. This is reflected by the estimated bias and RMSE shown in Figure 5. This is expected since EWOC-POM is characterized by a conservative dose escalation when a patient experiences grade 2 toxicity. The highest absolute value of the bias is 0.04 and is achieved when $\gamma = 0.5, 0.7$ and $\rho_1 = 0.8$. This constitutes 4% of the range of the dose and is practically not significant. The percent of trials with estimated MTD within 5% of the dose range and 10% of the dose range of the true MTD γ under the nine scenarios are shown in columns 2–5 of Table 1. These results further confirm that the precision of the estimate of the MTD is similar between the two models, with a higher precision for EWOC achieved when $\gamma = 0.5$ and $\rho_1 = 0.8$. Values other than 5% and 10% of the dose range were also used, and the conclusions were essentially the same. We conclude that EWOC-POM maintains the efficiency of the trial relative to EWOC for all practical purpose.

These simulation results suggest that EWOC-POM is a good alternative design for cancer phase I clinical trials which takes into account the ethical consideration that dose escalation in the absence of DLT is mitigated by the occurrence of a grade 2 toxicity.

4.3. Model Robustness

Model robustness with respect to assumption of proportional odds model in (2.2) was assessed by simulating the toxicity responses from a nonproportional odds model:

$$P(Y \geq j \mid x) = F(\alpha_j + \beta_j x), \quad j = 1, 2. \tag{4.2}$$

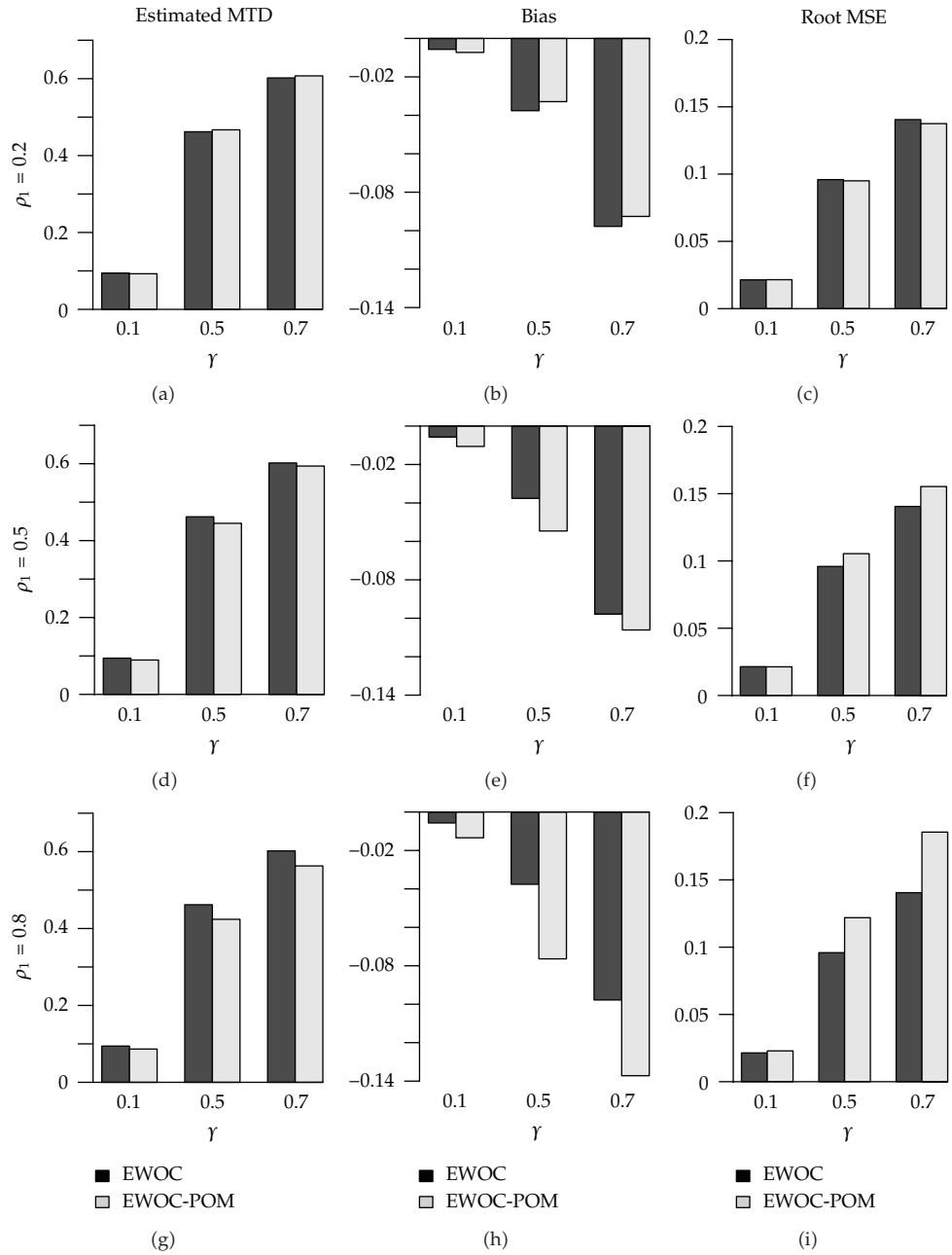


Figure 5: Summary statistics for trial efficiency for EWOC and EWOC-POM under all scenarios. Each graph represents a mean value obtained from all patients from all 1000 trials.

The same logistic link function $F(w) = 1/(1 + e^{-w})$ was used. We considered two different models M_1 and M_2 corresponding to the set of parameters $\alpha_2 = -3.94$, $\beta_1 = 22.36$, $\beta_2 = 32.36$ for model M_1 and $\alpha_2 = -1.94$, $\beta_1 = 22.36$, $\beta_2 = 12.36$ for M_2. For each model M_i, $i = 1, 2$, we considered three different values for the intercept term α_1, $\alpha_1 = -1.38, 0.00, 1.38$ which correspond to $\rho_1 = 0.2, 0.5, 0.8$. These parameters have been selected so that $\rho_0 = 0.020$ for

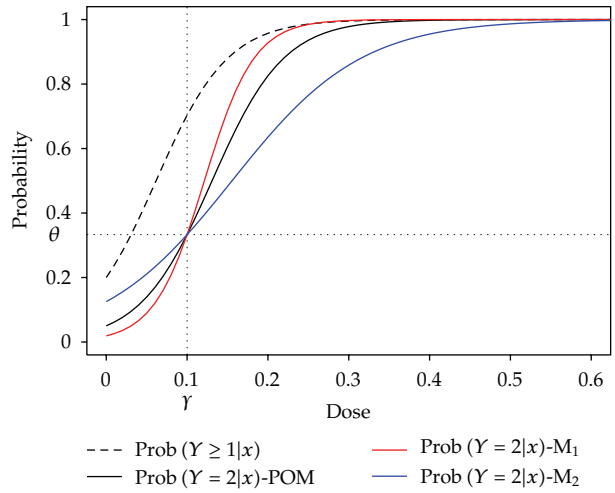


Figure 6: Dose-toxicity relationship under the proportional (EWOC-POM) and nonproportional odds models (M_1 and M_2) when the true MTD $\gamma = 0.1$. The dashed horizontal line corresponds to the target probability of DLT $\theta = 0.33$.

model M_1, $\rho_0 = 0.126$ for model M_2, and the true MTD is $\gamma = 0.1$. Figure 6 shows the graphs of the probabilities of DLT and probability of grade 2 or more toxicity as a function of dose for the proportional odds model POM and models M_1 and M_2 when $\rho_1 = 0.2$. For each scenario, we simulated 1000 trials with $n = 30$ patients where at each stage of the trial, the next dose is calculated using the trial design described in 2.1.3 as in Section 4.1 but the toxicity response is generated using the nonproportional odds model (4.2). The simulation results are summarized in Table 2. The maximum difference in proportion of patients exhibiting DLT (averaged across the simulated trials) between model M_i, $i = 1, 2$ and EWOC-POM is 3%. Under model M_2, the proportion of patients that are overdosed is higher than EWOC-POM, and this proportion is 13% higher when $\rho_1 = 0.2$.

The percent of trials with DLT rate exceeding 0.4 is less than 15% in all cases. This percent is highest for model M_2; however, this is still relatively small compared to the results obtained in [42]. The percent of trials with estimated MTD within 10% of the dose range of the true MTD is 100% between the three models and across all scenarios and very good within 5% of the dose range of the true MTD.

In summary, EWOC-POM seems to be robust to model misspecification when the true MTD is near the initial dose. On the other hand, the model is sensitive to model misspecification when the true MTD is high but the safety of the trial is not compromised. We also conducted similar simulations (results not shown) when the true MTD is $\gamma = 0.5, 0.7$. We found that under all scenarios, the proportion of patients exhibiting DLT is always less than 33% but the bias tends to be higher for high values of ρ_1 and γ. This is the case when the probability of DLT curve increases very slowly as a function of dose which results in a very slow dose escalation scheme.

5. Discussion

In this paper, we proposed a Bayesian adaptive design for dose-finding studies in cancer phase I clinical trials. The method addresses the ethical concern regarding dose escalation in

Table 2: Summary statistics for trial safety and efficiency under model misspecification when the true MTD $\gamma = 0.1$. Rows 2–6 give the summary statistics based on all patients from all 1000 trials.

Statistics	Model	ρ_1 0.2	0.5	0.8
Proportion of DLTs	EWOC-POM	0.337	0.312	0.299
	M_1	0.328	0.304	0.289
	M_2	0.357	0.342	0.331
Proportion overdosed	EWOC-POM	0.274	0.155	0.200
	M_1	0.240	0.203	0.173
	M_2	0.404	0.351	0.322
MTD estimate	EWOC-POM	0.093	0.090	0.087
	M_1	0.094	0.092	0.090
	M_2	0.095	0.090	0.086
Bias	EWOC-POM	−0.007	−0.010	−0.013
	M_1	−0.006	−0.007	−0.010
	M_2	−0.005	−0.010	−0.013
Root MSE	EWOC-POM	0.022	0.021	0.023
	M_1	0.016	0.016	0.017
	M_2	0.027	0.029	0.031
Percent of trial with DLT rate >40%	EWOC-POM	6.6	3.0	2.9
	M_1	1.5	1.2	0.8
	M_2	14.7	11.6	12.6
Percent of trial with estimated MTD within 0.05 of γ	EWOC-POM	98.4	97.5	96.4
	M_1	99.7	99.5	99.5
	M_2	91.1	91.2	88.1
Percent of trial with estimated MTD within 0.10 of γ	EWOC-POM	100.0	100.0	100.0
	M_1	100.0	100.0	100.0
	M_2	100.0	100.0	100.0

the absence of DLT. Specifically, if the current patient experiences drug-related grade 2 toxicity at the most, then it is ethical not to escalate the dose for the next patient by the same amount as the one had the current patient experienced a maximum of grade 0 or 1 toxicity. The method termed EWOC-POM is an extension of EWOC by accommodating an intermediate grade 2 toxicity to the model. We used a proportional odds model to describe the dose-toxicities relationship for simplicity. We proved that if the maximum grade of toxicity experienced by patient $k-1$ within one cycle of therapy is grade 2, then the dose allocated to patient k is lower than the dose that would have been given to patient k had the maximum grade of toxicity experienced by patient $k-1$ been grade 0 or 1. Furthermore, we also showed that the coherence properties of EWOC are maintained.

We studied design operating characteristics by simulating a large number of trials under different scenarios of the dose-toxicity relationships. EWOC-POM was compared to EWOC with respect to the primary goals of cancer phase I trials, safety and efficiency of the estimate of the MTD. We found that in general, the safety of the trial is not compromised when we account for an intermediate grade 2 toxicity. In particular, when the unknown MTD is close to the initial dose, a substantial number of patients are overdosed when using EWOC relative to EWOC-POM, and if the current patient experiences grade 2 toxicity, then the next patient is more likely to be overdosed using EWOC compared to EWOC-POM. The loss in efficiency of the estimate of the MTD by introducing an extra parameter to the model is very

marginal as was shown by the simulation results of the various scenarios. We also showed that the proportional odds assumption is robust to model misspecification when the true MTD is close to the minimum dose. However, the bias of the estimate of the MTD increases as a function of the MTD under model misspecification. In any case, safety of the trial as assessed by the number of patients exhibiting DLT and number of patients that are overdosed was never compromised.

We have shown that Theorem 3.1 holds under the proportional odds assumption using EWOC scheme with link functions satisfying monotonicity conditions as a function of the MTD. One can easily prove that a similar ethical constraint stated in Theorem 3.1 can be established using the Bayesian CRM originally proposed in O'Quigley et al. [3]. In fact, a more general theorem can be established for a Bayes estimate using a general convex loss function. Extensions of EWOC-POM to accommodate patients' baseline characteristics ([30]) and time to event DLT for late onset toxicities are being investigated. In conclusion, EWOC-POM is a good alternative design to EWOC if clinicians expect to see a substantial number of grade 2 toxicities induced by the agent or drug combinations under study.

Appendix

Proof of Theorem 3.1. To simplify notation and presentation of the proof, we assume that $X_{\min} = 0, \rho_0 \le \rho_1$ are fixed, and let us drop the nuisance parameters ρ_0 and ρ_1 from $F_1(\rho_0, \rho_1, \gamma; x)$, and $F_2(\rho_0, \rho_1, \gamma; x)$. Let $L_k(\gamma) = L_k(\rho_0, \rho_1, \gamma \mid D_k)$, and $\pi(\gamma)$ be a proper prior density for γ. If $Y_k = 0$, $L_k(\gamma) = L_{k-1}(\gamma)[1 - F_1(\gamma; x_k)]$, and if $Y_k = 1$, $L_k(\gamma) = L_{k-1}(\gamma)[F_1(\gamma; x_k) - F_2(\gamma; x_k)]$. Using Bayes' rule,

$$
\begin{aligned}
&\prod_k(t;0) - \prod_k(t;1) \\
&\quad = \frac{\int_0^t L_{k-1}(\gamma)\left[1 - F_1(\gamma; x_k)\right]\pi(\gamma)d\gamma}{\int_0^1 L_{k-1}(\gamma)\left[1 - F_1(\gamma; x_k)\right]\pi(\gamma)d\gamma} - \frac{\int_0^t L_{k-1}(\gamma)\left[F_1(\gamma; x_k) - F_2(\gamma; x_k)\right]\pi(\gamma)d\gamma}{\int_0^1 L_{k-1}(\gamma)\left[F_1(\gamma; x_k) - F_2(\gamma; x_k)\right]\pi(\gamma)d\gamma} \\
&\quad = A^{-1}\left[\int_0^t\int_0^1 L_{k-1}(\gamma)L_{k-1}(\gamma')\pi(\gamma)\pi(\gamma')\left[1 - F_1(\gamma; x_k)\right]\left[F_1(\gamma'; x_k) - F_2(\gamma'; x_k)\right]d\gamma' d\gamma\right] \\
&\qquad - A^{-1}\left[\int_0^t\int_0^1 L_{k-1}(\gamma)L_{k-1}(\gamma')\pi(\gamma)\pi(\gamma')\left[F_1(\gamma; x_k) - F_2(\gamma; x_k)\right]\left[1 - F_1(\gamma'; x_k)\right]d\gamma' d\gamma\right] \\
&\quad = A^{-1}\left[\int_0^t\int_t^1 L_{k-1}(\gamma)L_{k-1}(\gamma')\pi(\gamma)\pi(\gamma')\left[1 - F_1(\gamma; x_k)\right]\left[F_1(\gamma'; x_k) - F_2(\gamma'; x_k)\right]d\gamma' d\gamma\right] \\
&\qquad - A^{-1}\left[\int_0^t\int_t^1 L_{k-1}(\gamma)L_{k-1}(\gamma')\pi(\gamma)\pi(\gamma')\left[F_1(\gamma; x_k) - F_2(\gamma; x_k)\right]\left[1 - F_1(\gamma'; x_k)\right]d\gamma' d\gamma\right] \\
&\quad = A^{-1}\left[\int_0^t\int_t^1 L_{k-1}(\gamma)L_{k-1}(\gamma')\pi(\gamma)\pi(\gamma')\left\{\left[1 - F_1(\gamma; x_k)\right]\left[F_1(\gamma'; x_k) - F_2(\gamma'; x_k)\right]\right.\right. \\
&\qquad \left.\left. - \left[F_1(\gamma; x_k) - F_2(\gamma; x_k)\right]\left[1 - F_1(\gamma'; x_k)\right]\right\}d\gamma' d\gamma\right],
\end{aligned}
\tag{A.1}
$$

where

$$A = \iint_0^1 L_{k-1}(\gamma) L_{k-1}(\gamma') \left[1 - F_1(\gamma; x_k)\right] \left[F_1(\gamma'; x_k) - F_2(\gamma'; x_k)\right] \pi(\gamma) \pi(\gamma') d\gamma' d\gamma. \tag{A.2}$$

Since $\gamma \leq \gamma'$ and $(F_1 - F_2)/(1 - F_1)$ is monotonically decreasing in γ, then $[F_1(\gamma'; x_k) - F_2(\gamma'; x_k)]/[1 - F_1(\gamma'; x_k)] \leq [F_1(\gamma; x_k) - F_2(\gamma; x_k)]/[1 - F_1(\gamma; x_k)]$, which implies that $[1 - F_1(\gamma; x_k)][F_1(\gamma'; x_k) - F_2(\gamma'; x_k)] \leq [F_1(\gamma; x_k) - F_2(\gamma; x_k)][1 - F_1(\gamma'; x_k)]$. Hence, $\Pi_k(t; 0) \leq \Pi_k(t; 1)$, which implies that $\Pi_k^{-1}(\alpha; 0) \geq \Pi_k^{-1}(\alpha; 1)$, that is, $x_{k+1} \geq x'_{k+1}$. This completes the proof of Theorem 3.1. □

Proof of Theorem 3.2. Using the notation in the proof of Theorem 3.1, if $Y_k = 0$, then $L_k(\gamma) = L_{k-1}(\gamma)[1 - F_1(\gamma; x_k)]$, and if $Y_k = 2$, then $L_k(\gamma) = L_{k-1}(\gamma) F_2(\gamma; x_k)$. Since both F_1 and F_2 are monotonically decreasing in γ, then the result of the theorem follows from the proof of Theorem 1 in Tighiouart and Rogatko [23]. □

WinBUGS Code

```
# Assume that the toxicity response Y has 3 categories:
# Y = 1 if patient has grade 0 or 1 toxicity
# Y = 2 if patient has a grade 2 toxicity
# Y = 3 if patient has grade 3 or 4 toxicity, that is, DLT
# Dose is standardized between 0 and 1.
model {
for (i in 1 : N) {
# Likelihood
logit(eta[i,1])<- -(logit(rho1)) − (1/mtd)*(logit(theta) − logit(rho0))*X[i]
logit(eta[i,3])<- logit(rho0) + (1/mtd)*(logit(theta) − logit(rho0))*X[i]
# rho1 = P(grade 2 or more toxicity at initial dose)
# rho0 = P(DLT or grade 3 or 4 toxicity at initial dose)
# MTD = dose producing DLT in theta fraction of the population
                                }
for (i in 1 : N) {
p[i,1] <-eta[i,1]
p[i,3]<-eta[i,3]
p[i,2]<-1-(eta[i,1]+eta[i,3])
Y[i]~dcat(p[i,])
}
# Prior Distributions
rho0 ~ dunif(0, theta)
rho1 ~ dunif(rho0, 1)
```

mtd ~ dunif(0,1)
}
# Data
list($Y = c(1, 2, 2, 2, 3, 1, 2, 3, 3, 1, 2, 2, 1, 3)$,dose =
c(0.1,0.3262,0.3873,0.4390,0.4892,0.3810,0.4298,0.4681,0.3980,
0.3339,0.3650,0.3788,0.3986,0.4308),theta = 0.33333333, $N = 14$)
Initial values
list(rho0 = 0.05, rho1 = 0.15, mtd = 0.3).

Acknowledgments

This paper supported in part by the National Center for Research Resources, Grant UL1RR033176, and is now at the National Center for Advancing Translational Sciences, Grant UL1TR000124 (M. Tighiouart and A. Rogatko), Grant 5P01CA098912-05 (A. Rogatko) and P01 DK046763 (A. Rogatko). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References

[1] T. G. Roberts Jr., B. H. Goulart, L. Squitieri et al., "Trends in the risk and benefits to patients with cancer participating in phase 1 clinical trials," *Journal of the American Medical Association*, vol. 292, no. 17, pp. 2130–2140, 2004.
[2] National Cancer Institute, Common Toxicity Criteria for Adverse Events v3.0 (CTCAE), 2003, http://ctep.cancer.gov/reporting/ctc.html.
[3] J. O'Quigley, M. Pepe, and L. Fisher, "Continual reassessment method: a practical design for phase 1 clinical trials in cancer," *Biometrics*, vol. 46, no. 1, pp. 33–48, 1990.
[4] C. Gatsonis and J. B. Greenhouse, "Bayesian methods for phase I clinical trials," *Statistics in Medicine*, vol. 11, no. 10, pp. 1377–1389, 1992.
[5] S. D. Durham and N. Flournoy, "Random walks for quantile estimation," in *Statistical Design Theory and Related Topics*, pp. 467–476, Springer, New York, NY, USA, 1994.
[6] E. L. Korn, D. Midthune, T. T. Chen, L. V. Rubinstein, M. C. Christian, and R. M. Simon, "A comparison of two phase I trial designs," *Statistics in Medicine*, vol. 13, no. 18, pp. 1799–1806, 1994.
[7] J. Whitehead and H. Brunier, "Bayesian decision procedures for dose determining experiments," *Statistics in Medicine*, vol. 14, no. 9-10, pp. 885–893, 1995.
[8] J. Whitehead, "Bayesian decision procedures with application to dose-finding studies," *International Journal of Pharmaceutical Medicine*, vol. 11, no. 4, pp. 201–208, 1997.
[9] J. Babb, A. Rogatko, and S. Zacks, "Cancer phase I clinical trials: efficient dose escalation with overdose control," *Statistics in Medicine*, vol. 17, no. 10, pp. 1103–1120, 1998.
[10] M. Gasparini and J. Eisele, "A curve-free method for phase I clinical trials," *Biometrics*, vol. 56, no. 2, pp. 609–615, 2000.
[11] S. Mukhopadhyay, "Bayesian nonparametric inference on the dose level with specified response rate," *Biometrics*, vol. 56, no. 1, pp. 220–226, 2000.
[12] L. M. Haines, I. Perevozskaya, and W. F. Rosenberger, "Bayesian optimal designs for phase I clinical trials," *Biometrics*, vol. 59, no. 3, pp. 591–600, 2003.
[13] D. Faries, "Practical modifications of the continual reassessment method for phase I cancer clinical trials," *Journal of Biopharmaceutical Statistics*, vol. 4, no. 2, pp. 147–164, 1994.
[14] S. Moller, "An extension of the continual reassessment methods using a preliminary up-and-down design in a dose finding study in cancer patients, in order to investigate a greater range of doses," *Statistics in Medicine*, vol. 14, no. 9-10, pp. 911–922, 1995.
[15] S. N. Goodman, M. L. Zahurak, and S. Piantadosi, "Some practical improvements in the continual reassessment method for phase I studies," *Statistics in Medicine*, vol. 14, no. 11, pp. 1149–1161, 1995.

[16] J. O'Quigley and L. Z. Shen, "Continual reassessment method: a likelihood approach," *Biometrics*, vol. 52, no. 2, pp. 673–684, 1996.
[17] S. Piantadosi, J. D. Fisher, and S. Grossman, "Practical implementation of a modified continual reassessment method for dose-finding trials," *Cancer Chemotherapy and Pharmacology*, vol. 41, no. 6, pp. 429–436, 1998.
[18] Y. K. Cheung and R. Chappell, "Sequential designs for phase I clinical trials with late-onset toxicities," *Biometrics*, vol. 56, no. 4, pp. 1177–1182, 2000.
[19] B. E. Storer, "An evaluation of phase I clinical trial designs in the continuous dose-response setting," *Statistics in Medicine*, vol. 20, no. 16, pp. 2399–2408, 2001.
[20] D. H. Y. Leung and Y. G. Wang, "An extension of the continual reassessment method using decision theory," *Statistics in Medicine*, vol. 21, no. 1, pp. 51–63, 2002.
[21] S. Zacks, A. Rogatko, and J. Babb, "Optimal Bayesian-feasible dose escalation for cancer phase I trials," *Statistics & Probability Letters*, vol. 38, no. 3, pp. 215–220, 1998.
[22] M. Tighiouart, A. Rogatko, and J. S. Babb, "Flexible Bayesian methods for cancer phase I clinical trials. Dose escalation with overdose control," *Statistics in Medicine*, vol. 24, no. 14, pp. 2183–2196, 2005.
[23] M. Tighiouart and A. Rogatko, "Dose finding with escalation with overdose control (EWOC) in cancer clinical trials," *Statistical Science*, vol. 25, no. 2, pp. 217–226, 2010.
[24] S. Lonial, J. Kaufman, M. Tighiouart et al., "A phase I/II trial combining high-dose melphalan and autologous transplant with bortezomib for multiple myeloma: a dose- and schedule-finding study," *Clinical Cancer Research*, vol. 16, no. 20, pp. 5079–5086, 2010.
[25] R. Sinha, J. L. Kaufman, H. J. Khouri, N. King, P. J. Shenoy, C. Lewis et al., "A phase 1 dose escalation study of bortezomib combined with rituximab, cyclophosphamide, doxorubicin, modified vincristine, and prednisone for untreated follicular lymphoma and other low-grade B-cell lymphomas," *Cancer*, vol. 118, no. 14, pp. 3538–3548, 2011.
[26] H. Borghaei, K. Alpaugh, G. Hedlund et al., "Phase I dose escalation, pharmacokinetic and pharmacodynamic study of naptumomab estafenatox alone in patients with advanced cancer and with docetaxel in patients with advanced non-small-cell lung cancer," *Journal of Clinical Oncology*, vol. 27, no. 25, pp. 4116–4123, 2009.
[27] J. D. Cheng, J. S. Babb, C. Langer et al., "Individualized patient dosing in phase I clinical trials: the role of escalation with overdose control in PNU-214936," *Journal of Clinical Oncology*, vol. 22, no. 4, pp. 602–609, 2004.
[28] R. J. Schilder, J. M. Gallo, M. M. Millenson et al., "Phase I trial of multiple cycles of high-dose carboplatin, paclitaxel, and topotecan with peripheral-blood stem-cell support as front-line therapy," *Journal of Clinical Oncology*, vol. 19, no. 4, pp. 1183–1194, 2001.
[29] G. M. Freedman, N. J. Meropol, E. R. Sigurdson et al., "Phase I trial of preoperative hypofractionated intensity-modulated radiotherapy with incorporated boost and oral capecitabine in locally advanced rectal cancer," *International Journal of Radiation Oncology, Biology and Physics*, vol. 67, no. 5, pp. 1389–1393, 2007.
[30] J. S. Babb and A. Rogatko, "Patient specific dosing in a cancer phase I clinical trial," *Statistics in Medicine*, vol. 20, no. 14, pp. 2079–2090, 2001.
[31] M. Tighiouart and A. Rogatko, "Dose escalation with overdose control," in *Statistical Methods for Dose-Finding Experiments*, S. Chevret, Ed., pp. 173–188, John Wiley & Sons, 2006.
[32] M. Tighiouart and A. Rogatko, "Dose finding in oncology—parametric methods," in *Dose Finding in Drug Development*, N. Ting, Ed., pp. 59–72, Springer, New York, NY, USA, 2006.
[33] A. Rogatko and M. Tighiouart, "Novel and efficient translational clinical trial designs in advanced prostate cancer," in *Prostate Cancer*, L. Chung, W. Isaacs, and J. W. Simons, Eds., pp. 487–495, Humana Press, New Jersey, NJ, USA, 2007.
[34] N. Ting, *Dose Finding in Drug Development*, Springer, New York, NY, USA, 1st edition, 2006.
[35] S. Chevret, *Statistical Methods for Dose-Finding Experiments*, John Wiley & Sons, Chichester, UK, 2006.
[36] N. H. Gordon and J. K. V. Willson, "Using toxicity grades in the design and analysis of cancer phase I clinical trials," *Statistics in Medicine*, vol. 11, no. 16, pp. 2063–2075, 1992.
[37] C. Wang, T. T. Chen, and I. Tyan, "Designs for phase I cancer clinical trials with differentiation of graded toxicity," *Communications in Statistics*, vol. 29, no. 5-6, pp. 975–987, 2000.
[38] B. N. Bekele and P. F. Thall, "Dose-finding based on multiple toxicities in a soft tissue sarcoma trial," *Journal of the American Statistical Association*, vol. 99, no. 465, pp. 26–35, 2004.
[39] Z. Yuan, R. Chappell, and H. Bailey, "The continual reassessment method for multiple toxicity grades: a Bayesian quasi-likelihood approach," *Biometrics*, vol. 63, no. 1, pp. 173–179, 2007.

[40] R. F. Potthoff and S. L. George, "Flexible phase I clinical trials: allowing for nonbinary toxicity response and removal of other common limitations," *Statistics in Biopharmaceutical Research*, vol. 1, no. 3, pp. 213–228, 2009.

[41] B. N. Bekele, Y. S. Li, and Y. A. Ji, "Risk-group-specific dose finding based on an average toxicity score," *Biometrics*, vol. 66, no. 2, pp. 541–548, 2010.

[42] E. M. Van Meter, E. Garrett-Mayer, and D. Bandyopadhyay, "Proportional odds model for dose-finding clinical trial designs with ordinal toxicity grading," *Statistics in Medicine*, vol. 30, no. 17, pp. 2070–2080, 2011.

[43] E. M. Van Meter, E. Garrett-Mayer, and D. Bandyopadhyay, "Dose-finding clinical trial design for ordinal toxicity grades using the continuation ratio model: an extension of the continual reassessment method," *Clinical Trials*, vol. 9, no. 3, pp. 303–313, 2012.

[44] A. Iasonos, S. Zohar, and J. O'Quigley, "Incorporating lower grade toxicity information into dose finding designs," *Clinical Trials*, vol. 8, no. 4, pp. 370–379, 2011.

[45] S. M. Lee, B. Cheng, and Y. K. Cheung, "Continual reassessment method with multiple toxicity constraints," *Biostatistics*, vol. 12, no. 2, pp. 386–398, 2011.

[46] Z. Chen, M. Tighiouart, and J. Kowalski, "Dose escalation with overdose control using a quasi-continuous toxicity score in cancer phase I clinical trials," *Contemporary Clinical Trials*, vol. 33, no. 5, pp. 949–958, 2012.

[47] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *Journal of Chemical Physics*, vol. 21, no. 6, pp. 1087–1092, 1953.

[48] W. K. Hastings, "Monte carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.

[49] D. J. Lunn, A. Thomas, N. Best, and D. Spiegelhalter, "WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility," *Statistics and Computing*, vol. 10, no. 4, pp. 325–337, 2000.

[50] Y. K. Cheung, "Coherence principles in dose-finding studies," *Biometrika*, vol. 92, no. 4, pp. 863–873, 2005.

Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2012, Article ID 692725, 16 pages
doi:10.1155/2012/692725

Research Article

Number of Patients per Cohort and Sample Size Considerations Using Dose Escalation with Overdose Control

Mourad Tighiouart and André Rogatko

Samuel Oschin Comprehensive Cancer Institute, 8700 Beverly Boulevard, Los Angeles, CA 90048, USA

Correspondence should be addressed to Mourad Tighiouart, mourad.tighiouart@cshs.org

Received 29 June 2012; Accepted 15 September 2012

Academic Editor: Yichuan Zhao

The main objective of cancer phase I clinical trials is to determine a maximum tolerated dose (MTD) of a new experimental treatment. In practice, most of these trials are designed so that three patients per cohort are treated at the same dose level. In this paper, we compare the safety and efficiency of trials using the escalation with overdose control (EWOC) scheme designed with three or only one patient per cohort. We show through simulations that the number of patients per cohort does not impact the proportion of patients given therapeutic doses, safety of the trial, and efficiency of the estimate of the MTD. Additionally, we present guidelines and tabulated values on the number of patients needed to design a phase I cancer clinical trial using EWOC to achieve a given accuracy of the estimate of the MTD.

1. Introduction

Cancer phase I clinical trials are small studies whose main objective is to determine a maximum tolerated dose (MTD) of a new experimental drug or combination of known drugs for use in a phase II trial. Patients are typically accrued to the trial sequentially in cohorts of size m and dose level assignment to a given cohort of patients is dependent upon the dose levels and toxicity outcomes of the previously treated cohorts of patients. A large number of statistical methodologies which account for the sequential nature of the data generated by such designs have been proposed in the literature, see [1, 2] for a comprehensive review of such methods. In particular, the continual reassessment method (CRM) proposed by [3] and its modifications [4–8] and escalation with overdose control (EWOC) described in [9–15] are Bayesian adaptive designs that produce consistent sequences of doses and can be easily implemented in practice using published tutorials and free interactive software, see, for example, [16–19].

The work we present here has been motivated by the frequent requests by clinicians and review committees at Cancer Center Institutions the authors worked at on (1) the number of patients that should be included in each cohort, and (2) the number of patients required to conduct a phase I cancer clinical trial. Denote by ${}^{m}P$ a design that treats patients in successive cohorts of size m simultaneously at the same dose level. For a given fixed number of patients in the trial, an advantage of an ${}^{m}P$ design with $m > 1$ over a ${}^{1}P$ design is a shorter time of completion of the trial. However, it is not clear how the two designs compare with respect to safety of the trial and efficiency of the estimate of the MTD using EWOC. Goodman et al. [5] argue for the use of more than one patient per dose level in a modified version of the CRM to reduce the duration of the trial and toxicity incidence associated with the original CRM. In this paper, we compare a ${}^{3}P$ design with a ${}^{1}P$ design in terms of the number of patients given therapeutic doses, that is, doses in a neighborhood of the "true" MTD. In addition, safety of the trial and efficiency of the estimate of the MTD will be compared using extensive simulations.

The number of patients that are enrolled in a cancer phase I clinical trial is typically between 12 and 40 and trial duration depends on the study design and length of the study cycle to resolve toxicity outcome. An increasing number of clinicians we work with inquire about the number of patients they need to accrue in order to estimate the MTD with an acceptable degree of accuracy. We are not aware of any published methodologies for sample size determination (SSD) in cancer phase I clinical trials based on power calculation or precision of some Bayes estimates using either frequentist or Bayesian adaptive designs. As a point of fact, most sample size recommendations are based on prespecified stopping rules, see, for example, [20] on selecting the number of patients by considering different stopping rules using the CRM. Lin and Shih [21] and Ivanova [22] describe sample size recommendations based on the expected number of patients allocated to each dose selected from a set of prespecified dose levels.

In this paper, we address the SSD problem using the traditional approach; we estimate the sample size based on a desired accuracy of the Bayes estimate on the average. Specifically, we seek the smallest number of patients so that the posterior variance of the MTD on the average over all possible trials is no more than a specified margin. This procedure is not based on a specific stopping rule and consequently preserves the coherent nature of EWOC, see [15] for the coherence EWOC.

This paper is organized as follows. Section 2 describes dose escalation with overdose control using cohorts of size m. In Section 3, we present two criteria to sample size determination in this Bayesian setting. Comparisons of designs that treat cohorts of size $m > 1$ simultaneously over the ones that treat one patient at a time are presented in Section 4. In that section, we also give tabulated values relating the number of patients on the trial and the corresponding average posterior variance and length of the highest posterior density interval. Section 5 contains some concluding remarks and discussion.

2. ${}^{m}P$ Design Using EWOC

EWOC is a Bayesian adaptive design permitting precise determination of the MTD while directly controlling the likelihood of an overdose. It is the first statistical method to directly incorporate formal safety constraints into the design of cancer phase I clinical trials. Zacks et al. [10] and Tighiouart and Rogatko [15] discuss statistical properties and coherence of the method, and a comparison of EWOC with alternative phase I design methods is given

in [9]. Babb and Rogatko [11] provide a summary of Bayesian phase I design methods and Tighiouart et al. [12] studied the performance of EWOC under a richer class of prior distributions for the model parameters. The defining property of EWOC is that the expected proportion of patients treated at doses above the MTD is equal to a specified value α, the feasibility bound. This value is selected by the clinician and reflects his/her level of concern about overdosing. Zacks et al. [10] showed that among designs with this defining property, EWOC minimizes the average amount by which patients are underdosed. This means that EWOC approaches the MTD as rapidly as possible, while keeping the expected proportion of patients overdosed less than the value α. Zacks et al. [10] also showed that, as a trial progresses, the dose sequence defined by EWOC approaches the MTD (i.e., the sequence of recommended doses converges in probability to the MTD). Eventually, all patients beyond a certain time would be treated at doses sufficiently close to the MTD.

EWOC has been used to design over a dozen of phase I studies approved by the Research Review Committee and the Institute Review Board of the Fox Chase Cancer Center, Philadelphia, Winship Cancer Institute, Atlanta, and Cedars Sinai Medical Center, Los Angeles (see [23–29] for some of the published trials).

We adopt the-logistic-based model to represent the dose-toxicity relationship the following:

$$\text{Prob}(\text{DLT} \mid \text{Dose} = x) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)}, \tag{2.1}$$

where $(\beta_0, \beta_1) \in (-\infty, \infty) \times (0, \infty)$ so that the probability of dose limiting toxicity (DLT) is an increasing function of dose. The MTD γ is defined as the dose expected to produce DLT in a specified proportion θ of patients. Let ρ_0 be the probability of a DLT at the starting dose. To facilitate interpretation of model parameters by the clinicians, we further parameterize model (2.1) in terms of (ρ_0, γ), see [9, 12] for more details. Suppose we plan to enroll n patients in the trial in cohorts of size m. Dose levels in the trial are selected in the interval $[X_{\min}, X_{\max}]$ and an ${}^mP$ design proceeds as follows. We first specify prior distributions for ρ_0 and γ. Then, the first cohort of m patients receives the dose $x_1 = X_{\min}$. Let d_1 be the number of toxicities observed among the first m patients. The likelihood given the observed data thus far is

$$L_1(\rho_0, \gamma \mid D_1) = p(\rho_0, \gamma, x_1)^{d_1} \left(1 - p(\rho_0, \gamma, x_1)\right)^{m - d_1}, \tag{2.2}$$

where

$$p(\rho_0, \gamma, x_1) = \frac{\exp\left\{\ln\left[\rho_0/(1-\rho_0)\right] + \ln\left[\theta(1-\rho_0)/\rho_0(1-\theta)\right](x_1/\gamma)\right\}}{1 + \exp\left\{\ln\left[\rho_0/(1-\rho_0)\right] + \ln\left[\theta(1-\rho_0)/\rho_0(1-\theta)\right](x_1/\gamma)\right\}} \tag{2.3}$$

and $D_1 = \{(x_1, d_1)\}$. Let $\Pi_1(x)$ be the marginal posterior cumulative distribution function (cdf) of the MTD γ given D_1. The second cohort of m patients receives the dose $x_2 = \Pi_1^{-1}(\alpha)$ so that the posterior probability of exceeding the MTD is equal to the feasibility bound α. In

general, the likelihood of the data after observing the toxicity outcomes of the ith cohort of m patients is

$$L_i(\rho_0, \gamma \mid D_i) = \prod_{j=1}^{i} p(\rho_0, \gamma, x_j)^{d_j} (1 - p(\rho_0, \gamma, x_j))^{m-d_j}, \tag{2.4}$$

where x_j is the dose assigned to the jth cohort of m patients, $p(\rho_0, \gamma, x_j)$ is given by (2.3) with x_1 replaced by x_j, and $D_i = \{(x_1, d_1), \ldots, (x_i, d_i)\}$. The $(i+1)$st cohort of m patients receives the dose $x_{i+1} = \Pi_i^{-1}(\alpha)$ where $\Pi_i(x)$ is the marginal posterior cdf of γ given D_i. This process is repeated until a total of k cohorts are enrolled in the trial. This completes the description of an ${}^mP$ design. For a given sample size n, we propose to compare the performance of a ${}^1P$ with a ${}^3P$ design by estimating the percent of patients treated within a neighborhood of the true MTD. Other comparisons include safety and efficiency of the estimate of the MTD under the two designs.

3. Sample Size Determination

An increasing number of clinicians inquire about the number of patients they need to accrue in the design of cancer phase I trials to achieve a specific goal. Sample size recommendation based on the expected number of patients treated at each dose level in "3 + 3" designs and $A + B$ designs have been studied in [21, 22], respectively. However, these methods apply to a prespecified set of discrete doses and it is not clear how they can be applied to continuous doses. Unlike the frequentist approach, there is no consensus on a specific Bayesian method for the SSD problem, see Adcock [30] for a review of Bayesian approaches. In this paper, we present numerical results based on the posterior variance of the MTD and highest posterior density (HPD) interval, see [31].

Denote by $\mathrm{Var}(\gamma \mid D_n)$ the posterior variance of the MTD given that n patients have been accrued to the trial. The first criterion is to find the smallest n that satisfies

$$E_{D_n}\left[\mathrm{Var}(\gamma \mid D_n)\right] \leq \eta, \tag{3.1}$$

where the above expectation is taken with respect to the marginal distribution of the data and η is specified by the clinician. In other words, we require an estimate of the MTD within a given accuracy as measured by the posterior variance on the average overall possible trials. In the second criteria, we seek the smallest n such that

$$E_{D_n}[l(D_n)] \leq d, \tag{3.2}$$

where $l(D_n)$ is the length of the HPD interval $(a, a + l(D_n))$ determined by the constraint on the coverage probability

$$P\left[\gamma \in (a, a + l(D_n)) \mid D_n\right] = 1 - \alpha_1. \tag{3.3}$$

This is also known as the average length criteria (ALC) because for each realization of a trial D_n, the corresponding HPD interval is determined by (3.3) and the lengths of these

HPD intervals are averaged out with respect to the marginal distribution of the data in (3.2). The tolerance values of the average length of the HPD interval d and coverage probability 1-α_1 are prespecified by the clinician. Since both the posterior distribution of the MTD and marginal distribution of the data are intractable, Monte Carlo averages were used to estimate the left hand sides of (3.1) and (3.2). Details on the computation of $\mathrm{Var}(\gamma \mid D_n)$ and $l(D_n)$ can be found in [9, 18].

4. Numerical Results

The simulation results presented below all assume that the feasibility bound $\alpha = 0.25$ and that the dose levels are standardized so that the starting dose for each trial is $x_1 = 0$ and all subsequent dose levels are selected from the unit interval. Independent uniform prior distributions were put on the parameters ρ_0 and γ on the intervals $[0, \theta]$, $[0, 1]$, respectively.

4.1. Comparison of Designs ${}^3P$ with ${}^1P$

We simulate trials under different scenarios corresponding to different values of ρ_0 and γ. For the ${}^1P$ design, the first patient receives dose 0 and the next dose x_2 is determined as described in Section 2. The second response y_2 is then generated from the logistic model (2.3). This process is repeated until a trial of n patients is generated. The same process applies to the ${}^3P$ design except that 3 patients will be given the same dose at each stage of the trial and 3 responses are generated from model (2.3) independently instead of 1. Since $0 \leq \rho_0 \leq \theta$ and $0 \leq \gamma \leq 1$, we considered 12 scenarios corresponding to combinations of three values of ρ_0, $\{\theta/4, \theta/2, 3\theta/4\}$ with four values of γ, 0.2, 0.4, 0.6, and 0.8. We will refer to $\theta/4$, $\theta/2$, $3\theta/4$ as low, intermediate, and high values for ρ_0, respectively. Similarly, 0.2 and 0.4 will be referred to as low values for the MTD γ and 0.6 and 0.8 as high values. The same value $\theta = 0.3$ was used in all simulations. For each design, each sample size $n = 12$, 18, 24, 30, and each combination of (ρ_0, γ), we simulated 5000 trials and calculated the proportions of patients given therapeutic doses, that is, doses in an ε-neighborhood of the true MTD, for $\varepsilon = 0.05, 0.1, 0.15, 0.2$.

Table 1 gives the estimated proportions of patients given doses in an ε-neighborhood of the true MTD under designs ${}^1P$ and ${}^3P$ and the difference in these proportions between the two designs for low values of the true MTD γ and different sample sizes. Table 2 gives the corresponding estimates for high values of the true MTD and Table 3 gives the average of these estimates across the 12 combination of (ρ_0, γ). For low values of the true MTD, design ${}^1P$ assigns more patients to doses near the MTD than design ${}^3P$ in general and the difference can be as high as 16% for $\varepsilon = 0.05$, $(\rho_0, \gamma) = (0.4, 0.075)$, and $n = 12$. For high values of the MTD, Table 2 shows that design ${}^1P$ always assigns more patients to doses near the MTD than design ${}^3P$ and the highest difference is about 16% for $\varepsilon = 0.2$, $(\rho_0, \gamma) = (0.6, 0.075)$, and $n = 12$. The estimated difference in the proportions of patients given doses in an ε-neighborhood of the true MTD between the ${}^1P$ design and ${}^3P$ design averaged across the 12 entertained scenarios for (ρ_0, γ) for different sample sizes show that the proportion of patients given therapeutic doses under design ${}^1P$ is always greater than the corresponding proportion under design ${}^3P$, the largest of these differences is about 5%. The practical impact of this difference is unimportant because of the relatively small number of patients involved in phase I cancer clinical trials. In Tables 4 and 5, we present differences in (i) the proportions of patients exhibiting DLT, (ii) the proportions of patients given doses above the "true" MTD,

Table 1: Estimated proportions of patients given doses in an ε-neighborhood of the true MTD under designs ${}^1P$ and ${}^3P$ and differences between these proportions for small values of γ and selected values of ρ_0.

	Sample size											
	12			18			24			30		
	${}^1P$	${}^3P$	Diff.	${}^1P$	${}^3P$	Diff.	${}^1P$	${}^3P$	Diff.	${}^1P$	${}^3P$	Diff.
$(\gamma, \rho_0) = (0.2, 0.075)$												
$\varepsilon = 0.20$	0.9768	0.9829	−0.0061	0.9826	0.9854	−0.0028	0.9864	0.9890	−0.0026	0.9897	0.9910	−0.0013
$\varepsilon = 0.15$	0.9065	0.9352	−0.0287	0.9290	0.9465	−0.0175	0.9433	0.9598	−0.0165	0.9542	0.9661	−0.0119
$\varepsilon = 0.10$	0.7525	0.7004	0.0521	0.7958	0.7440	0.0518	0.8261	0.7897	0.0364	0.8569	0.8210	0.0359
$\varepsilon = 0.05$	0.3634	0.2521	0.1114	0.4306	0.3309	0.0997	0.4747	0.3997	0.0750	0.5265	0.4444	0.0821
$(\gamma, \rho_0) = (0.2, 0.15)$												
$\varepsilon = 0.20$	0.9340	0.9603	−0.0262	0.9430	0.9492	−0.0062	0.9517	0.9571	−0.0054	0.9592	0.9629	−0.0037
$\varepsilon = 0.15$	0.8153	0.8839	−0.0687	0.8378	0.8760	−0.0382	0.8610	0.8896	−0.0287	0.8784	0.8958	−0.0174
$\varepsilon = 0.10$	0.6297	0.6729	−0.0431	0.6507	0.6643	−0.0136	0.6831	0.6924	−0.0092	0.7032	0.7105	−0.0074
$\varepsilon = 0.05$	0.2752	0.2298	0.0454	0.3142	0.2659	0.0483	0.3478	0.3152	0.0327	0.3713	0.3477	0.0236
$(\gamma, \rho_0) = (0.2, 0.225)$												
$\varepsilon = 0.20$	0.8658	0.9351	−0.0693	0.8659	0.9032	−0.0373	0.8772	0.8926	−0.0155	0.8838	0.8984	−0.0146
$\varepsilon = 0.15$	0.7122	0.8463	−0.1341	0.7190	0.8036	−0.0846	0.7359	0.7900	−0.0541	0.7438	0.7931	−0.0493
$\varepsilon = 0.10$	0.5256	0.6533	−0.1277	0.5219	0.6063	−0.0845	0.5390	0.5865	−0.0475	0.5448	0.5915	−0.0467
$\varepsilon = 0.05$	0.2154	0.2102	0.0052	0.2279	0.2276	0.0003	0.2549	0.2371	0.0178	0.2626	0.2604	0.0022
$(\gamma, \rho_0) = (0.4, 0.075)$												
$\varepsilon = 0.20$	0.7855	0.7121	0.0734	0.8272	0.7762	0.0510	0.8542	0.8180	0.0362	0.8784	0.8504	0.0280
$\varepsilon = 0.15$	0.6939	0.5833	0.1106	0.7269	0.6541	0.0728	0.7477	0.6968	0.0509	0.7825	0.7376	0.0449
$\varepsilon = 0.10$	0.4885	0.5258	−0.0373	0.5267	0.5478	−0.0211	0.5474	0.5647	−0.0173	0.5874	0.5906	−0.0032
$\varepsilon = 0.05$	0.2773	0.1144	0.1629	0.2956	0.1759	0.1197	0.3050	0.2165	0.0885	0.3282	0.2506	0.0776
$(\gamma, \rho_0) = (0.4, 0.15)$												
$\varepsilon = 0.20$	0.7038	0.6680	0.0358	0.7433	0.7201	0.0231	0.7587	0.7541	0.0046	0.7748	0.7814	−0.0066
$\varepsilon = 0.15$	0.6031	0.5135	0.0897	0.6247	0.5757	0.0490	0.6348	0.6103	0.0246	0.6489	0.6417	0.0072
$\varepsilon = 0.10$	0.4129	0.4371	−0.0242	0.4348	0.4531	−0.0184	0.4458	0.4722	−0.0264	0.4580	0.4903	−0.0323
$\varepsilon = 0.05$	0.2332	0.1125	0.1207	0.2388	0.1648	0.0740	0.2416	0.1835	0.0581	0.2489	0.2100	0.0389
$(\gamma, \rho_0) = (0.4, 0.225)$												
$\varepsilon = 0.20$	0.6383	0.6107	0.0276	0.6484	0.6550	−0.0066	0.6573	0.6826	−0.0252	0.6608	0.6859	−0.0251
$\varepsilon = 0.15$	0.5377	0.4446	0.0932	0.5364	0.5042	0.0322	0.5389	0.5313	0.0076	0.5322	0.5349	−0.0027
$\varepsilon = 0.10$	0.3605	0.3556	0.0049	0.3651	0.3819	−0.0168	0.3680	0.3921	−0.0241	0.3627	0.3912	−0.0286
$\varepsilon = 0.05$	0.2008	0.1036	0.0972	0.1988	0.1467	0.0521	0.1982	0.1583	0.0400	0.1926	0.1665	0.0261

Table 2: Estimated proportions of patients given doses in an ε-neighborhood of the true MTD under designs ${}^1P$ and ${}^3P$ and differences between these proportions for high values of γ and selected values of ρ_0.

	Sample size											
	12			18			24			30		
	${}^1P$	${}^3P$	Diff.	${}^1P$	${}^3P$	Diff.	${}^1P$	${}^3P$	Diff.	${}^1P$	${}^3P$	Diff.
$(\gamma, \rho_0) = (0.6, 0.075)$												
$\varepsilon = 0.20$	0.4140	0.2552	0.1588	0.4940	0.4012	0.0928	0.5503	0.4825	0.0679	0.6103	0.5418	0.0685
$\varepsilon = 0.15$	0.2790	0.2224	0.0566	0.3542	0.3137	0.0406	0.4092	0.3814	0.0278	0.4692	0.4339	0.0353
$\varepsilon = 0.10$	0.1702	0.0455	0.1247	0.2273	0.1544	0.0729	0.2696	0.2153	0.0543	0.3172	0.2620	0.0552
$\varepsilon = 0.05$	0.0746	0.0448	0.0298	0.1090	0.0830	0.0260	0.1321	0.1132	0.0189	0.1567	0.1390	0.0177
$(\gamma, \rho_0) = (0.6, 0.15)$												
$\varepsilon = 0.20$	0.3104	0.1702	0.1402	0.3744	0.2755	0.0989	0.4140	0.3500	0.0641	0.4315	0.3768	0.0547
$\varepsilon = 0.15$	0.1988	0.1463	0.0525	0.2590	0.2008	0.0583	0.2965	0.2553	0.0412	0.3101	0.2770	0.0331
$\varepsilon = 0.10$	0.1173	0.0267	0.0906	0.1588	0.0920	0.0668	0.1897	0.1362	0.0535	0.2005	0.1548	0.0457
$\varepsilon = 0.05$	0.0499	0.0263	0.0236	0.0728	0.0474	0.0254	0.0909	0.0682	0.0226	0.0976	0.0770	0.0206
$(\gamma, \rho_0) = (0.6, 0.225)$												
$\varepsilon = 0.20$	0.2428	0.1095	0.1333	0.2883	0.1983	0.0900	0.3079	0.2444	0.0635	0.3331	0.2694	0.0636
$\varepsilon = 0.15$	0.1500	0.0924	0.0576	0.1932	0.1416	0.0516	0.2146	0.1716	0.0430	0.2336	0.1895	0.0441
$\varepsilon = 0.10$	0.0864	0.0157	0.0707	0.1171	0.0626	0.0546	0.1350	0.0881	0.0469	0.1484	0.1016	0.0468
$\varepsilon = 0.05$	0.0356	0.0153	0.0203	0.0531	0.0321	0.0209	0.0646	0.0452	0.0194	0.0721	0.0489	0.0231
$(\gamma, \rho_0) = (0.8, 0.075)$												
$\varepsilon = 0.20$	0.0582	0.0052	0.0530	0.1270	0.0736	0.0534	0.1886	0.1342	0.0544	0.2354	0.1852	0.0502
$\varepsilon = 0.15$	0.0220	0.0000	0.0220	0.0619	0.0431	0.0188	0.1039	0.0803	0.0235	0.1414	0.1145	0.0269
$\varepsilon = 0.10$	0.0017	0.0000	0.0017	0.0235	0.0092	0.0143	0.0478	0.0310	0.0168	0.0724	0.0529	0.0195
$\varepsilon = 0.05$	0.0000	0.0000	0.0000	0.0055	0.0014	0.0040	0.0165	0.0094	0.0072	0.0275	0.0180	0.0095
$(\gamma, \rho_0) = (0.8, 0.15)$												
$\varepsilon = 0.20$	0.0344	0.0034	0.0310	0.0729	0.0349	0.0380	0.1041	0.0630	0.0411	0.1288	0.0859	0.0429
$\varepsilon = 0.15$	0.0126	0.0000	0.0126	0.0336	0.0179	0.0158	0.0554	0.0360	0.0194	0.0718	0.0460	0.0258
$\varepsilon = 0.10$	0.0008	0.0000	0.0008	0.0120	0.0035	0.0085	0.0236	0.0131	0.0105	0.0336	0.0189	0.0147
$\varepsilon = 0.05$	0.0000	0.0000	0.0000	0.0027	0.0005	0.0022	0.0075	0.0037	0.0038	0.0119	0.0062	0.0057
$(\gamma, \rho_0) = (0.8, 0.225)$												
$\varepsilon = 0.20$	0.0219	0.0014	0.0205	0.0410	0.0166	0.0244	0.0559	0.0323	0.0236	0.0752	0.0474	0.0278
$\varepsilon = 0.15$	0.0079	0.0000	0.0079	0.0187	0.0095	0.0091	0.0272	0.0166	0.0106	0.0404	0.0262	0.0142
$\varepsilon = 0.10$	0.0005	0.0000	0.0005	0.0067	0.0020	0.0047	0.0112	0.0056	0.0056	0.0196	0.0111	0.0085
$\varepsilon = 0.05$	0.0000	0.0000	0.0000	0.0017	0.0003	0.0014	0.0037	0.0019	0.0018	0.0076	0.0037	0.0038

Table 3: Estimated proportions of patients given doses in an ε-neighborhood of the true MTD under designs ${}^1P$ and ${}^3P$ and differences between these proportions on the average.

		Sample size n							
		12		18		24		30	
ε	0.05	${}^1P$	0.1436	${}^1P$	0.1625	${}^1P$	0.1781	${}^1P$	0.1920
		${}^3P$	0.0924	${}^3P$	0.1230	${}^3P$	0.1459	${}^3P$	0.1644
		diff.	**0.0512**	**diff.**	**0.0395**	**diff.**	**0.0322**	**diff.**	**0.0276**
	0.10	${}^1P$	0.2956	${}^1P$	0.3200	${}^1P$	0.3405	${}^1P$	0.3587
		${}^3P$	0.2860	${}^3P$	0.3101	${}^3P$	0.3322	${}^3P$	0.3497
		diff.	**0.0096**	**diff.**	**0.0099**	**diff.**	**0.0083**	**diff.**	**0.0090**
	0.15	${}^1P$	0.4115	${}^1P$	0.4412	${}^1P$	0.4640	${}^1P$	0.4839
		${}^3P$	0.3888	${}^3P$	0.4239	${}^3P$	0.4516	${}^3P$	0.4714
		diff.	**0.0227**	**diff.**	**0.0173**	**diff.**	**0.0124**	**diff.**	**0.0125**
	0.20	${}^1P$	0.4988	${}^1P$	0.5340	${}^1P$	0.5589	${}^1P$	0.5801
		${}^3P$	0.4517	${}^3P$	0.4991	${}^3P$	0.5333	${}^3P$	0.5564
		diff.	**0.0471**	**diff.**	**0.0349**	**diff.**	**0.0256**	**diff.**	**0.0237**

(iii) the bias, and (iv) the mean square error between the ${}^1P$ and ${}^3P$ designs. Table 6 gives the average values of these statistics, averaged across the 12 entertained scenarios for (ρ_0, γ). Based on (i) and (ii), the results indicate that the two designs are equally safe and that no practical gain is achieved in terms of the efficiency of the estimate of the MTD according to (iii) and (iv). From an ethical point of view, we recommend the ${}^1P$ design to prevent the occurrence of three simultaneous DLTs if we were to use the ${}^3P$ design. This should be discussed with the clinician after assessing the importance of the length of the trial.

4.2. Sample Size Determination

In this section, we present tabulated values for average posterior standard deviation of the MTD and average length HPD interval that are achieved for even sample sizes $n = 6, \ldots, 40$ and selected values of θ, the target probability of DLT. Table 7 summarizes the results for $\theta = 0.3$. For a given sample size n, each entry in the table was calculated according to the following algorithm:

Set $j = 1$.

(i) Generate $(\rho_{0,j}, \gamma_j) \sim \mathit{Uniform}\ [0, \theta] \times [0, 1]$ and independently.

(ii) Simulate a trial of n patients $D_{n,j}$ according to the EWOC algorithm described in Section 4.1 with $(\rho_{0,j}, \gamma_j)$ as the true model parameters.

(iii) Calculate the posterior variance $\mathrm{Var}(\gamma \mid D_{n,j})$ and HPD $(a_j, a_j + l(D_{n,j}))$ using (3.3).

(iv) Repeat steps (i)–(iii) for $j = 2, \ldots, M$.

Table 4: Estimated proportions of patients exhibiting DLTs, treated above the MTD, MSE, and bias of the MTD under designs ^1P and ^3P for small values of γ and selected values of ρ_0.

	Sample size											
	12			18			24			30		
	^1P	^3P	Diff.	^1P	^3P	Diff.	^1P	^3P	Diff.	^1P	^3P	Diff.
$(\gamma, \rho_0) = (0.2, 0.075)$												
Proportion of DLTs	0.3389	0.3456	−0.0067	0.3271	0.3346	−0.0075	0.3192	0.3254	−0.0062	0.3136	0.3173	−0.0037
Proportion above the MTD	0.4694	0.5095	−0.0401	0.4493	0.4701	−0.0208	0.4274	0.4419	−0.0145	0.4052	0.4140	−0.0088
MSE	0.0457	0.0467	−0.0010	0.0389	0.0365	0.0024	0.0338	0.0311	0.0027	0.0311	0.0274	0.0036
Bias	0.2013	0.1907	0.0106	0.1857	0.1676	0.0181	0.1738	0.1556	0.0182	0.1663	0.1467	0.0196
$(\gamma, \rho_0) = (0.2, 0.15)$												
Proportion of DLTs	0.3346	0.3289	0.0058	0.3305	0.3311	−0.0006	0.3262	0.3306	−0.0044	0.3211	0.3279	−0.0068
Proportion above the MTD	0.5219	0.5436	−0.0217	0.5185	0.5512	−0.0327	0.5073	0.5376	−0.0303	0.4928	0.5293	−0.0366
MSE	0.0558	0.0661	−0.0102	0.0468	0.0542	−0.0074	0.0402	0.0447	−0.0045	0.0367	0.0393	−0.0026
Bias	0.2164	0.2315	−0.0151	0.1972	0.2076	−0.0104	0.1825	0.1878	−0.0054	0.1741	0.1760	−0.0019
$(\gamma, \rho_0) = (0.2, 0.225)$												
Proportion of DLTs	0.3200	0.3121	0.0080	0.3210	0.3164	0.0047	0.3200	0.3183	0.0017	0.3181	0.3190	−0.0009
Proportion above the MTD	0.5573	0.5505	0.0069	0.5618	0.5766	−0.0148	0.5470	0.5839	−0.0369	0.5527	0.5836	−0.0310
MSE	0.0734	0.0849	−0.0115	0.0636	0.0742	−0.0106	0.0559	0.0672	−0.0113	0.0513	0.0593	−0.0080
Bias	0.2403	0.2625	−0.0222	0.2206	0.2416	−0.0210	0.2049	0.2271	−0.0222	0.1952	0.2126	−0.0174
$(\gamma, \rho_0) = (0.4, 0.075)$												
Proportion of DLTs	0.2461	0.2222	0.0239	0.2545	0.2414	0.0131	0.2594	0.2511	0.0082	0.2630	0.2581	0.0049
Proportion above the MTD	0.2444	0.1786	0.0658	0.2803	0.2474	0.0329	0.2807	0.2701	0.0105	0.2954	0.2843	0.0111
MSE	0.0276	0.0316	−0.0040	0.0262	0.0291	−0.0029	0.0243	0.0266	−0.0023	0.0232	0.0242	−0.0011
Bias	0.1371	0.1493	−0.0122	0.1353	0.1403	−0.0050	0.1311	0.1342	−0.0031	0.1302	0.1292	0.0010
$(\gamma, \rho_0) = (0.4, 0.15)$												
Proportion of DLTs	0.2583	0.2408	0.0175	0.2626	0.2548	0.0078	0.2669	0.2607	0.0063	0.2689	0.2646	0.0043
Proportion above the MTD	0.2202	0.1352	0.0850	0.2493	0.1917	0.0576	0.2663	0.2294	0.0369	0.2673	0.2489	0.0184
MSE	0.0278	0.0308	−0.0030	0.0269	0.0291	−0.0022	0.0254	0.0279	−0.0025	0.0237	0.0266	−0.0029
Bias	0.1167	0.1336	−0.0169	0.1153	0.1266	−0.0113	0.1107	0.1218	−0.0111	0.1070	0.1194	−0.0124
$(\gamma, \rho_0) = (0.4, 0.225)$												
Proportion of DLTs	0.2749	0.2692	0.0057	0.2798	0.2737	0.0060	0.2803	0.2761	0.0042	0.2809	0.2782	0.0027
Proportion above the MTD	0.2047	0.1037	0.1010	0.2216	0.1646	0.0570	0.2447	0.1987	0.0461	0.2461	0.2098	0.0364
MSE	0.0291	0.0281	0.0009	0.0275	0.0290	−0.0015	0.0284	0.0298	−0.0014	0.0273	0.0284	−0.0012
Bias	0.0941	0.1089	−0.0148	0.0833	0.1051	−0.0219	0.0814	0.0997	−0.0184	0.0758	0.0942	−0.0184

Table 5: Estimated proportions of patients exhibiting DLTs, treated above the MTD, MSE, and bias of the MTD under designs ${}^1P$ and ${}^3P$ for high values of γ and selected values of ρ_0.

	Sample size											
	12			18			24			30		
	${}^1P$	${}^3P$	Diff.	${}^1P$	${}^3P$	Diff.	${}^1P$	${}^3P$	Diff.	${}^1P$	${}^3P$	Diff.
$(\gamma, \rho_0) = (0.6, 0.075)$												
Proportion of DLTs	0.1866	0.1657	0.0209	0.2010	0.1863	0.0147	0.2108	0.1996	0.0112	0.2168	0.2080	0.0088
Proportion above the MTD	0.0258	0.0001	0.0257	0.0529	0.0341	0.0188	0.0787	0.0603	0.0185	0.0974	0.0817	0.0157
MSE	0.0099	0.0092	0.0008	0.0102	0.0104	−0.0002	0.0103	0.0102	0.0001	0.0098	0.0107	−0.0008
Bias	0.0065	0.0097	−0.0033	0.0195	0.0238	−0.0043	0.0287	0.0331	−0.0044	0.0397	0.0401	−0.0003
$(\gamma, \rho_0) = (0.6, 0.15)$												
Proportion of DLTs	0.2250	0.2121	0.0129	0.2286	0.2226	0.0059	0.2342	0.2270	0.0072	0.2369	0.2341	0.0027
Proportion above the MTD	0.0184	0.0000	0.0184	0.0384	0.0188	0.0197	0.0535	0.0339	0.0196	0.0666	0.0450	0.0216
MSE	0.0178	0.0140	0.0038	0.0169	0.0152	0.0017	0.0167	0.0152	0.0015	0.0168	0.0153	0.0015
Bias	−0.0449	−0.0356	−0.0092	−0.0317	−0.0278	−0.0039	−0.0271	−0.0177	−0.0093	−0.0223	−0.0173	−0.0050
$(\gamma, \rho_0) = (0.6, 0.225)$												
Proportion of DLTs	0.2609	0.2570	0.0039	0.2604	0.2593	0.0011	0.2643	0.2601	0.0042	0.2632	0.2636	−0.0004
Proportion above the MTD	0.0132	0.0001	0.0132	0.0291	0.0120	0.0171	0.0392	0.0236	0.0156	0.0481	0.0323	0.0158
MSE	0.0289	0.0228	0.0061	0.0306	0.0251	0.0055	0.0321	0.0250	0.0071	0.0327	0.0267	0.0060
Bias	−0.0898	−0.0794	−0.0104	−0.0870	−0.0762	−0.0109	−0.0905	−0.0718	−0.0186	−0.0863	−0.0746	−0.0117
$(\gamma, \rho_0) = (0.8, 0.075)$												
Proportion of DLTs	0.1549	0.1387	0.0162	0.1678	0.1541	0.0136	0.1774	0.1663	0.0111	0.1846	0.1762	0.0084
Proportion above the MTD	0.0000	0.0000	0.0000	0.0001	0.0000	0.0001	0.0022	0.0009	0.0013	0.0052	0.0032	0.0020
MSE	0.0347	0.0349	−0.0003	0.0276	0.0273	0.0003	0.0226	0.0221	0.0004	0.0184	0.0192	−0.0008
Bias	−0.1575	−0.1622	0.0047	−0.1323	−0.1326	0.0003	−0.1128	−0.1125	−0.0003	−0.0977	−0.0996	0.0019
$(\gamma, \rho_0) = (0.8, 0.15)$												
Proportion of DLTs	0.2042	0.1933	0.0109	0.2099	0.2026	0.0073	0.2143	0.2083	0.0060	0.2179	0.2135	0.0044
Proportion above the MTD	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0012	0.0003	0.0008	0.0020	0.0013	0.0007
MSE	0.0641	0.0592	0.0049	0.0578	0.0542	0.0037	0.0543	0.0503	0.0040	0.0514	0.0464	0.0050
Bias	−0.2192	−0.2159	−0.0034	−0.2025	−0.1983	−0.0042	−0.1911	−0.1869	−0.0041	−0.1818	−0.1776	−0.0042
$(\gamma, \rho_0) = (0.8, 0.225)$												
Proportion of DLTs	0.2502	0.2469	0.0032	0.2533	0.2505	0.0028	0.2555	0.2522	0.0033	0.2548	0.2532	0.0016
Proportion above the MTD	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0004	0.0002	0.0001	0.0013	0.0008	0.0005
MSE	0.0979	0.0877	0.0102	0.0995	0.0900	0.0095	0.1011	0.0883	0.0128	0.0989	0.0887	0.0102
Bias	−0.2760	−0.2673	−0.0087	−0.2750	−0.2640	−0.0110	−0.2747	−0.2595	−0.0152	−0.2685	−0.2564	−0.0121

Table 6: Estimated proportions of patients exhibiting DLTs, treated above the MTD, MSE, and bias of the MTD under designs ${}^1P$ and ${}^3P$ and differences between these proportions on the average.

	Sample size n							
	12		18		24		30	
Proportion of DLTs	${}^1P$	0.2546	${}^1P$	0.2580	${}^1P$	0.2607	${}^1P$	0.2616
	${}^3P$	0.2444	${}^3P$	0.2523	${}^3P$	0.2563	${}^3P$	0.2595
	diff.	**0.0102**	**diff.**	**0.0057**	**diff.**	**0.0044**	**diff.**	**0.0021**
Proportion above the MTD	${}^1P$	0.1895	${}^1P$	0.2001	${}^1P$	0.2040	${}^1P$	0.2067
	${}^3P$	0.1685	${}^3P$	0.1888	${}^3P$	0.1984	${}^3P$	0.2029
	diff.	**0.0210**	**diff.**	**0.0113**	**diff.**	**0.0056**	**diff.**	**0.0038**
MSE	${}^1P$	0.0427	${}^1P$	0.0394	${}^1P$	0.0371	${}^1P$	0.0351
	${}^3P$	0.0429	${}^3P$	0.0395	${}^3P$	0.0365	${}^3P$	0.0344
	diff.	**−0.0002**	**diff.**	**−0.0001**	**diff.**	**0.0006**	**diff.**	**0.0007**
Bias	${}^1P$	0.0186	${}^1P$	0.0190	${}^1P$	0.0181	${}^1P$	0.0193
	${}^3P$	0.0271	${}^3P$	0.0261	${}^3P$	0.0259	${}^3P$	0.0244
	diff.	**−0.0085**	**diff.**	**−0.0071**	**diff.**	**−0.0078**	**diff.**	**−0.0051**

The left hand sides of (3.1) and (3.2) are estimated by

$$E_{D_n}\left[\mathrm{Var}(\gamma \mid D_n)\right] \approx \frac{1}{M}\sum_{j=1}^{M}\mathrm{Var}(\gamma \mid D_{n,j}),$$
$$E_{D_n}[l(D_n)] \approx \frac{1}{M}\sum_{j=1}^{M} l(D_{n,j}). \tag{4.1}$$

In the numerical results presented here, we took $M = 1000$. When $\theta = 0.3$, Table 7 shows that with 6 patients, we can estimate the MTD with an average posterior standard deviation equal to 25% of the range of the dose and that a 17% decrease in the average posterior standard deviation is achieved when increasing the sample size from 6 to 40 patients. Similarly, the average length of the 90% HPD interval is 74% of the dose range when 6 patients are enrolled in the phase I trial and a reduction of 16% of this length is achieved when increasing the number of patients from 6 to 40. Figures 1 and 2 show the average posterior standard deviation and average lengths of the 95% HPD intervals as functions of the sample size n and target probability of DLT θ.

4.3. Illustrative Example

A randomized phase I clinical trial of the combination bortezomib and melphalan as conditioning for autologous stem cell transplant in patients with multiple myeloma was designed using EWOC and the results published in [27]. patients are randomized to arm A where a fixed dose of melphalan ($100\,\mathrm{mg/m^2}$) is given before bortezomib and arm B where the same fixed dose of melphalan is given after bortezomib. The doses available for bortezomib are 0.4, 0.7, 1.0, 1.3, and $1.6\,\mathrm{mg/m^2}$ with the first patient in either arm receiving $1.0\,\mathrm{mg/m^2}$. For each arm, the MTD is defined to be the dose level of bortezomib that when

Table 7: Average posterior standard deviation and average length of HPD of the posterior distribution of the MTD that are achieved for a given sample size for $\theta = 0.3$.

n	Mean SD	Length of 90% HPD	Length of 95% HPD
6	0.2453	0.7386	0.8161
8	0.2399	0.7238	0.8040
10	0.2351	0.7111	0.7925
12	0.2309	0.6985	0.7818
14	0.2281	0.6913	0.7755
16	0.2248	0.6821	0.7678
18	0.2221	0.6748	0.7608
20	0.2197	0.6673	0.7546
22	0.2176	0.6624	0.7500
24	0.2153	0.6557	0.7439
26	0.2136	0.6505	0.7395
28	0.2119	0.6455	0.7352
30	0.2102	0.6410	0.7313
32	0.2085	0.6350	0.7257
34	0.2072	0.6313	0.7221
36	0.2057	0.6262	0.7176
38	0.2050	0.6240	0.7162
40	0.2036	0.6200	0.7123

administered in combination with $100\,\text{mg}/\text{m}^2$ of melphalan (either before or after) to a patient results in a probability equal to $\theta = 0.33$ that a dose limiting toxicity will be manifest. In this trial, we start at $\alpha = 0.3$ and increase α in small increments of 0.05 until $\alpha = 0.5$, this value being a compromise between the therapeutic aspect of the Bortezomib and its toxic side effects. Since the doses in this trial are discrete, the dose allocated to the next patient is obtained by rounding down the dose recommended by EWOC algorithm to the nearest discrete dose, see [9, 15] on how to conduct a trial in the presence of a prespecified set of discrete doses.

Figure 3 shows all the possible dose sequences that could be realized for the first four patients, assuming that only one patient is treated at each dose and a selected situation for patient 5. The principal investigator (PI) wanted to determine the number of patients to accrue in each arm so that the posterior standard deviation of the MTD is no more than one-fifth the range of the dose level. This statistical constraint combined with the logistics such as availability of the resources for the PI, number of patients available, and limits on the duration of the trial leads us to select 20 patients per arm. In fact, a sample size of 20 results in an average posterior standard deviation $E_{D_{20}}[(\text{Var}(\gamma \mid D_{20}))^{1/2}] \approx 0.228$; This is just below one-fifth the range of dose levels 0.4–1.6.

5. Concluding Remarks

The objectives of this paper are to provide a rational for the choice of cohort sizes and number of patients to accrue in a phase I cancer clinical trial when the Bayesian adaptive design EWOC is used. In these trials, patients are typically enrolled in cohorts of size three for

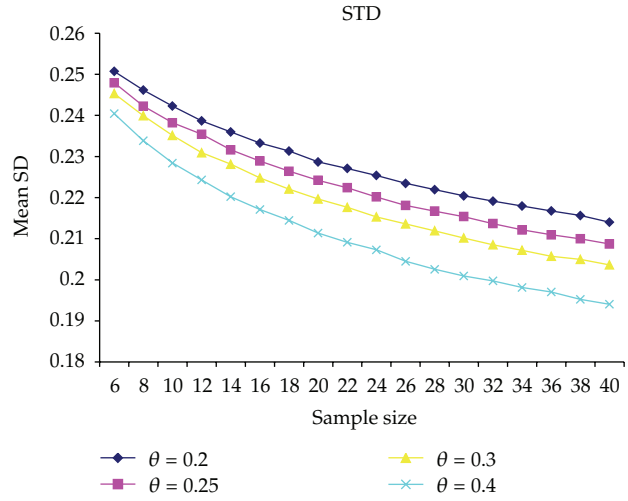


Figure 1: Estimated mean posterior standard deviation as a function of the number of patients accrued to the trial for different target probabilities of DLT θ.

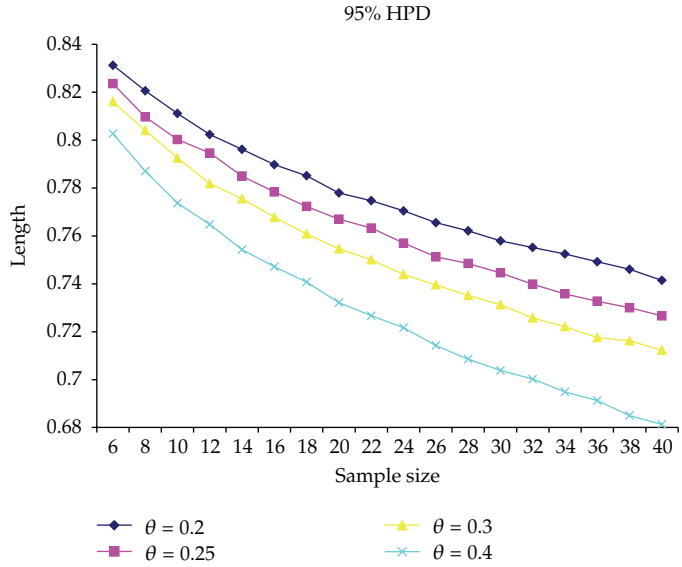


Figure 2: Estimated mean length of HPD of the posterior distribution of the MTD as a function of the number of patients accrued to the trial for different target probabilities of DLT θ.

no apparent reason other than being in agreement with the traditional "3 + 3" design and shortening the duration of the trial. We have shown through simulations that the two designs are equally safe and that no practical gain is achieved in terms of the efficiency of the estimate of the MTD. Depending on how important the length of the trial is to the clinician and the institution, we recommend using one patient per dose level to avoid seeing simultaneous toxic events when a group of patients is treated at the same dose level as was the case in

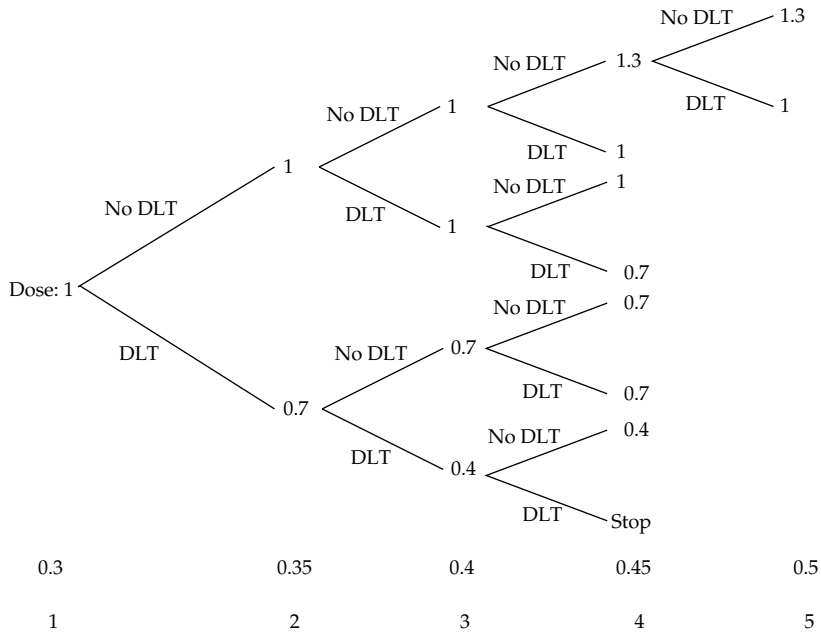


Figure 3: All the possible dose sequences that could be realized for the first four patients and a selected situation for patient 5. It assumes no simultaneous treatment of patients.

a recent phase I trial of the drug TGN1412, see [32]. In that trial, six volunteers were given what was believed to be a safe dose of an anti-inflammatory drug TGN1412. Shortly after, all 6 were admitted into intensive care due to severe reactions including swelling of the head and neck.

The simulation results were obtained by generating the toxicity responses using the logistic model (2.3). This assumption may not be true in practice and the operating characteristics of EWOC may be sensitive to model misspecification. However, for the purpose of model comparisons between ${}^{1}P$ and ${}^{3}P$ designs, any model misspecification for the probability of toxicity response will affect the two designs the same way.

In the second part of the paper, we addressed the SSD problem by giving tabulated values of the number of patients to accrue in a cancer phase I clinical trial as a function of the posterior standard deviation and length of the HPD interval of the MTD on the average over all possible trials. Although this aspect of the trial never received much emphasis in the literature due to the relatively small number of patients and logistical issues associated with such trials, we felt that providing a measure of the accuracy of the estimate of the MTD that can be achieved for a given sample size would help the clinicians understand what can and cannot be achieved during this phase of the trial. Our results show that in general, there is 17% decrease in the average posterior standard deviation of the MTD when the sample size increases from 6 to 40 patients and that for a sample size of 20 patients, the average posterior standard deviation of the MTD is about one-fifth the range of the dose levels. Although this decrease in the average posterior standard deviation seems modest, we note that this is dependent upon the use of prior distribution for the MTD. A more informative prior based on past data will result in smaller average posterior standard deviations and narrower HPD intervals.

Acknowledgments

This paper is supported in part by the National Center for Research Resources, Grant UL1RR033176, and is now at the National Center for Advancing Translational Sciences, Grant UL1TR000124 (M. Tighiouart and A. Rogatko), Grant 5P01CA098912-05 (A. Rogatko), and Grant P01 DK046763 (A. Rogatko). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References

[1] W. F. Rosenberger and L. M. Haines, "Competing designs for phase I clinical trials: a review," *Statistics in Medicine*, vol. 21, no. 18, pp. 2757–2770, 2002.
[2] N. Ting, "Dose finding in drug development," in *Dose Finding in Drug Development*, Springer, New York, NY, USA, 2006.
[3] J. O'Quigley, M. Pepe, and L. Fisher, "Continual reassessment method: a practical design for phase 1 clinical trials in cancer," *Biometrics*, vol. 46, no. 1, pp. 33–48, 1990.
[4] D. Faries, "Practical modifications of the continual reassessment method for phase I cancer clinical trials," *Journal of Biopharmaceutical Statistics*, vol. 4, no. 2, pp. 147–164, 1994.
[5] S. N. Goodman, M. L. Zahurak, and S. Piantadosi, "Some practical improvements in the continual reassessment method for phase I studies," *Statistics in Medicine*, vol. 14, no. 11, pp. 1149–1161, 1995.
[6] S. Moller, "An extension of the continual reassessment methods using a preliminary up-and-down design in a dose finding study in cancer patients, in order to investigate a greater range of doses," *Statistics in Medicine*, vol. 14, no. 9-10, pp. 911–922, 1995.
[7] S. Piantadosi, J. D. Fisher, and S. Grossman, "Practical implementation of a modified continual reassessment method for dose-finding trials," *Cancer Chemotherapy and Pharmacology*, vol. 41, no. 6, pp. 429–436, 1998.
[8] B. E. Storer, "An evaluation of phase I clinical trial designs in the continuous dose-response setting," *Statistics in Medicine*, vol. 20, no. 16, pp. 2399–2408, 2001.
[9] J. Babb, A. Rogatko, and S. Zacks, "Cancer Phase I clinical Trials: efficient dose escalation with overdose control," *Statistics in Medicine*, vol. 17, pp. 1103–1120, 1998.
[10] S. Zacks, A. Rogatko, and J. Babb, "Optimal Bayesian-feasible dose escalation for cancer phase I trials," *Statistics and Probability Letters*, vol. 38, no. 3, pp. 215–220, 1998.
[11] J. S. Babb and A. Rogatko, "Patient specific dosing in a cancer phase I clinical trial," *Statistics in Medicine*, vol. 20, no. 14, pp. 2079–2090, 2001.
[12] M. Tighiouart, A. Rogatko, and J. S. Babb, "Flexible Bayesian methods for cancer phase I clinical trials. Dose escalation with overdose control," *Statistics in Medicine*, vol. 24, no. 14, pp. 2183–2196, 2005.
[13] M. Tighiouart and A. Rogatko, "Dose finding in oncology—parametric methods," in *Dose Finding in Oncology—Parametric Methods*, N. Ting, Ed., Springer, New York, NY, USA, 2006.
[14] M. Tighiouart and A. Rogatko, "Dose escalation with overdose control," in *Dose Escalation With Overdose Control*, S. Chevret, Ed., John Wiley & Sons, New York, NY, USA, 2006.
[15] M. Tighiouart and A. Rogatko, "Dose finding with escalation with overdose control (EWOC) in cancer clinical trials," *Statistical Science*, vol. 25, no. 2, pp. 217–226, 2010.
[16] E. Garrett-Mayer, "The continual reassessment method for dose-finding studies: a tutorial," *Clinical Trials*, vol. 3, no. 1, pp. 57–71, 2006.
[17] S. Zohar, A. Latouche, M. Taconnet, and S. Chevret, "Software to compute and conduct sequential Bayesian phase I or II dose-ranging clinical trials with stopping rules," *Computer Methods and Programs in Biomedicine*, vol. 72, no. 2, pp. 117–125, 2003.
[18] Z. Xu, M. Tighiouart, and A. Rogatko, "EWOC 2.0: interactive software for dose escalation in cancer phase I clinical trials," *Drug Information Journal*, vol. 41, no. 2, pp. 221–228, 2007.
[19] A. Rogatko, M. Tighiouart, and Z. Xu, "EWOC 3.1 application software," 2009, http://biostatistics.csmc.edu/ewoc/index.php.
[20] S. Zohar and S. Chevret, "The continual reassessment method: comparison of Bayesian stopping rules for dose-ranging studies," *Statistics in Medicine*, vol. 20, no. 19, pp. 2827–2843, 2001.
[21] Y. Lin and W. J. Shih, "Statistical properties of the traditional algorithm-based designs for phase I cancer clinical trials," *Biostatistics*, vol. 2, pp. 203–215, 2001.
[22] A. Ivanova, "Escalation, group and A + B designs for dose-finding trials," *Statistics in Medicine*, vol. 25, no. 21, pp. 3668–3678, 2006.

[23] H. Borghaei, K. Alpaugh, G. Hedlund et al., "Phase I dose escalation, pharmacokinetic and pharmacodynamic study of naptumomab estafenatox alone in patients with advanced cancer and with docetaxel in patients with advanced non-small-cell lung cancer," *Journal of Clinical Oncology*, vol. 27, no. 25, pp. 4116–4123, 2009.

[24] J. D. Cheng, J. S. Babb, C. Langer et al., "Individualized patient dosing in phase I clinical trials: the role of Escalation with Overdose Control in PNU-214936," *Journal of Clinical Oncology*, vol. 22, no. 4, pp. 602–609, 2004.

[25] G. M. Freedman, N. J. Meropol, E. R. Sigurdson et al., "Phase I trial of preoperative hypofractionated intensity-modulated radiotherapy with incorporated boost and oral capecitabine in locally advanced rectal cancer," *International Journal of Radiation Oncology Biology Physics*, vol. 67, no. 5, pp. 1389–1393, 2007.

[26] N. Haas, B. Roth, C. Garay et al., "Phase I trial of weekly paclitaxel plus oral estramustine phosphate in patients with hormone-refractory prostate cancer," *Urology*, vol. 58, no. 1, pp. 59–64, 2001.

[27] S. Lonial, J. Kaufman, M. Tighiouart et al., "A phase I/II trial combining high-dose melphalan and autologous transplant with bortezomib for multiple myeloma: a dose- and schedule-finding study," *Clinical Cancer Research*, vol. 16, no. 20, pp. 5079–5086, 2010.

[28] R. J. Schilder, J. M. Gallo, M. M. Millenson et al., "Phase I trial of multiple cycles of high-dose carboplatin, paclitaxel, and topotecan with peripheral-blood stem-cell support as front-line therapy," *Journal of Clinical Oncology*, vol. 19, no. 4, pp. 1183–1194, 2001.

[29] R. Sinha, J. L. Kaufman, N. King et al., "A phase 1 dose escalation study of bortezomib combined with rituximab, cyclophosphamide, doxorubicin, modified vincristine, and prednisone for untreated follicular lymphoma and other low-grade B-cell lymphomas," *Cancer*, vol. 118, no. 14, pp. 3538–3548, 2012.

[30] C. J. Adcock, "Sample size determination: a review," *Journal of the Royal Statistical Society D*, vol. 46, no. 2, pp. 261–283, 1997.

[31] L. Joseph, D. B. Wolfson, and R. Duberger, "Sample-size calculations for binomial proportions via highest posterior density intervals," *Statistician*, vol. 44, pp. 143–154, 1995.

[32] S. Senn, D. Amin, R. A. Bailey et al., "Statistical issues in first-in-man studies," *Journal of the Royal Statistical Society A*, vol. 170, pp. 517–579, 2007.

Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2012, Article ID 436239, 15 pages
doi:10.1155/2012/436239

Research Article

Two-Stage Adaptive Optimal Design with Fixed First-Stage Sample Size

Adam Lane and Nancy Flournoy

Department of Statistics, University of Missouri, 146 Middlebush Hall, Columbia, MO 65203, USA

Correspondence should be addressed to Adam Lane, aclpp9@mail.missouri.edu

Received 1 June 2012; Accepted 5 September 2012

Academic Editor: Zhengjia Chen

In adaptive optimal procedures, the design at each stage is an estimate of the optimal design based on all previous data. Asymptotics for regular models with fixed number of stages are straightforward if one assumes the sample size of each stage goes to infinity with the overall sample size. However, it is not uncommon for a small pilot study of fixed size to be followed by a much larger experiment. We study the large sample behavior of such studies. For simplicity, we assume a nonlinear regression model with normal errors. We show that the distribution of the maximum likelihood estimates converges to a scale mixture family of normal random variables. Then, for a one parameter exponential mean function we derive the asymptotic distribution of the maximum likelihood estimate explicitly and present a simulation to compare the characteristics of this asymptotic distribution with some commonly used alternatives.

1. Introduction

Elfving [1] introduced a geometric approach for determining a c-optimal design for linear regression models. Kiefer and Wolfowitz [2] developed the celebrated equivalence theorem which provides an efficient method for verifying if a design is D-optimal, again for a linear model. These two results were generalized by Chernoff [3] and White [4] to include nonlinear models, respectively. See Bartroff [5], O'Brien and Funk [6], and references therein for extensions to the geometric and equivalence approaches. Researchers in optimal design have built an impressive body of theoretical and practical tools for linear models based on these early results. However, advances for nonlinear models have not kept pace.

One reason for the prevalence of the linear assumption in optimal design is that the problem can be explicitly described. The goal of optimal design is to determine precise experiments. Define an approximate design, proposed by Kiefer and Wolfowitz [7], as $\xi = \{x_i, w_i\}_1^K$, where ξ is a probability measure on $\mathcal{X}$ consisting of support points $x_i \in \mathcal{X}$ and

corresponding design weights w_i; w_i are rational and defined on the interval $[0,1]$ and $\sum w_i = 1$. Then the optimal design problem is to find the design that maximizes precision for a given experimental interest. Typically, this precision is achieved by maximizing some concave function, ϕ, of Fisher's information matrix. For example, when the estimation of all the parameters is the primary interest then the D-optimality criteria, where ϕ is equal to the determinant of the inverse of Fisher's information, are the most popular method. See Pukelsheim [8] for a detailed discussion of common optimality criteria.

The basic principles for nonlinear models are the same as for linear models except Fisher's information will be a function of the model parameters. As a result, optimal designs depend on the parameters and thus are only optimal in the neighborhood of the true parameters. The term locally optimal design is commonly used for nonlinear optimal designs to reflect this dependence on the parameters of interest.

To overcome this dependence Fisher [9] and Chernoff [3] suggest using expert knowledge to approximate the locally optimal design. Ford et al. [10] suggest optimal designs in nonlinear problems are to be used to provide a benchmark or to construct sequential or adaptive designs. Atkinson et al. [11] suggest using a polynomial expansion to approximate the nonlinear model with a linear one.

Stein [12] provides the earliest two-stage procedure in which the information from the first stage is used to determine design features for the second stage. In this paper we examine a two-stage adaptive optimal design procedure. An adaptive optimal design uses the data from all previous stages to estimate the locally optimal design of the current stage. Many, including Box and Hunter [13], Fedorov [14], White [15], and Silvey [16], have suggested using such designs. Recently, Lane et al. [17], Dragalin et al. [18], Fedorov et al. [19], Yao and Flournoy [20], and so forth have investigated the properties and performance of these procedures.

Lane et al. [17] show that the optimal stage-one sample size is of the order $\sqrt{n}$, where n is the overall sample size, in a two-stage regression model. Luc Pranzato obtains this relationship for a more general model (personal communication, 2012). However, in certain experiments, for example, early phase clinical trials or bioassay studies, it is common to use designs with very small stage-one sample sizes. Current literature has characterized the adaptive optimal design procedure under the assumption that both stage-one and stage-two sample sizes are large.

In this paper we characterize the asymptotic distribution of the maximum likelihood estimate (MLE) when the stage-one sample size is fixed. The distribution for a nonlinear regression model with normal errors and a one parameter exponential mean function is derived explicitly. Then for a specific numeric example the differences between the finite stage-one sample distribution are compared with other candidate approximate distributions.

2. Adaptive Optimal Procedure for a Two-Stage Nonlinear Regression Model with Normal Errors

2.1. The Model

Let $\{y_{ij}\}_{1,1}^{n_i,2}$ be observations from a two-stage experiment, where n_i is the number of observations and x_i is the single-dose level used for the ith stage, $i = 1, 2$. Assume that

$$y_{ij} = \eta(x_i, \theta) + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \mathcal{N}\left(0, \sigma^2\right), \tag{2.1}$$

where $\eta(x,\theta)$ is some nonlinear mean function. In most practical examples it is necessary to consider a bounded design space, that is, $x_i \in \mathcal{X} = [a,b]$, $-\infty < a < b < \infty$. It is assumed that y_{ij} are independent conditional on treatment x_i, where x_1 is fixed and x_2 is selected adaptively. Denote the adaptive design by $\xi_A = \{x_i, w_i\}_1^2$, where $w_i = n_i/n$.

The likelihood for model (2.1) is

$$\begin{aligned}&\mathcal{L}_n(\theta \mid \overline{y}_1, \overline{y}_2)\\ &\quad\propto \exp\left\{-\frac{n_1}{2\sigma^2}\left(\overline{y}_1 - \eta(x_1,\theta)\right)^2 - \frac{n_2}{2\sigma^2}\left(\overline{y}_2 - \eta(x_2,\theta)\right)^2\right\},\end{aligned} \tag{2.2}$$

where $\overline{y}_i = n_i^{-1}\sum_1^{n_i} y_{ij}$ are the stage specific sample means, and the total score function is

$$\begin{aligned}S &= \frac{d}{d\theta}\ln \mathcal{L}_n(\theta \mid \overline{y}_1, \overline{y}_2) = \frac{n_1}{2\sigma^2}\left(\overline{y}_1 - \eta(x_1,\theta)\right)\frac{d\eta(x_1,\theta)}{d\theta}\\ &\quad + \frac{n_2}{2\sigma^2}\left(\overline{y}_2 - \eta(x_2,\theta)\right)\frac{d\eta(x_2,\theta)}{d\theta} = S_1 + S_2,\end{aligned} \tag{2.3}$$

where S_i represents the score function for the ith stage.

2.2. The Adaptive Optimal Procedure

Fix the first stage design point x_1 and let $\widetilde{\theta}_{n_1}$ represent an estimate based on the first-stage complete sufficient statistic $\overline{y}_1$. The locally optimal design point for the second stage is

$$x^* = \arg\max_{x\in\mathcal{X}} \operatorname*{Var}_{\overline{y}_2|x}(S_2) = \arg\max_{x\in\mathcal{X}}\left(\frac{d\eta(x,\theta)}{d\theta}\right)^2, \tag{2.4}$$

which is commonly estimated by $x^*|_{\theta=\widetilde{\theta}_{n_1}}$ for use in stage 2. Because the adaptive optimal design literature assumes n_1 is large, the MLE of the second stage design point, $x^*|_{\theta=\widehat{\theta}_{n_1}}$, where $\widehat{\theta}_{n_1}$ is the MLE of θ based on the first stage data, is traditionally used to estimate x^*.

However, when n_1 is small the bias of the MLE can be considerable. Therefore, for some mean functions η using a different estimate would be beneficial. In general, the adaptively selected stage two treatment is

$$x_2 = \arg\max_{x\in\mathcal{X}}\left(\frac{d\eta(x,\theta)}{d\theta}\right)^2\Bigg|_{\theta=\widetilde{\theta}_{n_1}}. \tag{2.5}$$

2.3. Fisher's Information

Since $x_1 \in \mathcal{X} = [a,b]$, a bounded design space, but $y \in \mathbb{R}$, there is a positive probability that x_2 will equal a or b. Denote these probabilities as $\pi_a = P(x_2 = a)$ and $\pi_b = P(x_2 = b)$, respectively. Then the per subject information can be written as

$$M(\xi_A,\theta) = \frac{1}{n}\operatorname{Var}(S) = \frac{1}{\sigma^2}\Bigg[w_1\left(\frac{d\eta(x_1,\theta)}{d\theta}\right)^2 + w_2\left(\pi_a\left(\frac{d\eta(a,\theta)}{d\theta}\right)^2 + \pi_b\left(\frac{d\eta(b,\theta)}{d\theta}\right)^2 + E_{x_2}\left[\left(\frac{d\eta(x_2,\theta)}{d\theta}\right)^2 I(a<x_2<b)\right]\right)\Bigg], \tag{2.6}$$

where x_2 is the random variable defined by the onto transformation (2.5) of $\overline{y}_1$.

3. Asymptotic Properties

We examine three different ways of deriving an asymptotic distribution of the final MLE which may be used for inference at the end of the study. The first is under the assumption that both n_1 and n_2 are large. The second considers the data from the second stage alone. Finally, assume a fixed first-stage sample size and a large second-stage sample size.

3.1. Large Stage-1 and Stage-2 Sample Sizes

If $d\eta(x_2,\theta)/d\theta$ is bounded and continuous and provided common regularity conditions that hold,

$$\sqrt{n}\left(\widehat{\theta}_n - \theta_t\right) \xrightarrow{\mathfrak{D}} \mathcal{N}\left(0, M^{-1}(\xi^*,\theta)\right), \tag{3.1}$$

as $n_1 \to \infty$ and $n_2 \to \infty$, where $\xi^* = \{(x_1,n_1),(x^*,n_2)\}$. This result is used to justify the common practice of using $x^*|_{\theta=\widehat{\theta}_{n_1}}$ to estimate x^* in order to make inferences about θ. However, if $d\eta(x_2,\theta)/d\theta$ is not bounded and continuous then it is very difficult to obtain the result in (3.1) and for certain mean functions the result will not hold. In such cases the asymptotic variance in (3.1) must be replaced with $\lim_{n_1\to\infty} M^{-1}(\xi_A,\theta)$. Lane et al. [17] examine using the exact Fisher's information for an adaptive design ξ_A, $M(\xi_A,\theta)$, instead of $M(\xi^*,\theta)$ in (3.1) to obtain an alternative approximation of the variance of the MLE $\widehat{\theta}_n$.

3.2. Distribution of the MLE If Only Second-Stage Data Are Considered

Often pilot data are discarded after being used to design a second experiment then the derivation of the distribution of the MLE using only the second-stage data takes if x_2 to be fixed:

$$\sqrt{n_2}\left(\widehat{\theta}_{n_2} - \theta_t\right) \xrightarrow{\mathfrak{D}} \mathcal{N}\left(0, M_2^{-1}(x_2,\theta)\right), \tag{3.2}$$

as $n_2 \to \infty$, where $M_2(x_2,\theta) = \sigma^{-2}(d\eta(x_2,\theta)/d\theta)^2$. The estimate $\widehat{\theta}_{n_2}$ will likely perform poorly in comparison to $\widehat{\theta}_n$ if n_1 and n_2 are relatively of the same size but conceivably may perform quite well when n_1 is much smaller than n. For this reason it represents an informative benchmark distribution.

3.3. Fixed First-Stage Sample Size; Large Second-Stage Sample Size

When the first-stage sample size is fixed and the second stage is large we have the following result.

Theorem 3.1. *For model* (2.1) *with* x_2 *as defined in* (2.5) *if* $d\eta/d\theta \neq 0$ *for all* $x \in \mathcal{X}$, $\theta \in \Theta$, x_2 *is an onto function of* $\overline{y}_1$, $|d\eta/d\theta| < \infty$ *and provided common regularity conditions,*

$$\sqrt{n}\left(\widehat{\theta}_n - \theta_t\right) \xrightarrow{\mathfrak{D}} UQ \tag{3.3}$$

as $n_2 \to \infty$, *where* $Q \sim \mathcal{N}(0,\sigma^2)$ *and* $U = ((d\eta(x_2,\theta))/d\theta)^{-1}$ *is a random function of* $\overline{y}_1$.

Proof. As in classical large sample theory (cf. Ferguson [21] and Lehmann [22]):

$$\sqrt{n}\left(\widehat{\theta}_n - \theta\right) \approx \frac{(1/\sqrt{n})S}{-(1/n)(d/d\theta)S}, \tag{3.4}$$

since $S(\widehat{\theta}_n)$ can be expanded around $S(\theta_t)$ as

$$S\left(\widehat{\theta}_n\right) = S(\theta_t) + \left(\widehat{\theta}_n - \theta_t\right)\frac{d}{d\theta}S(\theta_t) + \frac{1}{2}\left(\widehat{\theta}_n - \theta_t\right)^2\frac{d^2}{d\theta^2}S(\theta^*), \tag{3.5}$$

where θ_t is the true value of the parameter and $\theta^* \in (\theta_t, \widehat{\theta}_n)$. Solving for $\sqrt{n}(\widehat{\theta}_n - \theta_t)$ gives

$$\sqrt{n}\left(\widehat{\theta}_n - \theta_t\right) = \frac{(1/\sqrt{n})S(\theta_t)}{-(1/n)\left((d/d\theta)S(\theta_t) + (1/2)\left(\widehat{\theta}_n - \theta_t\right)(d^2/d\theta^2)S(\theta^*)\right)}. \tag{3.6}$$

It can be shown that $\widehat{\theta}_n$ is consistent for θ_t if $n_2 \to \infty$ and $n_1/n \to 0$ which gives the result in (3.4).

Now, decompose the right hand side of (3.4) as

$$\begin{aligned}\frac{(1/\sqrt{n})S}{-(1/n)(d/d\theta)S} &= \frac{(1/\sqrt{n})(S_1+S_2)}{-(1/n)((d/d\theta)S_1 + (d/d\theta)S_2)} \\ &= \frac{(1/\sqrt{n})S_1}{-(1/n)((d/d\theta)S_1 + (d/d\theta)S_2)} + \frac{(1/\sqrt{n})S_2}{-(1/n)((d/d\theta)S_1 + (d/d\theta)S_2)}.\end{aligned} \tag{3.7}$$

As $n_2 \to \infty$, $S_1/\sqrt{n} \to 0$, $(n_2/n) \to 1$, and $(1/n)(d/d\theta)S_2 \nrightarrow 0$ as $n \to \infty$. Thus, the first term in (3.7) goes to 0 as $n \to \infty$. Write the second term in (3.7) as

$$\frac{(1/\sqrt{n})S_2}{-(1/n)((d/d\theta)S_1 + (d/d\theta)S_2)} = \left(-\frac{(1/n)(d/d\theta)S_1}{(1/\sqrt{n})S_2} - \frac{(1/n)(d/d\theta)S_2}{(1/\sqrt{n})S_2}\right)^{-1}. \tag{3.8}$$

Further as $n_2 \to \infty$, $(1/n)(d/d\theta)S_1 \to 0$ and $(1/\sqrt{n})S_2 \nrightarrow 0$,

$$\begin{aligned} \frac{(1/n)(d/d\theta)S_1}{(1/\sqrt{n})S_2} &\xrightarrow{p} 0, \\ \frac{(1/n)(d/d\theta)S_2}{(1/\sqrt{n})S_2} &= \frac{(1/n)(\overline{y}_2 - \eta(x_2,\theta))(d^2\eta(x_2,\theta)/d\theta^2) + w_2((d\eta(x_2,\theta))/d\theta)^2}{(1/\sqrt{n})n_2(\overline{y}_2 - \eta(x_2,\theta))((d\eta(x_2,\theta))/d\theta)} \\ &= \left(\frac{1}{\sqrt{n}}\right)\frac{d^2\eta(x_2,\theta)/d\theta^2}{d\eta(x_2,\theta)/d\theta} + \frac{\sqrt{w_2}(d\eta(x_2,\theta))/d\theta}{\sqrt{n_2}(\overline{y}_2 - \eta(x_2,\theta))}. \end{aligned} \tag{3.9}$$

The first term in (3.9) goes to 0. To evaluate the second term, it is important to recognize that $\varepsilon_{i2} = \overline{y}_2 - \eta(x_2,\theta) \sim \mathcal{N}(0,\sigma^2/n_2)$ and $\overline{y}_1 \sim \mathcal{N}(0,\sigma^2/n_1)$ are independent and thus

$$\overline{y}_2 - \eta(x_2,\theta), \qquad \frac{d\eta(x_2,\theta)}{d\theta} \tag{3.10}$$

are independent. Because of this independence,

$$\sqrt{n_2}(\overline{y}_2 - \eta(x_2,\theta))\left(\frac{d\eta(x_2,\theta)}{d\theta}\right)^{-1} \sim UQ, \tag{3.11}$$

where U is a random function of $\overline{y}_1$ and $Q \sim \mathcal{N}(0,\sigma^2)$ as determined by $((d\eta(x_2,\theta))/d\theta)^{-1}$. Now, with $\sqrt{w_2} \to 1$ as $n_2 \to \infty$ the result follows from an application of Slutsky's theorem. □

Remark 3.2. Provided $d\eta(x,\theta)/d\theta$ is bounded and continuous UQ is the asymptotic distribution of $\sqrt{n}(\widehat{\theta}_n - \theta_t)$ as $n \to \infty$. The important case for this exposition is presented in Theorem 3.1. However, the two other potential cases can be shown easily.

Case 1. $n_1 \to \infty$, $n_2 \to \infty$, and $n \to \infty$. As $n_1 \to \infty$, $x_2 \to x^*$ which implies that $U \to ((d(x^*,\theta))/d\theta)^{-1}$, a constant, and thus UQ converges to asymptotic distribution of $\sqrt{n}(\widehat{\theta}_n - \theta_t)$ given in (3.1).

Case 2. $n_1 \to \infty, n_2$ fixed, and $n \to \infty$. Just as in Case 1, $U \to M^{-1}(x^*,\theta)$, where $M(x^*,\theta) = ((d\eta(x^*,\theta))/d\theta)^{-2}$. Note that $M(x^*,\theta)$ differs from $M(\xi^*,\theta)$ which depends on x_1 and x^*.

Therefore $UQ \rightarrow \mathcal{N}(0, \sigma^2 M^{-1}(x^*, \theta))$. Look back at (3.7) in the proof, but now take n_2 to be fixed; $(1/\sqrt{n})S_2 \rightarrow 0$ and $(1/\sqrt{n})(d/d\theta)S_2 \rightarrow 0$ and the only term left is

$$\frac{(1/\sqrt{n})S_1}{-(1/n)(d/d\theta)S_1}. \tag{3.12}$$

Consider the following: $(1/\sqrt{n})S_1 \rightarrow \mathcal{N}(0, \sigma^2 M^{-1}(x^*, \theta))$ and $(1/\sqrt{n})(d/d\theta)S_1 \rightarrow M^{-1}(x^*, \theta)$ as $n \rightarrow \infty$. Therefore, $\sqrt{n}(\widehat{\theta}_n - \theta_t) \rightarrow \mathcal{N}(0, \sigma^2 M^{-1}(x^*, \theta))$ as $n \rightarrow \infty$ which is equivalent to UQ.

4. Example: One Parameter Exponential Mean Function

In model (2.1) let $\eta(x, \theta) = e^{-\theta x}$, where $x \in \mathcal{X} = [a, b]$, $0 < a < b < \infty$ and $\theta \in (0, \infty)$. The simplicity of the exponential mean model facilitates our illustration, but it is also important in its own right. For example, Fisher [9] used a variant of this model to examine the information in serial dilutions. Cochran [23] further elaborated on Fisher's application using the same model.

For this illustration we use the MLE of the first-stage data to estimate the second-stage design point. Here,

$$\widehat{\theta}_{n_1} = \begin{cases} \dfrac{-\ln \overline{y}_1}{x}, & \text{if } \overline{y}_1 \in \left(e^{-\overline{\theta}x}, 1\right), \\ 0, & \text{if } \overline{y}_1 \geq 1, \\ \overline{\theta}, & \text{if } \overline{y}_1 \leq e^{-\overline{\theta}x}. \end{cases} \tag{4.1}$$

The adaptively selected second-stage treatment as given by (2.5) is

$$x_2 = \arg\max_{x \in \mathcal{X}} \left(x^2 e^{-2\theta x}\right) = \begin{cases} \widehat{\theta}_{n_1}^{-1}, & \text{if } \overline{y}_1 \in \left(e^{-a^{-1}x_1}, e^{-b^{-1}x_1}\right), \\ b, & \text{if } \overline{y}_1 \geq e^{-b^{-1}x_1}, \\ a, & \text{if } \overline{y}_1 \leq e^{-a^{-1}x_1}. \end{cases} \tag{4.2}$$

Thus, the exact per subject Fisher information is

$$\begin{aligned} M(\xi_A, \theta) = \frac{1}{n}\operatorname{Var}(S) &= w_1 x_1^2 e^{-2\theta x_1} + w_2 \pi_a a^2 e^{-2\theta a} \\ &\quad + w_2 \pi_b b^2 e^{-2\theta b} + w_2 E_{x_2}\left[x_2^2 e^{-2\theta x_2} \cdot I(a < x_2 < b)\right]. \end{aligned} \tag{4.3}$$

For this example $M(\xi_A, \theta) \rightarrow M(\xi^*, \theta)$ as $n_1 \rightarrow \infty$. For more detailed information on the derivations of (4.1), (4.2), and (4.3) see Lane et al. [17].

The asymptotic distributions of the MLE in Sections 3.1 and 3.2 can be derived easily. For the asymptotic distribution of the MLE in Section 3.3 consider the following corollary. For details on the functions h, v_1, and v_2 see the proof of the corollary.

Corollary 4.1. *If $\eta(x,\theta) = e^{-\theta x}$ in model* (2.1) *then*

$$\sqrt{n}\left(\widehat{\theta}_n - \theta\right) \xrightarrow{\mathfrak{D}} UQ \tag{4.4}$$

as $n \to \infty$, where UQ is defined by

$$P(UQ \le t) = \begin{cases} P\left(U \ge \dfrac{t}{q} \mid -\infty < q \le 0\right)\Phi(q), & \text{if } t \in (-\infty, 0), \\ P\left(U \le \dfrac{t}{1} \mid 0 < q \le \infty\right)(1 - \Phi(q)), & \text{if } t \in (0, \infty), \end{cases} \tag{4.5}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. Let $\Psi(q) = \Phi(\sqrt{n}(q - \eta(x,\theta))/\sigma)$ and $h(s) = s^{-1}e^{\theta s}$. Then if $h(a) < h(b)$,

$$\begin{aligned} P\left(U \ge \frac{t}{q} \mid -\infty < q \le 0\right)\Phi(q) &= \Phi\left(\frac{t}{\sigma h(1/\theta)}\right) \\ &\quad + [1 - (\Psi(v_2(h(a))) - \Psi(v_1(h(a))))] \\ &\quad \times \left[\Phi\left(\frac{t}{\sigma h(a)}\right) - \Phi\left(\frac{t}{\sigma h(1/\theta)}\right)\right] \\ &\quad + [\Psi(v_2(h(b))) - \Psi(v_2(h(a)))] \\ &\quad \times \left[\Phi\left(\frac{t}{\sigma h(b)}\right) - \Phi\left(\frac{t}{\sigma h(a)}\right)\right], \\ P\left(U \le \frac{t}{q} \mid 0 < q \le \infty\right)(1 - \Phi(q)) &= \Phi\left(\frac{t}{\sigma h(b)}\right) \\ &\quad + [\Psi(v_2(h(a))) - \Psi(v_1(h(a)))] \\ &\quad \times \left[\Phi\left(\frac{t}{\sigma h(1/\theta)}\right) - \Phi\left(\frac{t}{\sigma h(a)}\right)\right] \\ &\quad + [1 - (\Psi(v_2(h(b))) - \Psi(v_2(h(a))))] \\ &\quad \times \left[\Phi\left(\frac{t}{\sigma h(a)}\right) - \Phi\left(\frac{t}{\sigma h(b)}\right)\right]. \end{aligned} \tag{4.6}$$

If $h(b) < h(a)$, then

$$\begin{aligned} P\left(U \ge \frac{t}{q} \mid -\infty < q \le 0\right)\Phi(q) &= \Phi\left(\frac{t}{\sigma h(1/\theta)}\right) \\ &\quad + [1 - (\Psi(v_2(h(b))) - \Psi(v_1(h(b))))] \\ &\quad \times \left[\Phi\left(\frac{t}{\sigma h(b)}\right) - \Phi\left(\frac{t}{\sigma h(1/\theta)}\right)\right] \\ &\quad + [\Psi(v_1(h(b))) - \Psi(v_1(h(a)))] \\ &\quad \times \left[\Phi\left(\frac{t}{\sigma h(b)}\right) - \Phi\left(\frac{t}{\sigma h(a)}\right)\right], \end{aligned}$$

$$P\left(U \le \frac{t}{q} \mid 0 < q \le \infty\right)(1 - \Phi(q)) = \Phi\left(\frac{t}{\sigma h(a)}\right) + [\Psi(v_2(h(b))) - \Psi(v_1(h(b)))] \times \left[\Phi\left(\frac{t}{\sigma h(1/\theta)}\right) - \Phi\left(\frac{t}{\sigma h(b)}\right)\right] + [1 - (\Psi(v_1(h(b))) - \Psi(v_1(h(a))))] \times \left[\Phi\left(\frac{t}{\sigma h(b)}\right) - \Phi\left(\frac{t}{\sigma h(a)}\right)\right]. \tag{4.7}$$

Proof. First, we find the distribution of U where $U = h(z)$ and the random variable z is defined by

$$z = \begin{cases} -\dfrac{x_1}{\ln \overline{y}_1}, & \text{if } \overline{y}_1 \in \left(e^{-x_1/a}, e^{-x_1/b}\right), \\ -\dfrac{x_1}{\ln a}, & \text{if } \overline{y}_1 \le e^{-x_1/a}, \\ -\dfrac{x_1}{\ln b}, & \text{if } \overline{y}_1 \ge e^{-x_1/b}. \end{cases} \tag{4.8}$$

Figure 1 illustrates the map from U to $z \in [a, b]$ where $\theta = 1$, $\sigma = .5$, $a = .25$, and $b = 4$.

Lambert's product log function (cf. Corless et al. [24]) is defined as the solutions to

$$we^w = c \tag{4.9}$$

for some constant c. Denote the solutions to (4.9) by $W(w)$. Let

$$V(c) = \arg_{\overline{y}_1}\left\{\left(-\frac{x_1}{\ln \overline{y}_1}\right)^{-1} \exp\left\{\theta \frac{-x_1}{\ln \overline{y}_1}\right\} = c\right\}. \tag{4.10}$$

Then

$$V(c) = \exp\left\{\frac{\theta x_1}{W(-\theta/c)}\right\}. \tag{4.11}$$

The W function is real valued on $w \ge -1/e$, single valued at $w = -1/e$, and double valued on $w \in (-1/e, 0)$. $U \in \{\theta e, \max\{h(a), h(b)\}\}$, $x_1 \in [a, b]$, $0 < a < b < \infty$. Therefore $V(c)$ is real valued for all $\theta \in (0, \infty)$. For simplicity, define $v_1 = \min V(c)$ and $v_2 = \max V(c)$ for a given c.

We present the proof for the cumulative distribution function (CDF) of U and the CDF of UQ for the case where $x^* \in [a, b]$ and $h(a) < h(b)$. The derivation of the distributions under alternative cases is tedious and does not differ greatly from this case.

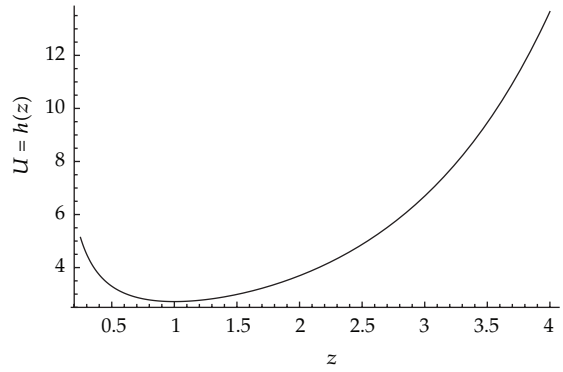


Figure 1: Map of $z = -x_1/\ln \overline{y}_1$ for $\theta = 1$, $a = .25$, and $b = 4$.

Note in this case the domain of U is $[h(1/\theta) = \theta e, h(b)]$. If $h(1/\theta) < U < h(a)$, then

$$P(U \le t_1) = P\left(h\left(\overline{y}_1\right) < t\right) = P\left(v_1(t_1) < \overline{y}_1 < v_2(t_1)\right) = \Psi(v_2(t_1)) - \Psi(v_1(t_1)). \tag{4.12}$$

If $U = h(a)$, then

$$P(U \le h(a)) = \Psi(v_2(h(a))). \tag{4.13}$$

If $U \in (h(a), h(b))$, then

$$P(U \le t_1) = P\left(v_1(t_1) < \overline{y}_1 < v_2(t_1)\right). \tag{4.14}$$

However, since $t_1 < h(a)$ $P(\overline{y}_1 < v_1(t_1)) = 0$,

$$P(U \le t_1) = \Psi(v_2(t_1)). \tag{4.15}$$

If $U \ge h(b)$, then

$$P(U \le h(b)) = 1. \tag{4.16}$$

Thus,

$$P(U \le t_1) = \begin{cases} 0, & \text{if } t_1 \le h\left(\dfrac{1}{\theta}\right), \\ \Psi(v_2(t_1)) - \Psi(v_2(t_1)), & \text{if } t_1 \in \left(h\left(\dfrac{1}{\theta}\right), h(a)\right), \\ \Psi\left(e^{-x_1/a}\right), & \text{if } t_1 = h(a), \\ \Psi(v_2(t_1)), & \text{if } t_1 \in (h(a), h(b)), \\ 1, & \text{if } h(b) \le t_1 \le \infty. \end{cases} \tag{4.17}$$

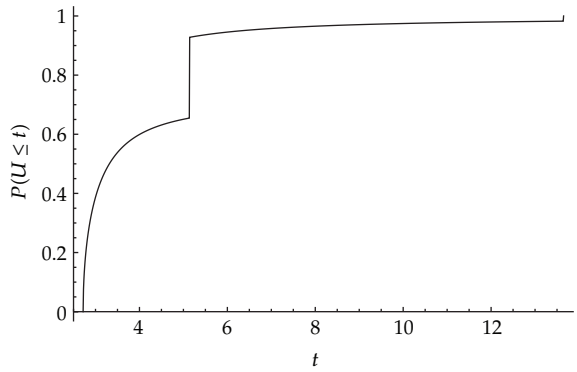


Figure 2: CDF of U for $\theta = 1$, $x_1 = 2$, $n_1 = 5$, $\sigma = .5$, $a = .25$, and $b = 4$.

Figure 2 plots the CDF of U for $\theta = 1$, $x_1 = 2$, $n_1 = 5$, $\sigma = .5$, $a = .25$, and $b = 4$. The distribution is a piecewise function with discontinuities at the boundary points a and b.

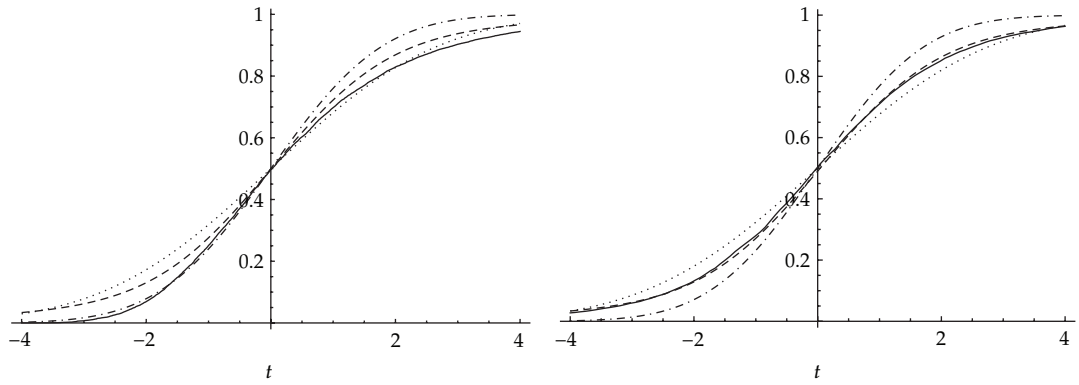
Now consider the distribution of UQ. Recall $q \sim \mathcal{N}(0, \sigma^2)$ and U and Q are independent. If $t \in (-\infty, 0)$, then

$$\begin{aligned} P(UQ \le t) = {} & P\left(U \ge \frac{t}{q} \mid 0 < \frac{t}{q} \le h\left(\frac{1}{\theta}\right)\right) P\left(0 < \frac{t}{q} \le h\left(\frac{1}{\theta}\right)\right) \\ & + P\left(U \ge \frac{t}{q} \mid h\left(\frac{1}{\theta}\right) < \frac{t}{q} \le h(a)\right) P\left(h\left(\frac{1}{\theta}\right) < \frac{t}{q} \le h(a)\right) \\ & + P\left(U = h(a) \mid \frac{t}{q} = h(a)\right) P\left(\frac{t}{q} = h(a)\right) \\ & + P\left(U \ge \frac{t}{q} \mid h(a) < \frac{t}{q} \le h(b)\right) P\left(h(a) < \frac{t}{q} \le h(b)\right) \\ & + P\left(U = h(b) \mid \frac{t}{q} = h(b)\right) P\left(\frac{t}{q} = h(b)\right) \\ & + P\left(U \ge \frac{t}{q} \mid h(b) < \frac{t}{q} \le \infty\right) P\left(h(b) < \frac{t}{q} < \infty\right). \end{aligned} \tag{4.18}$$

The distribution is symmetric, thus the derivation of the CDF if $t \in (0, \infty)$ is analogous. □

4.1. Comparisons of Asymptotic Distributions

First, consider the distribution described in (3.1) using $M(\xi_A, \theta)$ in place of $M(\xi^*, \theta)$ and the distribution described in (3.2). When n_1 is significantly smaller than n_2, $M(\xi_A, \theta)$ and $M(x_2, \theta)$ can differ significantly as a function of $\overline{y}_1$. This is primarily because $M(x_2, \theta)$ is a function of x_2, whereas $M(\xi_A, \theta)$ is an average over $\overline{y}_1$. Through simulation it can be seen that a $\mathcal{N}(0, M^{-1}(x_2, \theta))$ is a better approximate distribution of $\sqrt{n}(\widehat{\theta}_n - \theta)$ than $\mathcal{N}(0, M^{-1}(\xi_A, \theta))$ for only a small interval of x_2, and this interval has a very small probability. For these reasons the distribution of the MLE using only the second stage data as described in Section 3.2 is not considered further.



(a) $n = 30$ (b) $n = 1000$

Figure 3: In each plot the solid line represents the CDF of $\sqrt{n}(\widehat{\theta}_n - \theta)$ obtained via Monte Carlo simulations. The dotted-dashed line is the $P(T_1 \leq t)$, where $T_1 \sim \mathcal{N}(0, M^{-1}(\xi^*, \theta))$. The dotted line is the $P(T_2 \leq t)$, where $T_2 \sim \mathcal{N}(0, M^{-1}(\xi_A, \theta))$. The dashed line is the $P(T_3 \leq t)$, where $T_3 \sim Q$. Values $\theta = 1$, $x_1 = 2$, $n_1 = 5$, $\sigma = .5$, $a = .25$, and $b = 4$ were used.

Now for a set of numeric examples consider three distributions: (3.1), (3.1) using $M(\xi_A, \theta)$ in place of $M(\xi^*, \theta)$ and the distribution of UQ defined in (3.3). An asymptotic distribution can be justified in inference if it is approximately equal to the true distribution. In this case the true distribution is that of $\sqrt{n}(\widehat{\theta}_n - \theta)$. However, $\widehat{\theta}_n$ does not have a closed form and thus its distribution cannot be obtained analytically or numerically. To approximate this distribution 10,000 Monte Carlo simulations have been completed for each example to create a benchmark distribution.

Figure 3 plots the three different candidate approximate distributions, found exactly using numerical methods, together with the distribution of $\sqrt{n}(\widehat{\theta}_n - \theta)$ approximated using Monte Carlo simulations, for $\theta = 1$, x_1, $\sigma = .5$, $a = .25$, $b = 4$, $n_1 = 5$, and $n = \{30, 1000\}$. Note the y-axis represents $P(T_i \leq t)$, $i = 1, 2, 3$, where T_1 is $\mathcal{N}(0, M^{-1}(\xi^*, \theta))$, T_2 is $\mathcal{N}(0, M^{-1}(\xi_A, \theta))$, and T_3 is UQ. When $n = 30$ it is difficult, graphically, to determine if T_2 or T_3 provides a better approximation for $\sqrt{n}(\widehat{\theta}_n - \theta)$. It seems that if $t \in (-4, 0)$ the distribution T_3 is preferable to T_2; however, when $t \in (0, 4)$ the opposite appears to be the case. It is fairly clear that for this example T_1 performs poorly.

When $n = 1000$, it is clear that T_3 is much closer to $\sqrt{n}(\widehat{\theta}_n - \theta)$ than both T_1 and T_2. Further, comparing the two plots one can see how the distribution of $\sqrt{n}(\widehat{\theta}_n - \theta)$ has nearly converged to UQ but still differs from those T_1 and T_2 significantly, as predicted by Theorem 3.1 and Corollary 4.1.

Using only graphics it is difficult to assess which of T_1, T_2, and T_3 is nearest $\sqrt{n}(\widehat{\theta}_n - \theta)$ for a variety of cases. To get a better understanding, the integrated absolute difference of the CDFs of T_1, T_2, and T_3 versus that of $\sqrt{n}(\widehat{\theta}_n - \theta)$ for $x_1 = 2$, $\sigma = .5$, $a = .25$, $b = 4$, $n = \{5, 10, 15\}$, and $n = \{30, 50, 100, 400\}$ is presented in Table 1. First, consider the table where $\theta = .5$. The locally optimal stage-1 design point is $x_1 = 2$ when $\theta = .5$; as a result this scenario is the most generous to distribution T_1. However, even for this ideal scenario T_3 outperforms T_1 and T_2 for all values of n_1. In many cases the difference between T_3 and T_1 is quite severe. In this scenario T_3 outperforms T_2; however, the differences are not great.

Table 1: Integrated absolute difference of the cumulative distributions (×100) of $T_1 \sim \mathcal{N}(0, M^{-1}(\xi^*, \theta))$, $T_2 \sim \mathcal{N}(0, M^{-1}(\xi_A, \theta))$, and $T_3 \sim UQ$ versus the approximate cumulative distribution of $\sqrt{n}(\widehat{\theta}_n - \theta)$ obtained via Monte Carlo simulations for various n_1 and various moderate sizes of n. The values $\theta = 1$, $x_1 = 2$, $\sigma = .5$, $a = .25$, and $b = 4$ were used.

(a) ($\theta = .5$)

n_1	$n = 30$			$n = 50$			$n = 100$			$n = 400$		
	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3
5	19	24	11	16	17	8	15	16	7	14	15	5
10	11	13	8	9	12	7	9	11	6	7	12	4
15	9	10	8	8	9	6	6	9	5	4	10	3

(b) ($\theta = 1$)

n_1	$n = 30$			$n = 50$			$n = 100$			$n = 400$		
	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3
5	30	33	25	30	27	19	34	24	12	39	21	6
10	40	40	32	26	27	22	23	28	16	26	20	8
15	34	34	33	27	28	24	21	23	17	18	20	9

(c) ($\theta = 1.5$)

n_1	$n = 30$			$n = 50$			$n = 100$			$n = 400$		
	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3
5	32	33	31	39	25	25	42	21	23	42	17	21
10	34	33	22	27	25	16	32	22	10	35	19	12
15	35	35	32	26	26	21	26	22	13	28	21	7

Next, examine the results for $\theta = 1$ and $\theta = 1.5$. Once again T_3 outperforms T_1 and T_2 in all but 2 cases, where in many cases its advantage is quite significant. Also note that T_2 outperforms T_1 about half the time when $\theta = 1$ and the majority of the time when $\theta = 1.5$. This supports our observation that when the distance between x_1 and x^* increases the performance of T_1 compared with T_2 and T_3 worsens which indicates a lack of robustness for the commonly used distribution T_1. This lack of robustness is not evident for T_1 and T_2.

One final comparison is motivated by the fact that if $n_1 \to \infty$, T_1, T_2, and T_3 have the same asymptotic distribution. Although our method is motivated by the scenario where n_1 is a small pilot study, there is no theoretical reason that T_3 will not perform competitively when n_1 is large. Table 2 presents the integrated differences for the distributions T_2 and T_3 from $\sqrt{n}(\widehat{\theta}_n - \theta)$ for $x_1 = 2$, $\theta = 1$, $\sigma = .5$, $a = .25$, $b = 4$, $n_1 = \{50, 100, 200\}$, and $n = \{400, 1000\}$. T_1 is not included in the table due to the lack of robustness; it can perform better or worse than the other two distributions based on the value of θ. Even with larger values of n_1, T_3 performs slightly better when $n_1 = 50$ and 100 and only slightly worse when $n = 200$ indicating that using T_3 is robust for moderately large n_1.

5. Discussion

Assuming a finite first-stage sample size and a large second-stage sample size, we have shown for a general nonlinear one parameter regression model with normal errors that the asymptotic distribution of the MLE is a scale mixture distribution. We considered only one parameter for simplicity and clarity of exposition.

Table 2: Integrated absolute difference of the cumulative distributions (×100) of $T_1 \sim \mathcal{N}(0, M^{-1}(\xi^*, \theta))$, $T_2 \sim \mathcal{N}(0, M^{-1}(\xi_A, \theta))$, and $T_3 \sim UQ$ versus the approximate cumulative distribution of $\sqrt{n}(\widehat{\theta}_n - \theta)$ obtained via Monte Carlo simulations for various n_1 and various large sizes of n. The values $\theta = 1$, $x_1 = 2$, $\sigma = .5$, $a = .25$, and $b = 4$ were used.

n_1	$n = 400$		$n = 1000$	
	T_2	T_3	T_2	T_3
50	13	9	13	5
100	10	9	8	4
200	11	14	4	7

For the one parameter exponential mean function, the distribution of the adaptively selected second-stage treatment and the asymptotic distribution of the MLE were derived assuming a finite first-stage sample size and a large second-stage sample size. Then the performance of the normalized asymptotic distribution of the MLE, UQ, was analyzed and compared to popular alternatives for a set of simulations.

The distribution of UQ was shown to represent a considerable improvement over the other proposed distributions when n_1 was considerably smaller than n. This was true even when n_1 is moderately large in size.

Since the optimal choice of n_1 was shown to be of the order $\sqrt{n}$ for this model in Lane et al. [17], the usefulness of these findings could have significant implications for many combinations of n_1 and n.

Suppose it is desired that $P(D_1 \leq \sqrt{n}(\widehat{\theta}n - \theta) \leq D_2 = 1 - \alpha)$, where α is the desired confidence level and θ_t is the true parameter. If one was to use the large sample approximate distribution given in (3.1), D_1 and D_2, and therefore n, cannot be determined until after stage 1. However, using (3.1) with $M(\xi_A, \theta)$ in place of $M(\xi^*, \theta)$ or by using UQ on can compute the overall sample size necessary to solve for D_1 and D_2 before stage one is initiated. One could determine n initially using (3.1) with $M(\xi_A, \theta)$ or UQ and then update this calculation after stage-1 data is available. Such same size recalculation requires additional theoretical justification and investigation of their practical usefulness.

We have not, in this paper, addressed the efficiency of the estimate $\widehat{\theta}_n$. One additional way to improve inference would be to find biased adjusted estimates $\widetilde{\theta}_n$ that are superior to $\widehat{\theta}_n$ for finite samples. We have not investigated the impact on inference of estimating the variances in the distributions of UQ, $\mathcal{N}(0, M^{-1}(\xi^*, \theta))$, $\mathcal{N}(0, M^{-1}(\xi_A, \theta)$, and $\mathcal{N}(0, M^{-1}(x_2, \theta))$. Instead, the distributions themselves are compared. For some details on the question of estimation and consistency see Lane et al. [17] and Yao and Flournoy [20].

Acknowledgment

The authors would like to thank the reviewers for their helpful comments and suggestions.

References

[1] G. Elfving, "Optimum allocation in linear regression theory," *The Annals of Mathematical Statistics*, vol. 23, pp. 255–262, 1952.

[2] J. Kiefer and J. Wolfowitz, "The equivalence of two extremum problems," *Canadian Journal of Mathematics*, vol. 12, pp. 363–366, 1960.

[3] H. Chernoff, "Locally optimal designs for estimating parameters," *The Annals of Mathematical Statistics*, vol. 24, no. 4, pp. 586–602, 1953.

[4] L. V. White, "An extension of the general equivalence theorem to nonlinear models," *Biometrika*, vol. 60, pp. 345–348, 1973.
[5] J. Bartroff, "A new characterization of Elfving's method for high dimensional computation," *Journal of Statistical Planning and Inference*, vol. 142, no. 4, pp. 863–871, 2012.
[6] T. E. O'Brien and G. M. Funk, "A gentle introduction to optimal design for regression models," *The American Statistician*, vol. 57, no. 4, pp. 265–267, 2003.
[7] J. Kiefer and J. Wolfowitz, "Optimum designs in regression problems," *The Annals of Mathematical Statistics*, vol. 30, no. 2, pp. 271–294, 1959.
[8] F. Pukelsheim, *Optimal Design of Experiments*, vol. 50 of *Classics in Applied Mathematics*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pa, USA, 2006.
[9] R. A. Fisher, *The Design of Experiments*, Oliver and Boyd, Edinburgh, UK, 1947.
[10] I. Ford, B. Torsney, and C. F. J. Wu, "The use of a canonical form in the construction of locally optimal designs for nonlinear problems," *Journal of the Royal Statistical Society B*, vol. 54, no. 2, pp. 569–583, 1992.
[11] A. C. Atkinson, A. N. Donev, and R. D. Tobias, *Optimum Experimental Designs with SAS*, vol. 34 of *Oxford Statistical Science Series*, Oxford University Press, Oxford, UK, 2007.
[12] C. Stein, "A two-sample test for a linear hypothesis whose power is independent of the variance," *The Annals of Mathematical Statistics*, vol. 16, no. 3, pp. 243–258, 1945.
[13] G. E. P. Box and W. G. Hunter, "Sequential design of experiments for nonlinear models," in *Proceedings of the Scientific Computing Symposium: Statistics*, J. J. Korth, Ed., pp. 113–137, IBM, White Plains, NY, USA, 1965.
[14] V. V. Fedorov, *Theory of Optimal Experiments*, Academic Press, New York, NY, USA, 1972.
[15] L. V. White, *The optimal design of experiments for estimation of nonlinear models [Ph.D. thesis]*, University of London, London, UK, 1975.
[16] S. D. Silvey, *Optimal Design: An Introduction to the Theory for Parameter Estimation*, Chapman & Hall, London, UK, 1980.
[17] A. Lane, P. Yao, and N. Flournoy, Information in a Two-Stage Adaptive Optimal Design, Isaac Newton Institute for the Mathematical Sciences Preprint Series, Cambridge, UK, 2012, http://www.newton.ac.uk/preprints/NI12012.pdf.
[18] V. Dragalin, V. Fedorov, and Y. Wu, "Adaptive designs for selecting drug combinations based on efficacy-toxicity response," *Journal of Statistical Planning and Inference*, vol. 138, no. 2, pp. 352–373, 2008.
[19] V. Fedorov, N. Flournoy, Y. Wu, and R. Zhang, *Best Intention Designs in Dose-Finding Studies*, Isaac Newton Institute Preprint Series, Isaac Newton Institute for Mathematical Sciences, 2011, http://www.newton.ac.uk/preprints.html.
[20] P. Yao and N. Flournoy, "Information in a two-stage adaptive optimal design for normal random variables having a one parameter exponential mean function," in *MoDa 9—Advances in Model-Oriented Design and Analysis*, A. Giovagnoli, A. C. Atkinson, B. Torsney, and C. May, Eds., pp. 229–236, Springer, New York, NY, USA, 2010.
[21] T. S. Ferguson, *A Course in Large Sample Theory*, Texts in Statistical Science Series, Chapman & Hall/CRC, Boca Raton, Fla, USA, 1996.
[22] E. L. Lehmann, *Elements of Large-Sample Theory*, Springer Texts in Statistics, Springer, New York, NY, USA, 1999.
[23] W. Cochran, "Experiments for nonlinear functions," *The Journal of the American Statistical Association*, vol. 68, no. 344, pp. 771–781, 1973.
[24] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, "On the Lambert *W* function," *Advances in Computational Mathematics*, vol. 5, no. 4, pp. 329–359, 1996.

Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2012, Article ID 567819, 10 pages
doi:10.1155/2012/567819

Research Article

Incorporating a Patient Dichotomous Characteristic in Cancer Phase I Clinical Trials Using Escalation with Overdose Control

Mourad Tighiouart, Galen Cook-Wiens, and André Rogatko

Biostatistics and Bioinformatics Research Center, Samuel Oschin Comprehensive Cancer Institute, 8700 Beverly Boulevard, PACT, Suite 900C, Los Angeles, CA 90048, USA

Correspondence should be addressed to Mourad Tighiouart, mourad.tighiouart@cshs.org

Received 29 June 2012; Accepted 5 September 2012

Academic Editor: Yichuan Zhao

We describe a design for cancer phase I clinical trials that takes into account patients heterogeneity thought to be related to treatment susceptibility. The goal is to estimate the maximum tolerated dose (MTD) given patient's specific dichotomous covariate value. The design is Bayesian adaptive and is an extension of escalation with overdose control (EWOC). We will assess the performance of this method by comparing the following designs via extensive simulations: (1) design using a covariate; patients are accrued to the trial sequentially and the dose given to a patient depends on his/her baseline covariate value, (2) design ignoring the covariate; patients are accrued to the trial sequentially and the dose given to a patient does not depend on his/her baseline covariate value, and (3) design using separate trials; in each group, patients are accrued to the trial sequentially and EWOC is implemented in each group. These designs are compared with respect to safety of the trial and efficiency of the estimates of the MTDs via extensive simulations. We found that ignoring a significant baseline binary covariate in the model results in a substantial number of patients being overdosed. On the other hand, accounting for a nonsignificant covariate in the model has practically no effect on the safety of the trial and efficiency of the estimates of the MTDs.

1. Introduction

The main objective of cancer phase I clinical trials is to determine a maximum tolerated dose (MTD) of a new experimental drug or combination of known drugs for use in a phase II trial. These trials enroll advanced stage cancer patients who have exhausted all standard therapies sequentially in cohorts of size one or more patients and dose level assignment to a given cohort of patients is dependent upon the dose levels and toxicity outcomes of the previously treated cohorts of patients. Adaptive statistical designs for cancer phase I clinical trials have been studied extensively in the last two decades, see for example, O'Quigley et al.

[1], Durham and Flournoy [2], Korn et al. [3], Whitehead [4], Babb et al. [5], Gasparini and Eisele [6], Mukhopadhyay [7], and Haines et al. [8]. See also Ting [9] and Chevret [10] for a more comprehensive review of these statistical designs.

A key assumption implied by the definition of the phase I target dose (MTD) is that every subgroup of the patient population has the same MTD. That is, it is assumed that the patient population is homogeneous in terms of treatment tolerance and every patient should be treated at the same dose. As a result, no allowance is made for individual patient differences in susceptibility to treatment. Recent progress in our understanding of pharmacokinetics and the genetics of drug metabolism has led to the development of new strategies of drug allocation that accommodate individual patient needs, see [11–13]. For example, Newell [14] showed how impaired renal function can result in reduced clearance of carboplatin and a dosing formulae based on renal function was developed. In this paper, we present design operating characteristics of a design proposed by Babb et al. [5] known as escalation with overdose control (EWOC) by accounting for patients heterogeneity thought to be related to treatment susceptibility. In the case of a binary covariate, we will assess the performance of this method by comparing the following designs via extensive simulations: (1) design using a covariate; patients are accrued to the trial sequentially and the dose given to a patient depends on his/her covariate value, (2) design ignoring the covariate; patients are accrued to the trial sequentially and the dose given to a patient does not depend on his/her covariate value, and (3) design using separate trials; in each group, patients are accrued to the trial sequentially and EWOC is implemented in each group. O'Quigley et al. [15] investigated the performance of a two-stage continual reassessment method (CRM) using a binary covariate. They considered 3 different models for the dose-toxicity relationship and maximum likelihood method was used to estimate the model parameters. This required starting the escalation scheme using some ad hoc mechanism until the first toxicity is observed. They found that significant gains can be made using the two-sample CRM when there are group imbalances. However, there may not be enough patients in one group to detect that effect. O'Quigley and Paoletti [16] considered a two-group CRM design incorporating ordering of the two groups with respect to treatment tolerability in designing a phase I trial. Babb and Rogatko [17] extended EWOC to allow the utilization of information concerning individual patient differences in susceptibility to treatment. This was applied to a trial involving patients with advanced adenocarcinomas of gastrointestinal origin treated with PNU-214565 (PNU). PNU is a murine Fab fragment of the monoclonal antibody 5T4 fused to a mutated superantigen staphylococcal enterotoxin A (SEA). Preclinical testing demonstrated that the action of PNU is moderated by the neutralizing capacity of anti-SEA antibodies. Consequently, dose levels were adjusted during the trial according to each patient's pretreatment plasma concentration of anti-SEA antibodies. However, design operating characteristics were not studied.

This paper is organized as follows. Section 2 describes the model likelihood and prior distributions and the conduct of the trial using EWOC scheme for three different designs. We present some simulation results in Section 3 and concluding remarks are presented in Section 4.

2. Method

2.1. Model

In this section, we describe a Bayesian adaptive design which accounts for patient heterogeneity thought to be related to treatment susceptibility. Let $X_{\min}$ and $X_{\max}$ denote

the minimum and maximum dose levels available for use in the trial. Clinicians choose these levels in the belief that $X_{\min}$ is safe when administered to humans and $X_{\max}$ is too toxic, see [18] how these levels were selected for a real prospective trial. Denote by Z the observable baseline binary covariate taking values 0 or 1 and let

$$P_z(x) = \text{Prob}(\text{DLT} \mid \text{Dose} = x, Z = z), \tag{2.1}$$

be the probability of dose limiting toxicity (DLT) for a patient with baseline covariate z and treated with dose level x. For simplicity, we consider the logistic model to describe the dose-toxicity relation

$$P_z(x) = \frac{\exp(\beta_0 + \beta_1 x + \eta z)}{1 + \exp(\beta_0 + \beta_1 x + \eta z)}. \tag{2.2}$$

We assume that $\beta_1 \geq 0$ so that $P_z(x)$ is an increasing function of dose x. This is a reasonable assumption for cytotoxic agents. Model (2.2) implies a constant odds ratio of toxicity between the two groups of patients in the sense that this odds ratio does not depend on the dose level.

The MTD for a patient with covariate value z is defined as the dose $\gamma(z)$ that results in a probability equal to $\theta(z)$ that a DLT will manifest within one cycle of therapy. The value chosen for the target probability of DLT $\theta(z)$ would depend on the nature and consequences of the dose-limiting toxicity; it would be set relatively high when the DLT is a transient, correctable or nonfatal condition, and low when it is life threatening or lethal [5]. We will assume that $\theta(z)$ is constant in z although the methodology can be adapted to different target probabilities of toxicities. In practice, clinicians use a constant target probability of DLT θ since we do not know a priori how the treatment under study affects the different groups of patients defined by their baseline covariate value. It follows from the dose-toxicity model (2.2) that the MTD is

$$\gamma(z) = \frac{1}{\beta_1}(\text{logit}(\theta) - \alpha - \eta z). \tag{2.3}$$

Let $\rho_{0,z}$ be the probability of DLT at the initial dose given to a patient with covariate value z. In the statistical design of a phase I clinical trial, it is convenient to specify the prior distribution on parameters the clinicians can easily interpret. For instance, Babb and Rogatko [17] reparameterized model (2.2) in terms of the MTD associated with the maximum-anticipated plasma concentration of anti-SEA antibodies and the probabilities of DLT when the minimum allowable dose is administered to patients with pretreatment anti-SEA concentrations selected to span the range of this covariate. Here, we reparameterize model (2.2) in terms of $\gamma_0 = \gamma(0)$, $\gamma_1 = \gamma(1)$, and $\rho_{0,0}$. We chose this reparameterization because the MTDs for each group are the parameters of interest. However, other parameterizations such

as difference between the MTDs in both groups are possible. Using the definition of the MTDs and probability of toxicity at the initial dose x_1, one can show that

$$\begin{aligned}\beta_0 &= \frac{1}{(\gamma_0 - x_1)}\left[\gamma_0 \text{logit}(\rho_{0,0}) - x_1 \text{logit}(\theta)\right],\\ \beta_1 &= \frac{1}{(\gamma_0 - x_1)}\left[\text{logit}(\theta) - \text{logit}(\rho_{0,0})\right],\\ \eta &= \frac{(\gamma_0 - \gamma_1)}{(\gamma_0 - x_1)}\left[\text{logit}(\theta) - \text{logit}(\rho_{0,0})\right].\end{aligned} \tag{2.4}$$

We note that the probability of DLT does not depend on the parameter γ_1 when $z = 0$. Denote by $p(\rho_{0,0}, \gamma_0, \gamma_1, x_1)$ and $p(\rho_{0,0}, \gamma_0, x_0)$ the probabilities of DLT for a patient with covariate value 1 and 0, respectively. These probabilities are obtained using the dose-toxicity model (2.2) with $\beta_0,\ \beta_1,\ \eta$ given by (2.4).

2.2. Likelihood

Suppose that after the lth patient with baseline covariate value z is treated with dose $x_{z,l}$, there are m_l patients with covariate value $z = 0$ and k_l patients with covariate value $z = 1$. Let $y_{z,l}$ be the toxicity outcome (1 for DLT and 0 for no DLT) for the ith patient with covariate value z. The likelihood of the data is

$$\begin{aligned}L(\rho_{0,0}, \gamma_0, \gamma_1 \mid D_l) = &\prod_{i=1}^{m_l}\left[p(\rho_{0,0}, \gamma_0, x_{0,i})^{y_{0,i}}\left(1 - p(\rho_{0,0}, \gamma_0, x_{0,i})\right)^{1-y_{0,i}}\right]\\ &\times \prod_{j=1}^{k_l}\left[p(\rho_{0,0}, \gamma_0, \gamma_1, x_{1,i})^{y_{1,i}}\left(1 - p(\rho_{0,0}, \gamma_0, \gamma_1, x_{1,i})\right)^{1-y_{1,i}}\right],\end{aligned} \tag{2.5}$$

where $D_l = \{(x_{0,1}, y_{0,1}), \ldots, (x_{0,m_l}, y_{0,m_l}), (x_{1,1}, y_{1,1}), \ldots, (x_{1,k_l}, y_{0,k_l})\}$ and $m_l + k_l = l$.

Let $h(\rho_{0,0}, \gamma_0, \gamma_1)$ be a prior distribution on the parameters $\rho_{0,0}$, γ_0, and γ_1. The posterior distribution is

$$\pi(\rho_{0,0}, \gamma_0, \gamma_1 \mid D_l) = c(D_l)L(\rho_{0,0}, \gamma_0, \gamma_1 \mid D_l)h(\rho_{0,0}, \gamma_0, \gamma_1), \tag{2.6}$$

where $c(D_l)$ is a normalizing constant. This joint posterior is clearly intractable and WinBugs and a Markov chain Monte Carlo sampler will be devised to estimate features of this joint posterior distribution as in Tighiouart et al. [18].

2.3. Prior Distributions

Another advantage of the reparameterization in (2.4) is the natural specification of vague but proper prior densities for the model parameters. Indeed, under the assumption that γ_0, γ_1 belong to $[X_{\min}, X_{\max}]$ with prior probability 1 and no prior assumptions on whether one group can tolerate higher doses better than the other, we can take $(\gamma_0, \gamma_1) \sim$ Uniform

$[X_{\min}, X_{\max}]^2$ and γ_0 independent of γ_1. If on the other hand, we have a priori belief that one group can tolerate higher doses better than the other group for example, then (γ_0, γ_1) can be taken to be uniform on the triangle $X_{\min} < \gamma_0 < \gamma_1 < X_{\max}$. Design operating characteristics should be performed when designing prospective trials when considering informative priors. The prior distribution for $\rho_{0,0}$ is taken as a uniform in $[0, \theta]$, which reflects a lack of prior knowledge regarding the probability of DLT at the initial dose.

2.4. Trial Design

Denote by A and B the two groups of patients corresponding to covariate values 0 and 1, respectively. We assume that the support of the MTDs γ_0 and γ_1 are contained in $[X_{\min}, X_{\max}]$. That is, we assume that dose levels $X_{\min}$ and $X_{\max}$ are identified a priori such that γ_0, γ_1 belong to $[X_{\min}, X_{\max}]$ with prior (and hence posterior) probability 1. We note that if the prior distribution $\pi(\gamma_1)$ is independent of the joint prior distribution of $(\rho_{0,0}, \gamma_0)$, then $\pi(\gamma_1)$ is never updated unless a patient in group B is enrolled in the trial. In the case of such priors, the trial proceeds as follows.

The first patient in either group receives the dose $x_1 = X_{\min}$. Let $\Pi_{z,1}$ be the marginal posterior cdf of the MTD γ_z, $z = 0, 1$. Suppose that the first patient belongs to group A. If the second patient belongs to group A, then he or she will receive the dose $x_{0,2} = \Pi_{0,1}^{-1}(\alpha)$ so that the posterior probability of exceeding the MTD γ_0 is equal to the feasibility bound α. If the second patient belongs to group B, then he or she will receive the dose $x_1 = X_{\min}$. In general, the first time a patient is assigned to a given group always receives $x_1 = X_{\min}$ no matter how many patients have been enrolled in the other group. Once l patients have been enrolled in the trial with at least one patient treated in each group, the $l + 1$-st patient with covariate value z receives the dose $x_{z,l+1} = \Pi_{z,l}^{-1}(\alpha)$. The trial proceeds until a total of $n$patients have been accrued. At the end of the trial, we estimate the MTD as $\widehat{\gamma}_z = \Pi_{z,n}^{-1}(\alpha)$, $z = 0, 1$.

3. Simulation Studies

3.1. Comparison of Three Designs

In order to assess the operating characteristics of this design when designing a prospective trial, we explored the behavior of this method when we adjust for a significant covariate. We also evaluated the performance of this design when adjusting for a nonsignificant baseline covariate. Finally, its performance was also explored when two parallel trials are used instead of adjusting for a binary baseline covariate. Therefore, we study design operating characteristics by comparing the following designs.

(i) Design using a covariate; patients are accrued to the trial sequentially and the dose given to a patient is calculated assuming model (2.2).

(ii) Design ignoring the covariate; patients are accrued to the trial sequentially and the dose given to a patient is calculating assuming model (2.2) without the covariate, that is, as in the original EWOC.

(iii) Design using separate trials; in each group, patients are accrued to the trial sequentially and EWOC is implemented in each group.

Comparisons will be carried out under several scenarios for the true values of the MTDs γ_0 and γ_1. Since the main goal of cancer phase I clinical trials is to efficiently estimate the MTD

while protecting patients from potentially toxic side effects, we will assess the safety of the trial and efficiency of the estimate of the MTDs by simulating a large number of trials M under each model and compare the proportion of patients exhibiting DLT, the average bias $\text{bias}_{\text{ave}} = M^{-1}\sum_{i=1}^{M}\widehat{\gamma}_{z,i} - \gamma_{z,\text{true}}$ and the estimated mean square error $\text{MSE} = M^{-1}\sum_{i=1}^{M}(\widehat{\gamma}_{z,i} - \gamma_{z,\text{true}})^2$, where $\widehat{\gamma}_{z,i}$ is the MCMC estimate of the Bayes estimate of the marginal posterior distribution of the MTD at the end of the ith trial, $z = 0, 1$. In addition, the models are further compared with respect to the proportion of patients that were overdosed. Here, a patient with baseline covariate z is overdosed if this patient has been given a dose x such that $x > x^*$, where x^* is defined as the dose for which $P(\text{DLT} \mid x^*, z) = \theta + 0.05$. This probability is calculated using the parameter values from the corresponding scenario.

3.2. Simulation Setup

The simulation results presented below all assume that the feasibility bound $\alpha = 0.25$ and that the dose levels are standardized so that the starting dose x_1 equals to the minimum dose for each trial $X_{\min} = 0$ and all subsequent dose levels are selected from the unit interval. The target probability of DLT is fixed at $\theta = 0.33$, $\rho_{0,0} = 0.05$, and the total sample size is $n = 42$. We consider several scenarios corresponding to combinations of four possible values of γ_0, γ_1, 0.2, 0.4, 0.6, and 0.8. In all simulations, the prior distributions for $\rho_{0,0}, \gamma_0, \gamma_1$ were taken as uniform in $[0, \theta] \times [X_{\min}, X_{\max}]^2$ with $\rho_{0,0}, \gamma_0, \gamma_1$ independent a priori.

For design (1), a patient is randomly selected from either group A or B with equal probability so that the total number of patients in group A, m, equals to the total number of patients in group B, k. For each pair (γ_0, γ_1) in $\{0.2, 0.4, 0.6, 0.8\} \times \{0.2, 0.4, 0.6, 0.8\}$, we simulate 1000 trials and calculate the proportion of patients that were overdosed, the proportion of patients exhibiting DLT, the average bias, and the estimated MSE. For design (2), the covariate for each patient is recorded but it is not taken into account when calculating the dose level for the next patient. Again, we simulate 1000 trials and calculate the proportion of patients that were overdosed, the proportion of patients exhibiting DLT, the average bias, and the estimated MSE. For the third design, separate trials are simulated in each group and the summary statistics are calculated based on 1000 simulated trials in each group. In all cases, the responses $y_{z,i}$ are generated from model (2.2).

3.3. Results

Table 1 gives the overall proportion of patients exhibiting DLT, the proportion exhibiting DLT in each group, the proportion of patients in each group that are overdosed, the bias, and MSE of the estimates of the MTDs when design (i) in Section 3.1 is used. Table 2 gives the summary statistics corresponding to the safety of the trial when design (ii) is used. The overall proportion of patients exhibiting DLT is always less than $\theta = 0.33$ under all entertained scenarios and it is uniformly lower for a design which accounts for the baseline covariate relative to the design ignoring this covariate. The same conclusion holds when comparison of these two designs is carried out within each group. On the other hand, the proportion of patients being overdosed in group A is much higher when the two groups of patients differ in their susceptibility to treatment and this difference is not taken into account. This proportion can be as high as 16% in the case where $(\gamma_0, \gamma_1) = (0.4, 0.8)$. This is not surprising because when a difference in the MTDs is not taken into account in the model, then the sequence of doses generated by the design tends to cluster around a weighted average of

Table 1: EWOC with Covariate. Design operating characteristic with respect to safety and efficiency of the trial.

(γ_0, γ_1)	0.2, 0.4	0.2, 0.6	0.2, 0.8	0.4, 0.6	0.4, 0.8	0.6, 0.8
Proportion of DLTs	0.3032	0.2735	0.2505	0.2442	0.2231	0.1954
Proportion of DLTs in group A	0.3058	0.2758	0.2495	0.2432	0.2230	0.1932
Proportion of DLTs in group B	0.3007	0.2713	0.2514	0.2451	0.2232	0.1975
Proportion overdosed in group A	0.5958	0.6236	0.6199	0.3174	0.3738	0.1029
Proportion overdosed in group B	0.0934	0.0448	0.0102	0.0373	0.0044	0.0019
Bias (γ_1)	−0.0090	−0.0122	−0.0174	−0.0326	−0.0432	−0.0910
Bias of (γ_2)	−0.0585	−0.1218	−0.2185	−0.1075	−0.2014	−0.2013
MSE (γ_1)	0.0484	0.0501	0.0505	0.0916	0.0968	0.1476
MSE (γ_2)	0.1068	0.1711	0.2622	0.1635	0.2451	0.2437

Table 2: EWOC with no Covariate. Design operating characteristic with respect to safety of the trial.

(γ_0, γ_1)	0.2, 0.4	0.2, 0.6	0.2, 0.8	0.4, 0.6	0.4, 0.8	0.6, 0.8
Proportion of DLTs	0.3156	0.3172	0.3143	0.2824	0.2702	0.2370
Proportion of DLTs in group A	0.3170	0.3180	0.3105	0.2856	0.2737	0.2389
Proportion of DLTs in group B	0.3142	0.3164	0.3180	0.2791	0.2668	0.2350
Proportion overdosed in group A	0.6495	0.7415	0.775	0.4091	0.5298	0.1761
Proportion overdosed in group B	0.0184	0.0006	0.0000	0.0109	0.0000	0.0003

the two "true" MTDs, where the weights depends on the number of patients in each group. If on the other hand, the models accounts for the difference in the MTDs, then the distribution of the sequence of doses is bimodal clustering around the two "true" MTDs, as displayed in Figure 1, which shows the histogram of all doses with fitted density (dashed line) when $(\gamma_0, \gamma_1) = (0.3, 0.6)$. The difference in the proportion of patients being overdosed in group B between the two designs is practically negligible. When the two MTDs are equal and the design does account for the baseline covariate, Tables 3 and 4 show that the proportions of DLTs (overall and within each group) is no more than the target probability of DLT θ, and the differences in these proportions between the two designs are practically not important. No design is uniformly better than the other in terms of the proportion of patients being overdosed. For instance, when $\gamma_0 = \gamma_1 = 0.2$, the proportion of patients in group A is 0.336 when a covariate is used and this proportion is 0.275 when the covariate is not taken into account. On the other hand, when $\gamma_0 = \gamma_1 = 0.4$, the proportion of patients in group A is 0.179 when a covariate is used but this proportion is 0.21 when the covariate is not taken into account. In fact, these proportions are equal on the average across the four scenarios for the true value of the MTD $\gamma = \gamma_0 = \gamma_1$. Tables 3 and 4 also show that the bias and MSE of the estimates of the MTD is higher when a nonsignificant covariate is used in the model, with the higher values obtained the true MTD is high, $\gamma = 0.6, 0.8$.

Table 5 gives the summary statistics when separate trials enrolling $n = 21$ patients are used. As before, the proportion of patients exhibiting DLT does not exceed the target probability of DLT θ. When the true MTDs are the same, the overall proportion of patients that are overdosed using a model with a baseline covariate (Table 3) is lower than the corresponding proportion if parallel trials were used (Table 5) when $\gamma = 0.2, 0.4$. The differences in these proportions for $\gamma = 0.6, 0.8$ are negligible. The bias of the estimate of the MTD is about the same for both designs. In the case where the MTDs γ_0 and γ_1 are

Table 3: EWOC with Covariate. Design operating characteristic with respect to safety and efficiency of the trial.

$\gamma_0 = \gamma_1$	0.2	0.4	0.6	0.8
Proportion of DLTs	0.3337	0.2721	0.2183	0.1759
Proportion of DLTs in group A	0.3319	0.2751	0.2222	0.1764
Proportion of DLTs in group B	0.3356	0.2691	0.2143	0.1754
Proportion overdosed in group A	0.3361	0.1790	0.04995	0.0008
Proportion overdosed in group B	0.3346	0.1802	0.0497	0.0010
Bias (γ_1)	−0.0134	−0.0387	−0.0965	−0.2002
MSE (γ_1)	0.0716	0.1373	0.2142	0.3425

Table 4: EWOC with no Covariate. Design operating characteristic with respect to safety and efficiency of the trial.

$\gamma_0 = \gamma_1$	0.2	0.4	0.6	0.8
Proportion of DLTs	0.3243	0.2939	0.2527	0.2109
Proportion of DLTs in group A	0.3251	0.2944	0.2515	0.2083
Proportion of DLTs in group B	0.3236	0.2934	0.2539	0.2134
Proportion overdosed in group A	0.2755	0.2108	0.0774	0.0023
Proportion overdosed in group B	0.2761	0.2050	0.0749	0.0026
Bias (γ_1)	−0.0119	−0.0197	−0.0537	−0.1198
MSE (γ_1)	0.0353	0.0684	0.1032	0.1537

Table 5: EWOC. Design operating characteristic with respect to safety and efficiency of the trial.

γ	0.2	0.4	0.6	0.8
Proportion of DLTs	0.3372	0.2778	0.2162	0.1737
Proportion overdosed	0.3684	0.2067	0.0417	0.0001
Bias (γ)	−0.0086	−0.0249	−0.08124	−0.1915
MSE (γ)	0.0464	0.0858	0.1381	0.2246

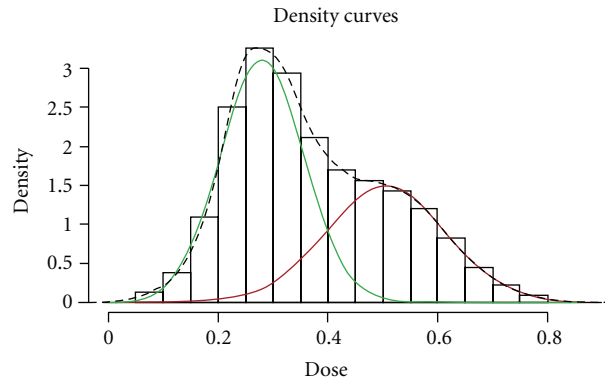


Figure 1: Histogram and fitted density (dashed line) of dose allocations for patients 2 through 80 based on 1000 simulated trials using the model with covariate. The true MTDs are $\gamma_0 = 0.3$, $\gamma_1 = 0.6$, and $\rho_0 = 0.05$, $\theta = 0.33$, $\alpha = 0.25$.

different, more patients are overdosed using a model with a baseline covariate compared to using parallel trials and the differences in the bias and MSE are negligible, see Tables 1 and 5.

Based on these results, we recommend adjusting for a baseline covariate thought to be related to treatment susceptibility when designing a cancer phase I trials whenever possible. We stand to lose little if we were to use a design with a covariate when in fact there is no difference between the MTDs of the two groups.

4. Discussion

We have presented design operating characteristics of a Bayesian adaptive design which accounts for a patient dichotomous baseline covariate using EWOC scheme. The design is suitable for cancer phase I clinical trials where the goal is to estimate the conditional MTD given patients' covariate value.

We have found that if the two MTDs are different and the design does not adjust for this heterogeneity, then the trial will result in more patients being overdosed. If the two MTDs are the same and the design adjusts for patients' heterogeneity, then slightly more patients can be overdosed if the true MTD is low relative to a design with no covariate but these proportions are equal on the average across the four scenarios for the true value of the MTD. Thus, we stand to lose little if we do include a statistically nonsignificant covariate in the model. Incidentally, this conclusion is in agreement with the findings in O'Quigley et al. [15]. We carried out other simulations (results not shown) for various sample sizes, allocation ratios, probability of DLT at the initial dose ρ_0. The results and conclusions were essentially the same. Ratain et al. [13] showed the importance of including patient's plasma concentration of anti-SEA antibodies in order to determine the MTD of the agent PNU-214565 as a function of this continuous baseline covariate. In a similar trial, Tighiouart and Rogatko [16] showed how more patients were overdosed when a baseline covariate, cancer type, was not accounted for in the model. Indeed, a retrospective analysis of a cancer phase I trial using a baseline continuous covariate showed that nonsmall cell cancer patients and pancreatic patients were treated at suboptimal doses whereas renal cell carcinoma patients were overdosed, with 36.4% experiencing DLT; the target probability of DLT was $\theta = 0.2$. This last example is in agreement with the simulation results we obtained in this paper. We are in the process of determining model operating characteristics in the presence of a continuous covariate, more than one covariate, and interaction term. An important question is to decide whether or not to include patients' covariate values during the trial. Although the previous results seem to indicate that we stand to lose little in terms of the proportion of patients being overdoses and the efficiency of the estimate of the MTDs when covariate information is taken into account in the model when in fact, this covariate is not predictive of DLT, determining the value of this covariate may involve a monetary cost. This is the case when patients need to be genotyped and certain biomarker expressions need to be determined.

Acknowledgments

Supported in part by the National Center for Research Resources, Grant UL1RR033176, and is now at the National Center for Advancing Translational Sciences, Grant UL1TR000124 (M. Tighiouart and A. Rogatko), Grant UL1TR000124 (M. Tighiouart and A. Rogatko), Grant 5P01CA098912-05 (A. Rogatko) and P01 DK046763 (A. Rogatko). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References

[1] J. O'Quigley, M. Pepe, and L. Fisher, "Continual reassessment method: a practical design for phase 1 clinical trials in cancer," *Biometrics*, vol. 46, no. 1, pp. 33–48, 1990.

[2] S. D. Durham and N. Flournoy, *Random Walks for Quantile Estimation*, Springer, New York, NY, USA, 1994.

[3] E. L. Korn, D. Midthune, T. T. Chen, L. V. Rubinstein, M. C. Christian, and R. M. Simon, "A comparison of two phase I trial designs," *Statistics in Medicine*, vol. 13, no. 18, pp. 1799–1806, 1994.

[4] J. Whitehead, "Bayesian decision procedures with application to dose-finding studies," *International Journal of Pharmaceutical Medicine*, vol. 11, no. 4, pp. 201–208, 1997.

[5] J. Babb, A. Rogatko, and S. Zacks, "Cancer Phase I clinical Trials: efficient dose escalation with overdose control," *Statistics in Medicine*, vol. 17, pp. 1103–1120, 1998.

[6] M. Gasparini and J. Eisele, "A curve-free method for phase I clinical trials," *Biometrics*, vol. 56, no. 2, pp. 609–615, 2000.

[7] S. Mukhopadhyay, "Bayesian nonparametric inference on the dose level with specified response rate," *Biometrics*, vol. 56, no. 1, pp. 220–226, 2000.

[8] L. M. Haines, I. Perevozskaya, and W. F. Rosenberger, "Bayesian optimal designs for phase I clinical trials," *Biometrics*, vol. 59, no. 3, pp. 591–600, 2003.

[9] N. Ting, *Dose Finding in Drug Development*, Springer, New York, NY, USA, 1st edition, 2006.

[10] S. Chevret, *Statistical Methods for Dose-Finding Experiments*, John Wiley & Sons, Chichester, UK, 2006.

[11] G. Decoster, G. Stein, and E. E. Holdener, "Original article: responses and toxic deaths in Phase I clinical trials," *Annals of Oncology*, vol. 1, no. 3, pp. 175–181, 1990.

[12] M. J. Ratain, R. Mick, R. L. Schilsky, and M. Siegler, "Statistical and ethical issues in the design and conduct of phase I and II clinical trials of new anticancer agents," *Journal of the National Cancer Institute*, vol. 85, no. 20, pp. 1637–1643, 1993.

[13] M. J. Ratain, R. Mick, L. Janisch et al., "Individualized dosing of amonafide based on a pharmacodynamic model incorporating acetylator phenotype and gender," *Pharmacogenetics*, vol. 6, no. 1, pp. 93–101, 1996.

[14] D. R. Newell, "Pharmacologically based phase I trials in cancer chemotherapy," *Hematology*, vol. 8, pp. 257–275, 1994.

[15] J. O'Quigley, L. Z. Shen, and A. Gamst, "Two-sample continual reassessment method," *Journal of Biopharmaceutical Statistics*, vol. 9, no. 1, pp. 17–44, 1999.

[16] J. O'Quigley and X. Paoletti, "Continual reassessment method for ordered groups," *Biometrics*, vol. 59, no. 2, pp. 430–440, 2003.

[17] J. S. Babb and A. Rogatko, "Patient specific dosing in a cancer phase I clinical trial," *Statistics in Medicine*, vol. 20, no. 14, pp. 2079–2090, 2001.

[18] M. Tighiouart, A. Rogatko, and J. S. Babb, "Flexible Bayesian methods for cancer phase I clinical trials. Dose escalation with overdose control," *Statistics in Medicine*, vol. 24, no. 14, pp. 2183–2196, 2005.