Published in final edited form as:

J Biopharm Stat. 2019 ; 29(4): 648–662. doi:10.1080/10543406.2019.1632881.

On the Relative Efficiency of Model-Assisted Designs: A Conditional Approach

Ruitao Lin, Ying Yuan

Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA.

Abstract

In phase I dose-finding trials, model-assisted designs are a novel class of designs that combine the simplicity of algorithm-based methods with the superior performance of model-based methods. Examples of model-assisted designs include the modified toxicity probability (mTPI), Bayesian optimal interval (BOIN) and keyboard designs. To achieve simplicity, these model-assisted methods model only “local” data observed at the current dose, typically using a binomial model, to guide dose assignments. This potentially causes efficiency loss, however, by ignoring the data observed in other doses. To investigate this issue, we propose a conditional approach that utilizes the data from both current and adjacent (i.e., next higher or lower) doses to make the dose-assignment decisions. Specifically, we propose the conditional optimal interval (COIN) design, as the conditional approach extension of the BOIN design. We investigate the theoretical properties of the COIN design, and conduct extensive numerical studies to examine its performance in comparison with existing model-assisted designs. We also present the conditional approach to the keyboard design. We observe that the conditional approach improves patient allocation, but yields similar maximum tolerated dose (MTD) identification accuracy as the model-assisted designs, suggesting only minor efficiency loss using local data under the model-assisted designs.

Keywords

Adaptive designs; dose finding; maximum tolerated dose; model-assisted designs; phase I trial

1 Introduction

Phase I dose-finding trials in oncology aim to determine the maximum tolerated dose (MTD) of a new drug, which is the highest dose level with the probability of dose-limiting toxicity (DLT) closest to the given target toxicity probability.¹ Under the assumption of the monotonically increasing dose–toxicity curve, many different phase I dose-finding designs have been proposed to determine the dose levels for the sequentially enrolled patients. The algorithm-based methods, mostly the 3+3 design,² are dominant in practice. The 3+3 design is transparent and easy to implement, but has poor operating characteristics.³ Model-based designs are proposed to improve the performance of dose finding. The most well-known example is the continual reassessment method (CRM).⁴ The CRM produces better operating

characteristics than the 3+3 design, but is less transparent and more complicated to implement in practice. For a comprehensive overview of the recent development in phase I dose-finding trials, see Liu et al. (2013)⁵ and Yuan et al. (2016).⁶

Recently, model-assisted designs have emerged as an attractive approach for phase I clinical trials, as they combine the simplicity of algorithm-based designs with the superior performance of model-based designs.^{7–9} Similar to model-based designs, model-assisted designs use simple probability models for efficient decision making, while their dose escalation and de-escalation rules can be tabulated before the onset of a trial, as with algorithm-based designs. Ivanova et al. (2007)¹⁰ introduced a cumulative cohort design based on the asymptotic distribution of patient allocation. Ji et al. (2010)¹¹ proposed the modified toxicity probability interval (mTPI) design that utilizes unit probability mass (UPM) to guide the dose assignment. Liu and Yuan (2015)¹² proposed the Bayesian optimal interval (BOIN) design that makes the decision of dose escalation and de-escalation by simply comparing the observed toxicity rate at the current dose with two prespecified boundaries, optimized to minimize the incorrect decision of dose assignment. Yan et al. (2017)⁷ noted the overdosing issue of the mTPI design due to the use of UPM and proposed the keyboard design as a seamless improvement of mTPI to achieve higher accuracy of MTD identification and better overdose control. Mu et al. (2018)¹³ proposed a generalized BOIN design that can handle toxicity grades, binary or continuous toxicity endpoints under a unified framework. Yuan et al. (2018)¹⁴ developed the time-to-event BOIN design to handle late-onset toxicities or fast patient accrual. Lin (2018)¹⁵ generalized the BOIN design to handle multiple toxicities. Lin and Yin (2017),¹⁶ Zhang and Yuan (2016),¹⁷ and Pan, et al. (2018)¹⁸ extended the BOIN and keyboard designs to drug-combination trials. Zhou et al. (2018a, 2018b)⁸ reviewed and compared the performance of the model-assisted designs, and found that they—especially BOIN and keyboard designs—yield competitive performance, compared to that of the CRM. As a side note, the aforementioned designs sometimes are referred to as interval designs as they involve dividing toxicity probability into intervals. We suggest refraining from using that terminology because some typical model-based designs, e.g., Bayesian logistic regression model design (a modification of the CRM),¹⁹ also use the toxicity intervals for determining dose escalation and de-escalation. In general, we prefer “model-assisted design” as it more accurately describes the most important feature of this class of designs, i.e., they offer a middle ground between algorithm-based and model-based designs, and possess the features of them. In addition, we have found that the terminology “model-assisted design” is more accessible and better received by practitioners and non-statisticians.

Unlike the model-based design, such as the CRM, which assumes a dose–toxicity curve across all doses, the model-assisted designs consider only “local” data, i.e., the data observed at the current dose. They typically use a binomial model, which renders it possible to enumerate the dose escalation and de-escalation rules before the trial begins for simple implementation. A potential concern for model-assisted designs is whether using only “local” data at the current dose, while ignoring the data observed from the other doses, causes the loss of efficiency in MTD identification; if so, by how much? To investigate the relative efficiency of model-assisted designs, we propose a conditional approach that utilizes the data information from both current and adjacent (i.e., next higher or lower) doses. We

mainly focus on the BOIN and keyboard designs, as they have very similar operating characteristics and both outperform mTPI.^{7–9, 20} We propose the conditional optimal interval (COIN) design, as the conditional approach counterpart of BOIN. The COIN design is built upon the framework of BOIN, but each dose assignment decision of COIN is made by conditioning on the observed information from the next higher or next lower dose level. As a result, the COIN design utilizes more data in decision making than the BOIN design. We also briefly discuss the extension of the conditional approach to the keyboard design. We conduct extensive numerical studies to examine the performance of COIN in comparison with existing model-assisted designs.

The arrangement of this paper is as follows: Section 2 first provides a brief introduction of BOIN and then derives the proposed COIN design. Several properties of COIN are discussed. Section 3 presents various simulation results to demonstrate the efficiencies of COIN and model-assisted designs. Section 4 concludes with some remarks. All technical details are provided in the Appendix.

2 Methods

2.1 Bayesian Optimal Interval Design

As COIN is developed using a similar framework as BOIN, for ease of exposition, we first review the BOIN design to set up the notation and basis for introducing our new method. Consider a phase I dose-finding trial with a target toxicity rate of ϕ , J dose levels are prespecified for which the true toxicity rates are assumed to increase monotonically, i.e., $p_1 < \dots < p_J$. Suppose that the current dose level is j , at which n_j total number of patients have been treated and y_j cumulative number of DLTs are observed. Let $O_j = (y_j, n_j)$ denote the “local” data observed, and $\hat{p}_j = y_j/n_j$ denote the observed toxicity rate at the current dose. Let $\lambda_{Lj}(n_j, \phi)$ and $\lambda_{Uj}(n_j, \phi)$ respectively denote dose escalation and de-escalation boundaries that are arbitrary functions of dose level j (as shown by subscript j), n_j and ϕ .

The BOIN design starts with a class of designs with the following dose escalation and de-escalation rule:

- If $\hat{p}_j \leq \lambda_{Lj}(n_j, \phi)$, then the dose for the next cohort is escalated to level $j + 1$.
- If $\hat{p}_j \geq \lambda_{Uj}(n_j, \phi)$, then the dose for the next cohort is de-escalated to level $j - 1$.
- If $\lambda_{Lj}(n_j, \phi) < \hat{p}_j < \lambda_{Uj}(n_j, \phi)$ or the next assignment is outside of the prespecified dose range, then the next cohort is treated at the same dose level j .

Because $\lambda_{Lj}(n_j, \phi)$ and $\lambda_{Uj}(n_j, \phi)$ are arbitrary functions and free to take distinct values for each possible values of dose level j , n_j and ϕ , this class of designs are extremely flexible and contain all possible local-data designs that make the decision of dose assignment based on local data. In other words, the mTPI, BOIN and keyboard designs are all special cases of the above design.

Within the class of local-data designs, BOIN chooses the optimal pair of $\lambda_{Lj}(n_j, \phi)$ and $\lambda_{Uj}(n_j, \phi)$, which minimize the probability of incorrect decision of dose assignment, by

formulating the dose-finding problem in a Bayesian decision-making framework. Specifically, three hypotheses at dose level j are considered,

$$H_{0j}: p_j = \phi, H_{1j}: p_j = \phi_1, H_{2j}: p_j = \phi_2,$$

where $\phi_1 < \phi < \phi_2$ are design parameters that should be prespecified, H_{0j} means that the current level j is the target dose, H_{1j} denotes that level j sub-therapeutic such that dose escalation should be pursued, and H_{2j} indicates that level j is overly toxic such that dose de-escalation is needed. A Bayesian interpretation of the hypotheses is that a point mass distribution is assigned to p_j under H_{sj} , $s = 0, 1, 2$. By minimizing the probability of incorrect decisions, the optimal boundaries $\lambda_{Lj}(n_j, \phi)$ and $\lambda_{Uj}(n_j, \phi)$ are given by

$$\lambda_{Lj}(n_j, \phi) = \frac{\log\left(\frac{1-\phi_1}{1-\phi}\right) + n_j^{-1} \log \pi_{10j}}{\log\left\{\frac{\phi(1-\phi_1)}{\phi_1(1-\phi)}\right\}}, \lambda_{Uj}(n_j, \phi) = \frac{\log\left(\frac{1-\phi}{1-\phi_2}\right) + n_j^{-1} \log \pi_{02j}}{\log\left\{\frac{\phi_2(1-\phi)}{\phi(1-\phi_2)}\right\}}, \quad (2.1)$$

where $\pi_{stj} = \Pr(H_{sj})/\Pr(H_{tj})$, $s, t = 0, 1, 2$ denotes the prior odds of the hypothesis H_{sj} versus H_{tj} . When the three hypotheses are assumed to be equally probable *a priori*, we have $\pi_{10j} = \pi_{02j} = 1$, and thus the optimal dose escalation and de-escalation boundaries are independent of j and n_j . This makes the BOIN design particularly easy to implement as its dose assignment decision only involves comparing the observed toxicity rate $\hat{p}_j$ with a pair of fixed boundaries.

Liu and Yuan (2015)¹² showed that the optimal interval boundaries $\lambda_{Lj}(n_j, \phi)$ and $\lambda_{Uj}(n_j, \phi)$ also can be derived by considering the pairwise comparisons of H_{1j} versus H_{0j} and H_{0j} versus H_{2j} . As a result, in an equivalent way, the trial conducted using BOIN can proceed according to posterior hypothesis probabilities $\Pr(H_{sj} | O_j)$, $s = 0, 1, 2$,

- If $\hat{p}_j < \phi$ and $\Pr(H_{1j} | O_j) > \Pr(H_{0j} | O_j)$, the dose for the next cohort is escalated to level $j + 1$.
- If $\hat{p}_j \geq \phi$ and $\Pr(H_{2j} | O_j) > \Pr(H_{0j} | O_j)$, the dose for the next cohort is de-escalated to level $j - 1$.
- Otherwise, the next cohort is treated at the same dose level j .

At the end of the trial, the pool adjacent violators algorithm (PAVA)²¹ is performed on the estimated toxicity rates $\hat{p}_j$, $j = 1, \dots, J$ in order to maintain the monotonic constraint on the estimated dose-response curve. The dose level with the final estimated toxicity rate closest to the target toxicity rate ϕ is selected as the MTD. An interesting fact we note is that when the noninformative prior is used (i.e., three hypotheses are assumed to be equally probable *a priori*), because $\Pr(H_{1j} | O_j) > \Pr(H_{0j} | O_j)$ is equivalent to $\Pr(O_j | H_{1j}) > \Pr(O_j | H_{0j})$, BOIN also can be interpreted as a frequentist design that makes the decision of dose escalation/de-escalation based on the likelihood ratio test. Because the likelihood ratio test is uniformly

most powerful, BOIN also is optimal from the frequentist viewpoint among all local-data designs.

As described previously, the BOIN design only uses the local data to determine dose transition, and ignores the data observed in other doses. A natural question is whether this causes a loss of efficiency in finding the MTD and allocating patients to the MTD. This motivates the COIN design, which uses more data for decision making, as described below.

2.2 Conditional Optimal Interval Design

To account for more observed evidence, we consider borrowing information from adjacent dose levels for deciding dose escalation or de-escalation. Let p_{j-1} and p_{j+1} denote the toxicity rates of the next lower and higher dose levels for the current dose level, respectively. Let $O_j^E = (y_j, n_j, y_{j+1}, n_{j+1})$ denote the observed data of dose level j and $j+1$, and $O_j^D = (y_j, n_j, y_{j-1}, n_{j-1})$ denote the observed data of dose level j and $j-1$, where $y_0 = n_0 = y_{J+1} = n_{J+1} = 0$ is specified for notational simplicity. Following the fashion of BOIN, a pairwise comparison is considered for the decision of dose escalation:

$$H_{0j}^E: p_j = \phi, p_{j+1} = \phi_2 \text{ versus } H_{1j}^E: p_j = \phi_1, p_{j+1} = \phi, \quad (2.2)$$

where H_{0j}^E indicates that the current dose level is proper dosing while the next higher dose is too toxic such that dose retainment is required, and H_{1j}^E states that the current dose level is underdosing while the next higher dose is proper dosing such that dose escalation should be made. In hypotheses (2.2), we only assume point masses for p_j and p_{j+1} . Although for most of the time the true DLT rates cannot hit exactly the point masses specified the hypothesis, the hypothesis we construct only represents some typical scenarios when dose escalation or retainment should be made. This is in a similar fashion to the sample size calculation used in phase III trials. The hypotheses H_{0j}^E and H_{1j}^E only consider whether to escalate to dose level $j+1$ or to stay at dose level j , and no dose de-escalation is considered. Like the BOIN design, we model the data observed at levels j and $j+1$ using independent Binomial models. Then, the likelihood ratio for testing H_{1j}^E versus H_{0j}^E is

$$\frac{\Pr(O_j^E | H_{1j}^E)}{\Pr(O_j^E | H_{0j}^E)} = \frac{\phi_1^y \phi_1^{n-y} \phi_2^{y_{j+1}} (1-\phi_2)^{n_{j+1}-y_{j+1}}}{\phi^y \phi^{n-y} \phi^{y_{j+1}} (1-\phi)^{n_{j+1}-y_{j+1}}}.$$

Let λ_{Lj} be the lower interval boundary and let $\mathcal{S}$, $\mathcal{E}$, $\mathcal{D}$ respectively stand for dose stay, escalation and de-escalation, the probability of making an incorrect dose-assignment decision for testing H_{1j}^E versus H_{0j}^E is given by

$$\begin{aligned}
\alpha_E(\lambda_{Lj}) &= \Pr(H_{0j}^E) \Pr(\mathcal{E} | H_{0j}^E) + \Pr(H_{1j}^E) \Pr(\mathcal{S} \text{ or } \mathcal{D} | H_{1j}^E) \quad (2.3) \\
&= \Pr(H_{0j}^E) \Pr(Y_j \leq \lambda_{Lj} n_j, Y_{j+1} = y_{j+1} | H_{0j}^E) \\
&\quad + \Pr(H_{1j}^E) \Pr(Y_j > \lambda_{Lj} n_j, Y_{j+1} = y_{j+1} | H_{1j}^E)
\end{aligned}$$

For simplicity, we assume through this paper that $\Pr(H_{0j}^E) = \Pr(H_{1j}^E)$, i.e., $\pi_{10j} = 1$, but the proposed method can be easily extended to account for the case of $\pi_{10j} \neq 1$. For example, if some prior information indicates that the dose level j is very safe, then we can simply set $\pi_{10j} > 1$. The examination of the interplay between the prior specification π_{10j} and the performance of COIN, however, is beyond the scope of our paper.

Similar to the BOIN design,¹² the conditional optimal interval (COIN) boundary $\lambda_{Lj}(O_j^E, \phi)$ for the observed toxicity rate $\hat{p}_j$ can be derived by minimizing the incorrect decision probability $\alpha_E(\lambda_{Lj})$,

$$\lambda_{Lj}(O_j^E, \phi) = \frac{n_j \log\left(\frac{1-\phi_1}{1-\phi}\right) + \log \pi_{10j} + n_{j+1} \log\left(\frac{1-\phi}{1-\phi_2}\right) + y_{j+1} \log\left\{\frac{\phi(1-\phi_2)}{\phi_2(1-\phi)}\right\}}{n_j \log\left\{\frac{\phi(1-\phi_1)}{\phi_1(1-\phi)}\right\}}. \quad (2.4)$$

As a result, if $\hat{p}_j \leq \lambda_{Lj}(O_j^E, \phi)$, we escalate to dose level $j+1$ for the next patients. In addition, it can be shown that this dose-escalation rule is equivalent to the rule comparing $\Pr(H_{0j}^E | O_j^E)$ and $\Pr(H_{1j}^E | O_j^E)$: if $\Pr(H_{1j}^E | O_j^E) \geq \Pr(H_{0j}^E | O_j^E)$, then dose escalation should be made.

According to the formula of $\lambda_{Lj}(O_j^E, \phi)$, it is observed that when the dose level $j+1$ has not been explored with $n_{j+1} = y_{j+1} = 0$, the conditional optimal boundary $\lambda_{Lj}(O_j^E, \phi)$ reduce to the BOIN boundary $\lambda_{Lj}(n_j, \phi)$. In addition, since $\log\left\{\frac{\phi(1-\phi_2)}{\phi_2(1-\phi)}\right\}$ is negative, if y_{j+1} increases given a fixed n_{j+1} , in other words, the toxicity degree of the dose level $j+1$ increases, then the conditional optimal boundary $\lambda_{Lj}(O_j^E, \phi)$ decreases, which leads to an increasingly impossible dose escalation. Such a pairwise comparison is intuitive. For example, when we are making a decision on whether there is a need to escalate the dose, not only the information from the current dose j but also the information from the next higher level $j+1$ should be incorporated: if level $j+1$ is observed to be overly toxic, the chance of dose escalation should decrease. As a result, the escalation decision is made by conditioning on the information from the next higher dose. Although we only assign the point mass distribution to (p_j, p_{j+1}) under H_{sj}^E , $s = 0, 1$, it actually can deal with various dose-toxicity scenarios. For example, if $\hat{p}_j \approx \phi$ and $\hat{p}_{j+1} > \phi$, based on the binomial likelihood, $\Pr(H_{0j}^E | O_j^E)$ should be greater than $\Pr(H_{1j}^E | O_j^E)$; if $\hat{p}_j < \phi$ and $\hat{p}_{j+1} < \phi$, then $\Pr(H_{0j}^E | O_j^E)$

would be smaller than $\Pr(H_{0j}^E | O_j^E)$; if $\hat{p}_j < \phi$ and $\hat{p}_{j+1} > \phi$, the rank between the posteriors of H_{0j}^E and H_{1j}^E depends on the magnitudes of $\hat{p}_j$ and $\hat{p}_{j+1}$ as well as the sample size.

Similarly, a pairwise comparison is considered for dose de-escalation:

$$H_{0j}^D: p_j = \phi, p_{j-1} = \phi_1 \text{ versus } H_{2j}^D: p_j = \phi_2, p_{j-1} = \phi,$$

where H_{0j}^D indicates that the current dose level is proper dosing while the next lower dose is too safe such that dose retainment is required, and H_{2j}^D states that the current dose level is overdosing while the next lower dose is proper dosing such that dose de-escalation should be made. Similar to the lower interval boundary $\lambda_{Lj}(O_j^E, \phi)$, the upper interval boundary $\lambda_{Uj}(O_j^D, \phi)$ can be derived by minimizing the incorrect decision probability that mimics (2.3),

$$\lambda_{Uj}(O_j^D, \phi) = \frac{n_j \log\left(\frac{1-\phi}{1-\phi_2}\right) + \log \pi_{02j} + n_{j-1} \log\left(\frac{1-\phi_1}{1-\phi}\right) + y_{j-1} \log\left\{\frac{\phi_1(1-\phi)}{\phi(1-\phi_1)}\right\}}{n_j \log\left\{\frac{\phi_2(1-\phi)}{\phi(1-\phi_2)}\right\}}. \quad (2.5)$$

As a result, if $\hat{p}_j \geq \lambda_{Uj}(O_j^D, \phi)$, we de-escalate to dose level $j-1$ for the next patients. The optimal boundary $\lambda_{Uj}(O_j^D, \phi)$ depends on y_{j-1} and n_{j-1} : since $\log\left\{\frac{\phi_1(1-\phi)}{\phi(1-\phi_1)}\right\}$ is negative, if y_{j-1} is smaller given a fixed n_{j-1} , then $\lambda_{Uj}(O_j^D, \phi)$ becomes larger, making dose de-escalation more difficult. It can be shown that this de-escalation rule is equivalent to the rule comparing $\Pr(H_{0j}^D | O_j^D)$ and $\Pr(H_{2j}^D | O_j^D)$: if $\Pr(H_{2j}^D | O_j^D) \geq \Pr(H_{0j}^D | O_j^D)$, then dose de-escalation should be made.

For illustration, we consider the target toxicity rate ϕ to be 0.25 and set the design parameters ϕ_1 and ϕ_2 as the default values of BOIN, i.e., $\phi_1 = 0.6\phi$ and $\phi_2 = 1.4\phi$. The hypotheses of COIN and BOIN are assumed to be equally probable *a priori*. Based on the equations (2.1), (2.4), and (2.5), Figure 1 exhibits the optimal interval boundaries of COIN and BOIN under various scenarios. It can be seen that interval boundaries of BOIN remain unchanged, regardless the data from adjacent dose levels. On the other hand, not only the sample size of the current dose level but also the observed data from the next higher or lower dose level determine those of COIN. When n_j at the current dose level goes to infinity, the interval boundaries of COIN shrink to those of BOIN.

Liu and Yuan (2015)¹² pointed out that the BOIN design (as well as other interval designs) does not possess the short-memory coherence property proposed by Cheung (2005),²² but instead, is long-memory coherent. We show in the Appendix that, with additional information involved in the decision-making process, the proposed COIN enjoys the short-memory coherence.

Theorem 1. Suppose that the dose level d_n for the n th patient is j . The COIN design is coherent in escalation and de-escalation, i.e., for the $(n + 1)$ th patient, $P(d_{n+1} = j - 1 | t_n = 0, d_n = j, O_j^E) = 0$ and $P(d_{n+1} = j + 1 | t_n = 1, d_n = j, O_j^D) = 0$, where t_n is the toxicity outcome for the n th patient.

As a final remark in this section, it is possible to extend the proposed COIN design by borrowing information from more dose levels. For example, the observed data across the dose levels $j - 2, j + 1$ and the current dose level j can be integrated for deciding whether a dose escalation is needed. However, we find that the information of the dose levels is less relevant the further away it is from the current dose. For ease of implementation, the considered COIN method using only the current and the adjacent levels is more straightforward without compromising the performance. In addition, the proposed approach for deriving COIN is not bundled with BOIN but can be generalized based on other model-assisted designs, like the keyboard design. We outline the development of a conditional approach to the keyboard design in the Appendix.

3 Numerical Study

3.1 Fixed Scenarios: COIN versus BOIN

The primary objective is to investigate whether there is an efficiency gain for the model-assisted designs when more information from more dose levels are borrowed for decision making. To do so, we develop a conditional approach to the local-data model-assisted designs. In the present section, simulation studies are carried out to compare the performance of COIN and BOIN. Based on such a head-to-head comparison, we aim to investigate the relative efficiency of the designs that only use local data and those using more data. First, we consider the same four fixed scenarios of Liu and Yuan (2015).¹² Six dose levels with a target toxicity rate of 0.25 are studied. We assume various combinations of the sample size of 36 or 90 and the cohort size of 1 or 3. Equal prior probabilities are assigned to the prespecified hypotheses for both BOIN and COIN. In terms of the design parameters, we utilize $\phi_1 = (1 - \delta)\phi$ and $\phi_2 = (1 + \delta)\phi$, where δ is taken to be either 0.4 or 0.2. For a fair comparison, no dose elimination rule is imposed, and the final MTD selection scheme of BOIN is applied to both designs. A total of 10,000 trials are replicated for each scenario, and the empirical distribution of the final selected dose level and the empirical patient allocation distribution are summarized in order to quantify the efficiency of a trial.

According to the simulation results that are provided in Table 1 ($n = 36$) and 2 ($n = 90$), several findings are highlighted. First, for the considered four scenarios, the correct MTD selection percentage by BOIN is similar to that of COIN. This indicates that using more information for decision making in an interval design may not be beneficial to the MTD identification under the considered scenarios. Second, COIN generally leads to more patients being allocated to the MTD than BOIN. Such an advantage of COIN is more prominent when the sample size is large. For instance, in scenario 2 with 90 patients (Table 2), COIN allocates at least eight more patients to the MTD compared to BOIN. We also consider the selection and allocation percentages in the neighbor of the MTD, which is made up of the dose levels $j^* - 1, j^* + 1$ given that j^* is the MTD. COIN has higher selection and

allocation percentages than BOIN by summing over the results across the three dose levels in the neighbor of the MTD. Only using the information from the current dose level, BOIN has a larger probability to stay at much higher dose levels (i.e., more than one level away from the true MTD). For example, in scenario 1 of Table 1 ($n = 36$) with the cohort size of one, BOIN assigns one more patient to the third dose level than COIN. By contrast, COIN tends to be more concentrated in the neighbor of the MTD. Last, we note that BOIN is relatively more sensitive to the design parameters ϕ_1 and ϕ_2 . For example, in scenario 3 with a sample size of 90 (Table 2) and a cohort size of one, BOIN leads to a more variable empirical patient allocation distribution than COIN for different λ 's: BOIN with $\lambda = 0.2$ has five fewer patients allocated to the MTD and three more patients to the overly toxic levels than BOIN with $\lambda = 0.4$; while the difference between the empirical patient allocation distributions for COIN with $\lambda = 0.2$ and COIN with $\lambda = 0.4$ is within at most two patients. The simulation results indicate that the adjacent dose levels indeed contain extra information in deciding dose escalation or de-escalation. In terms of identifying the MTD, however, the efficiency gain by borrowing more information is minimal.

3.2 Random Scenarios: COIN versus BOIN and mTPI

In order to further demonstrate the efficiency gain of COIN, we perform an objective comparison of COIN, BOIN and mTPI, using randomly-generated scenarios with six or ten dose levels. The performance of the keyboard design is particularly similar to that of BOIN, so it is omitted. The target toxicity rate is considered to be 0.25, 0.30, or 0.35, and the MTD is set at one of the six or ten dose levels with an equal probability. 50,000 random dose-response scenarios are generated by controlling the average probability difference around the target to be 0.1 or 0.15.^{12, 24} That is, the average distance between the toxicity probabilities of the MTD and its adjacent dose levels are controlled to be 0.1 or 0.15. The sample size varies from 18 to 99 with one patient or three patients in a cohort. For mTPI, the trial is conducted by partitioning the unit interval $(0, 1)$ into three subintervals: $(0, \phi - 0.05)$, $(\phi - 0.05, \phi + 0.05)$, and $(\phi + 0.05, 1)$. The configurations of BOIN and COIN remain the same as in Section 3.1. To adequately quantify the operating characteristics of the dose-finding methods, four performance statistics are summarized: the percentage of MTD selection and the percentage of patients treated at the MTD, for which the larger, the better, reflecting the accuracy and efficiency of a design. The remaining two statistics, including the percentage of patients treated at overdoses, and the risk of high toxicity defined as the percentage of trials that lead to the DLT rate greater than ϕ , can quantify the safety aspect of a trial.

The simulation results with different target toxicity rates are respectively presented in Figures 2 to 4. When $\phi = 0.25$, COIN and BOIN perform similarly in terms of selecting the MTD, while COIN allocates more patients to the MTD than BOIN. In addition, it is observed that BOIN is safer than COIN because fewer patients are assigned to overdoses. When $\phi = 0.30$, COIN on average has the best performance among the considered designs. When $\phi = 0.35$, COIN and BOIN perform similarly in regard to correct MTD selection and safety, but COIN still dominates BOIN in the aspect of patient allocation. In all cases, mTPI possesses inferior operating characteristics to both COIN and BOIN in terms of assigning patients to overly toxic doses and the risk of high toxicity. In addition, when the cohort size

is one, mTPI tends to have the smallest correct selection percentage among the three considered designs.

4 Discussion and Concluding Remarks

This paper concerns the relative efficiency of model-assisted designs. In regard to this, a conditional optimal interval design is developed by integrating the observed data from the current and adjacent dose levels. There are several essential contrasts between BOIN and COIN: using information only from the current dose level versus using more information, unchanged versus data-adaptive interval boundaries, and (short-memory) incoherent versus coherent. Extensive numerical studies have been carried out to examine the operating characteristics of COIN and existing model-assisted designs, and the simulation results reveal several key discoveries: First, the efficiency of the model-assisted design depends on the choice of optimal interval. In general, we find that the BOIN design with the default design parameters can perform efficiently in most scenarios. Second, the inclusion of more observed data in the decision-making process might not benefit the MTD identification because COIN and BOIN share particularly similar correct MTD selection percentages according to the simulation study. In summary, we conclude that the BOIN design that is based on local data is sufficient enough for maintaining a desirable performance in most usual phase I dose-finding trials.

For a fixed sample size, (i.e., all trials will not be terminated until the exhaustion of the maximum sample size), the COIN design does not exhibit superiority over the BOIN design in selection of the MTD. Another simulation setting is to fix the maximum sample size and compare the performance of different designs with an early-stopping rule. For example, we fix the maximum sample size to be 36 and consider to terminate the trial when 12 patients are treated at a certain dose level. We conduct another simulation study using this early-stopping setting and the four scenarios examined in Section 3.1. The simulation results of COIN and BOIN based on 10000 replications are provided in Table 3. It can be seen that the COIN design generally costs fewer patients in a trial than the BOIN design. This phenomenon is as expected because COIN has a longer memory than BOIN,²³ then it is more likely to settle down at a certain dose level. In addition, COIN is slightly better than BOIN in terms of the correct selection percentage per patient and the correct allocation percentage.

Although COIN cannot beat BOIN in terms of the MTD identification, COIN is still more efficient in assigning patients to the MTD, as well as it usually induces more sensible dose-assignment decisions due to its coherent property. The proposed COIN design can be seen as a generalization of BOIN by adding more information from the adjacent levels. We do think there is a place where the COIN design can play a role. According to a recent review, the landscape of phase I trials in oncology is changing, and more and more attention is shifting to biomarker-based dose finding.²⁵ When the biomarkers can divide the population into several ordered groups, the idea of COIN can be applied. Liu and Yuan (2015)¹² showed that the asymptotic dose assignment of BOIN would converge to a dose that has a toxicity probability within the optimal toxicity interval. It is also of interest to investigate the limiting behavior of COIN.

Nowadays, phase I/II clinical trials are more and more popular since the toxicity and efficacy endpoints can be accounted for simultaneously in the decision-making process.⁶ It is possible to extend COIN design to incorporate both efficacy and toxicity. For example, the following hypotheses can be constructed for the efficacy probabilities q_{j-1} , q_j , q_{j+1} at dose levels $j-1$, j , and $j+1$, respectively.

$$M_{0j}: q_j > q_{j-1} \text{ and } q_j > q_{j+1} \text{ versus } M_{1j}: q_{j-1} > q_j > q_{j+1} \text{ versus } M_{2j}: q_{j+1} > q_j > q_{j-1},$$

where M_{0j} denotes the current dose level has the maximum efficacy probability among the adjacent doses, and M_{1j} and M_{2j} respectively denotes that dose level $j-1$ and $j+1$ have the maximum efficacy. Based on the observed data, it is possible to calculate the posterior probability of each hypothesis M_{kj} , $k = 0, 1, 2$. However, additional efforts are needed to deal with the tradeoff between efficacy and toxicity. This issue is an appropriate topic for future study.

Acknowledgement

We thank the associate editor, the two referees, and the editor for their many constructive and insightful comments that have led to significant improvements in the article. Yuan's research was partially supported by Award Number P50CA098258 and P50CA217685 from the National Cancer Institute.

Appendix

A Proof of Theorem 1

First, mimicking Liu and Yuan (2015),¹² we can show that the proposed COIN design is equivalent to the following procedure.

- If $y_j < \phi$ and $\Pr(H_{1j}^E | O_j^E) \geq \Pr(H_{0j}^E | O_j^E)$, the dose for the next cohort is escalated to level $j+1$.
- If $y_j \geq \phi$ and $\Pr(H_{2j}^D | O_j^D) \geq \Pr(H_{0j}^D | O_j^D)$, the dose for the next cohort is de-escalated to level $j-1$.
- otherwise, the next cohort is treated at the same dose level j .

Without loss of generality, suppose that a total of n patients have been treated and the current dose level is j , we show the coherence in de-escalation, i.e.,

$P(d_{n+1} = j+1 | t_n = 1, d_n = j, O_j^E) = 0$, where n_j , n_{j+1} and y_{j+1} are updated observations after obtaining the toxicity outcome of the n th patient. The coherence in escalation can be derived in a similar way.

In general, two possible cases would occur:

- i. $n_j = 1$ and $n_{j+1} = 0$, that is, the j th dose level has not been tested before the arrival of the n th patient.
- ii. $n_j > 1$, that is, the dose level j has been tested on previous patients before the arrival of the n th patient.

Under the case (i), it is immediately observed that $\hat{p}_j = 1$, which is strictly greater than $\lambda_{L,j}(O_j^E, \phi)$. As a result, the dose level will not be escalated. Under the case (ii), since the n th patient is treated at the j th dose level, we have

$$\Pr(H_{1j}^E | y_j - t_n, n_j - 1, y_{j+1}, n_{j+1}) < \Pr(H_{0j}^E | y_j - t_n, n_j - 1, y_{j+1}, n_{j+1}),$$

that is, the posterior probability of H_{1j}^E is less than that of H_{0j}^E based on the accumulated information up to the $(n-1)$ th patient. When $t_n = 1$, it can be shown that $\Pr(H_{1j}^E | O_j^E)$ is still less than $\Pr(H_{0j}^E | O_j^E)$, which means $P(d_{n+1} = j+1 | t_n = 1, d_n = j, O_j^E) = 0$.

B Conditional keyboard design

The keyboard design essentially determines which one of the following keys has the largest posterior probability

$$\mathcal{K}_{1j}: p_j \in (\phi - 2\epsilon_1 - \epsilon_2, \phi - \epsilon_1), \mathcal{K}_{0j}: p_j \in (\phi - \epsilon_1, \phi + \epsilon_2), \mathcal{K}_{2j}: p_j \in (\phi + \epsilon_2, \phi + \epsilon_1 + 2\epsilon_2),$$

where ϵ_1, ϵ_2 are two small positive values, $\phi + \epsilon_1 + 2\epsilon_2 < 1$ and $\phi - 2\epsilon_1 - \epsilon_2 > 0$, and the prior $p_j \sim \text{Unif}(0, 1)$. The dose assignment decisions in the keyboard design are then as follows:

- Escalate the dose to level $j+1$ if $1 = \arg\max_{i \in \{0,1,2\}} \{\Pr(\mathcal{K}_{ij} | O_j)\}$;
- De-escalate the dose to level $j-1$ if $2 = \arg\max_{i \in \{0,1,2\}} \{\Pr(\mathcal{K}_{ij} | O_j)\}$;
- Retain the current dose j if $0 = \arg\max_{i \in \{0,1,2\}} \{\Pr(\mathcal{K}_{ij} | O_j)\}$.

For the conditional keyboard design, we consider the following pairwise comparison for dose escalation,

$$\mathcal{K}_{0j}^E: p_j \in (\phi - \epsilon_1, \phi + \epsilon_2), p_{j+1} \in (\phi + \epsilon_2, \phi + \epsilon_1 + 2\epsilon_2);$$

$$\mathcal{K}_{1j}^E: p_j \in (\phi - 2\epsilon_1 - \epsilon_2, \phi - \epsilon_1), p_{j+1} \in (\phi - \epsilon_1, \phi + \epsilon_2),$$

and the following pairwise comparison for dose de-escalation,

$$\mathcal{K}_{0j}^D: p_j \in (\phi - \epsilon_1, \phi + \epsilon_2), p_{j-1} \in (\phi - 2\epsilon_1 - \epsilon_2, \phi - \epsilon_1);$$

$$\mathcal{K}_{2j}^D: p_j \in (\phi + \epsilon_2, \phi + \epsilon_1 + 2\epsilon_2), p_{j-1} \in (\phi - \epsilon_1, \phi + \epsilon_2).$$

Then the decision rules of the conditional keyboard design can be summarized as follows.

- Escalate the dose to level $j+1$ if $\Pr(\mathcal{K}_{1j}^E | O_j^E) \geq \Pr(\mathcal{K}_{0j}^E | O_j^E)$;
- De-escalate the dose to level $j-1$ if $\Pr(\mathcal{K}_{2j}^D | O_j^D) \geq \Pr(\mathcal{K}_{0j}^D | O_j^D)$;
- Retain the current dose j otherwise.

References

1. Ting N Dose Finding in Drug Development. Springer: New York, 2006.
2. Storer BE. Design and analysis of phase I clinical trials. Biometrics 1989; 45, 925–937. [PubMed: 2790129]
3. Ahn C An evaluation of phase I cancer clinical trial designs. Stat Med 1998; 17: 1537–1549. [PubMed: 9699228]
4. O’Quigley J, Pepe M, and Fisher L. Continual reassessment method: a practical design for phase I clinical trials in cancer. Biometrics 1990; 46, 33–48. [PubMed: 2350571]
5. Liu S, Cai C and Ning J. Up-and-down designs for phase I clinical trials. Contemp Clin Trials 2013; 36, 218–227. [PubMed: 23856381]
6. Yuan Y, Nguyen HQ and Thall PF. Bayesian Designs for Phase I–II Clinical Trials. Chapman & Hall/CRC: New York, 2016.
7. Yan F, Mandrekar S, and Yuan Y. Keyboard: A Novel Bayesian Toxicity Probability Interval Design for Phase I Clinical Trials. Clin Cancer Res 2017; 23, 3994–4003. [PubMed: 28546227]
8. Zhou H, Murray T, Pan H and Yuan Y. Comparative review of model-assisted designs for phase I clinical trials. Stat Med 2018; 37, 2208–2222. [PubMed: 29682777]
9. Zhou H, Yuan Y and Nie L. Accuracy, safety and reliability of novel Bayesian phase I trial designs. Clin Cancer Res 2018; 24, 4357–4364. [PubMed: 29661774]
10. Ivanova A, Flournoy N and Chung Y. Cumulative cohort design for dose finding. J Stat Plan Inference 2007; 137, 2316–2327.
11. Ji Y, Liu P, Li Y and Bekele BN. A modified toxicity probability interval method for dose-finding trials. Clin Trials 2010; 7, 653–663. [PubMed: 20935021]
12. Liu S and Yuan Y. Bayesian optimal interval designs for phase I clinical trials. J R Stat Soc Ser C 2015; 64, 507–523.
13. Mu R, Yuan Y, Xu J, Mandrekar SJ, Yin JY. gboin: A unified model-assisted phase I trial design accounting for toxicity grades, binary or continuous endpoint. J R Stat Soc Ser C Appl Stat 2018; 10.1111/rssc.12263
14. Yuan Y, Lin R, Li D, Nie L, and Warren KE Time-to-event Bayesian Optimal Interval Design to Accelerate Phase I Trials. Clin Cancer Res 2018; 24, 4921–4930. [PubMed: 29769209]
15. Lin R Bayesian optimal interval design with multiple toxicity constraints. Biometrics 2018; doi: 10.1111/biom.12912. [Epub ahead of print]
16. Lin R and Yin G. Bayesian optimal interval design for dose finding in drug-combination trials. Stat Methods Med Res 2017; 26, 2155–2167. [PubMed: 26178591]
17. Zhang L and Yuan Y. A practical Bayesian design to identify the maximum tolerated dose contour for drug combination trials. Stat Med 2016; 35, 4924–4936. [PubMed: 27580928]
18. Pan H, Lin R and Yuan Y. Statistical properties of the keyboard design with extension to drug-combination trials. 2018, under revision.
19. Neuenschwander B, Branson M, Gsponer T. Critical aspects of the Bayesian approach to phase I cancer trials. Stat Med 2008; 27, 2420–2439. [PubMed: 18344187]

20. Yuan Y, Hess KR, Hilsenbeck SG, and Gilbert MR. Bayesian optimal interval design: a simple and well-performing design for phase I oncology trials. *Clin Cancer Res* 2016; 22, 4291–4301. [PubMed: 27407096]
21. Bril G, Dykstra R, Pillers C, and Robertson T. Algorithm AS 206: isotonic regression in two independent variables. *J R Stat Soc Ser C* 1984; 33, 352–357.
22. Cheung YK. Coherence principles in dose-finding studies. *Biometrika* 2005; 92, 863–873.
23. Oron AP and Hoff PD. Small-sample behavior of novel phase I cancer trial designs. *Clin Trials* 2013; 10, 63–80. [PubMed: 23345304]
24. Paoletti X, O’Quigley J, and Maccario J. Design efficiency in dose finding studies. *Comput Stat Data Anal* 2004; 45, 197–214.
25. Wong KM, Capasso A and Eckhardt SG. The changing landscape of phase I trials in oncology. *Nat Rev Clin Oncol* 2016; 13, 106–117. [PubMed: 26552953]

Figure 1:

Comparison of the interval boundaries generated by COIN and BOIN. We assume the current dose is level j . n_j and n_{j+1} are the sample sizes at dose levels j and $j+1$, respectively, and y_{j+1} is the number of DLTs observed at dose level $j+1$. The dashed line corresponds to the BOIN design, while the solid line denotes the COIN design. The dotted line indicates the target toxicity rate of 0.25. This figure shows that the interval boundaries of BOIN are invariant to the data from adjacent dose levels, while those of COIN depends on the data from adjacent levels as well as the sample size at the current dose level.

Figure 2:

Simulation results based on 50,000 randomly generated dose–toxicity scenarios under different numbers of dose levels and cohort sizes with the target toxicity rate of 0.25. The symbols “*” and “+” respectively correspond to cases with average probability difference around the target being 0.15 and 0.10.

Figure 3: Simulation results based on 50,000 randomly generated dose–toxicity scenarios under different numbers of dose levels and cohort sizes with the target toxicity rate of 0.30. The symbols “*” and “+” respectively correspond to cases with average probability difference around the target being 0.15 and 0.10.

Figure 4: Simulation results based on 50,000 randomly generated dose-toxicity scenarios under different numbers of dose levels and cohort sizes with the target toxicity rate of 0.35. The symbols “*” and “+” respectively correspond to cases with average probability difference around the target being 0.15 and 0.10.

Table 1:

Operating characteristics of the BOIN and COIN designs under four prespecified dose-toxicity scenarios with a total sample size of 36. The target toxicity rate of 0.25 is bolded. The design parameters of these two designs are set as $\phi_1 = (1-\lambda)\phi$ and $\phi_2 = (1+\lambda)\phi$.

Design			Cohort size = 1						Cohort size = 3					
			1	2	3	4	5	6	1	2	3	4	5	6
Scenario 1		Pr(toxicity)	0.25	0.35	0.50	0.60	0.70	0.80	0.25	0.35	0.50	0.60	0.70	0.80
	$\lambda = 0.4$	BOIN	70.0	27.1	2.7	0.2	0.0	0.0	76.1	22.2	1.6	0.1	0.0	0.0
		Patients (#)	20.8	10.1	3.7	1.0	0.3	0.1	24.9	8.9	2.0	0.2	0.0	0.0
		COIN	68.6	28.5	2.6	0.2	0.0	0.0	70.5	27.1	2.2	0.1	0.0	0.0
		Selection (%)	20.8	10.8	3.2	0.9	0.3	0.1	23.1	10.3	2.3	0.3	0.0	0.0
		Patients (#)	67.5	28.9	3.1	0.4	0.0	0.0	72.1	25.4	2.3	0.2	0.0	0.0
$\lambda = 0.2$		BOIN	19.1	10.8	4.5	1.2	0.3	0.1	22.5	10.3	2.9	0.3	0.0	0.0
		Selection (%)	69.6	27.6	2.6	0.2	0.0	0.0	70.7	26.7	2.4	0.2	0.0	0.0
		Patients (#)	20.8	10.6	3.3	0.9	0.3	0.1	22.5	10.5	2.6	0.3	0.0	0.0
		COIN	0.10	0.25	0.40	0.60	0.70	0.80	0.10	0.25	0.40	0.60	0.70	0.80
		Pr(toxicity)	12.9	67.1	19.3	0.6	0.1	0.0	14.2	69.5	15.7	0.6	0.0	0.0
	$\lambda = 0.4$	BOIN	8.7	15.5	8.9	2.4	0.5	0.1	11.8	16.3	6.7	1.1	0.1	0.0
		Patients (#)	11.8	68.3	19.2	0.6	0.1	0.0	12.3	67.7	19.2	0.7	0.0	0.0
		COIN	6.9	17.8	8.9	1.9	0.4	0.1	8.4	18.2	8.1	1.2	0.1	0.0
		Selection (%)	12.1	66.6	20.1	1.0	0.2	0.0	13.7	67.6	17.9	0.7	0.0	0.0
$\lambda = 0.2$		BOIN	8.4	14.4	9.6	3.0	0.6	0.1	11.1	15.4	7.9	1.6	0.1	0.0
		Patients (#)	12.4	68.8	17.9	0.7	0.1	0.0	13.0	66.9	19.3	0.8	0.0	0.0
		COIN	7.3	17.5	8.7	1.9	0.5	0.1	8.5	17.7	8.4	1.3	0.1	0.0
		Patients (#)	0.05	0.10	0.25	0.32	0.50	0.60	0.05	0.10	0.25	0.32	0.50	0.60
		Pr(toxicity)	0.2	12.5	51.2	32.0	3.8	0.3	0.1	16.8	54.3	25.5	3.2	0.1
	$\lambda = 0.4$	BOIN	2.2	7.4	12.3	9.0	4.0	1.0	5.1	10.5	12.4	6.1	1.8	0.2
		Patients (#)	0.2	11.3	50.0	34.4	3.9	0.2	0.1	12.8	50.4	31.7	4.9	0.2
		COIN	1.8	6.1	13.3	10.4	3.5	0.9	3.8	7.7	13.8	8.1	2.3	0.3
		Selection (%)	0.2	11.7	49.7	33.5	4.7	0.3	0.1	15.7	50.3	29.4	4.2	0.3
$\lambda = 0.2$		BOIN	2.2	7.1	11.4	9.3	4.7	1.2	4.9	9.9	11.7	7.0	2.3	0.3
		Patients (#)												

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Design			Cohort size = 1						Cohort size = 3					
			1	2	3	4	5	6	1	2	3	4	5	6
	COIN	Selection (%)	0.2	11.8	50.1	33.5	4.1	0.3	0.1	13.4	49.6	31.5	5.0	0.4
		Patients (#)	1.8	6.3	13.1	10.3	3.6	0.9	3.8	7.7	13.3	8.3	2.6	0.3
Scenario 4		Pr(toxicity)	0.01	0.02	0.03	0.04	0.05	0.25	0.01	0.02	0.03	0.04	0.05	0.25
	$\lambda = 0.4$	BOIN	0.0	0.0	0.0	0.0	7.7	92.2	0.0	0.0	0.1	0.7	15.2	84.1
		Patients (#)	1.1	1.2	1.3	1.6	7.8	22.9	3.3	3.5	3.8	4.0	7.7	13.7
	COIN	Selection (%)	0.0	0.0	0.0	0.0	7.3	92.7	0.0	0.0	0.0	0.1	11.7	88.2
		Patients (#)	1.1	1.2	1.3	1.4	5.6	25.4	3.1	3.2	3.4	3.5	5.4	17.4
$\lambda = 0.2$	BOIN	Selection (%)	0.0	0.0	0.0	0.1	7.9	92.0	0.0	0.0	0.1	0.3	14.8	84.8
		Patients (#)	1.1	1.2	1.3	1.6	8.4	22.4	3.3	3.5	3.7	4.0	7.7	13.8
	COIN	Selection (%)	0.0	0.0	0.0	0.0	7.8	92.2	0.0	0.0	0.0	0.0	12.7	87.3
		Patients (#)	1.1	1.2	1.2	1.4	6.2	24.9	3.1	3.2	3.3	3.5	5.9	17.0

Table 2:

Operating characteristics of the BOIN and COIN designs under four prespecified dose-toxicity scenarios with a total sample size of 90. The target toxicity rate of 0.25 is bolded. The design parameters of these two designs are set as $\phi_1 = (1-\lambda)\phi$ and $\phi_2 = (1+\lambda)\phi$.

Design			Cohort size = 1						Cohort size = 3					
			1	2	3	4	5	6	1	2	3	4	5	6
Scenario 1		Pr(toxicity)	0.25	0.35	0.50	0.60	0.70	0.80	0.25	0.35	0.50	0.60	0.70	0.80
	$\lambda = 0.4$	BOIN	83.3	16.6	0.1	0.0	0.0	0.0	84.9	15.0	0.1	0.0	0.0	0.0
		Patients (#)	60.9	22.9	4.9	1.0	0.3	0.1	66.1	20.6	3.0	0.3	0.0	0.0
	COIN	Selection (%)	82.4	17.5	0.1	0.0	0.0	0.0	82.6	17.4	0.1	0.0	0.0	0.0
		Patients (#)	62.3	22.9	3.6	0.9	0.2	0.1	64.7	22.2	2.8	0.3	0.0	0.0
$\lambda = 0.2$	BOIN	Selection (%)	81.3	18.5	0.2	0.0	0.0	0.0	81.8	18.0	0.2	0.0	0.0	0.0
		Patients (#)	55.9	25.7	6.7	1.3	0.3	0.1	59.3	25.1	5.2	0.4	0.0	0.0
	COIN	Selection (%)	82.8	17.0	0.1	0.0	0.0	0.0	83.3	16.6	0.1	0.0	0.0	0.0
		Patients (#)	62.0	22.8	3.9	1.0	0.3	0.1	63.5	22.8	3.3	0.4	0.0	0.0
Scenario 2		Pr(toxicity)	0.10	0.25	0.40	0.60	0.70	0.80	0.10	0.25	0.40	0.60	0.70	0.80
$\lambda = 0.4$	BOIN	Selection (%)	3.7	88.6	7.7	0.1	0.0	0.0	3.8	88.4	7.8	0.0	0.0	0.0
		Patients (#)	17.5	50.1	19.0	2.8	0.5	0.1	20.1	51.2	16.9	1.6	0.1	0.0
	COIN	Selection (%)	3.6	89.0	7.3	0.0	0.0	0.0	3.6	88.8	7.6	0.0	0.0	0.0
		Patients (#)	11.3	60.3	15.9	1.9	0.4	0.1	12.2	60.4	15.9	1.3	0.1	0.0
$\lambda = 0.2$	BOIN	Selection (%)	3.8	88.0	8.1	0.1	0.0	0.0	3.9	87.7	8.3	0.1	0.0	0.0
		Patients (#)	19.3	43.8	22.5	3.7	0.6	0.1	21.7	44.3	21.2	2.7	0.1	0.0
	COIN	Selection (%)	4.3	88.7	6.9	0.0	0.0	0.0	3.7	89.3	7.0	0.1	0.0	0.0
		Patients (#)	13.0	58.0	16.4	2.0	0.5	0.1	13.6	58.0	16.6	1.6	0.1	0.0
Scenario 3		Pr(toxicity)	0.05	0.10	0.25	0.32	0.50	0.60	0.05	0.10	0.25	0.32	0.50	0.60
$\lambda = 0.4$	BOIN	Selection (%)	0.0	4.2	69.5	25.9	0.4	0.0	0.0	4.1	71.7	23.7	0.4	0.0
		Patients (#)	2.5	15.7	40.5	24.1	6.2	1.1	5.1	18.9	41.9	19.8	4.0	0.3
	COIN	Selection (%)	0.0	3.7	68.7	27.3	0.3	0.0	0.0	3.6	68.1	28.0	0.3	0.0
		Patients (#)	1.9	10.2	45.9	26.7	4.3	0.9	3.9	11.5	46.3	24.7	3.4	0.3
$\lambda = 0.2$	BOIN	Selection (%)	0.0	4.2	68.0	27.1	0.6	0.0	0.0	4.1	68.9	26.3	0.7	0.0
		Patients (#)	2.5	17.1	35.5	25.0	8.5	1.4	5.1	19.7	36.0	22.3	6.4	0.5

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Design			Cohort size = 1						Cohort size = 3					
			1	2	3	4	5	6	1	2	3	4	5	6
	COIN	Selection (%)	0.0	4.0	69.9	25.7	0.3	0.0	0.0	3.8	68.9	26.8	0.5	0.0
		Patients (#)	1.9	11.7	44.8	26.0	4.6	1.0	3.8	12.5	44.4	24.8	4.1	0.4
Scenario 4		Pr(toxicity)	0.01	0.02	0.03	0.04	0.05	0.25	0.01	0.02	0.03	0.04	0.05	0.25
	$\lambda = 0.4$	BOIN	0.0	0.0	0.0	0.0	0.6	99.4	0.0	0.0	0.0	0.0	1.1	98.9
		Patients (#)	1.1	1.2	1.3	1.7	16.3	68.4	3.3	3.5	3.8	4.0	17.0	58.4
	COIN	Selection (%)	0.0	0.0	0.0	0.0	0.6	99.4	0.0	0.0	0.0	0.0	0.7	99.3
		Patients (#)	1.1	1.2	1.3	1.5	8.1	76.9	3.1	3.2	3.4	3.5	8.5	68.3
$\lambda = 0.2$	BOIN	Selection (%)	0.0	0.0	0.0	0.0	0.6	99.4	0.0	0.0	0.0	0.0	0.9	99.1
		Patients (#)	1.1	1.2	1.3	1.6	20.2	64.6	3.3	3.5	3.7	4.0	19.8	55.7
	COIN	Selection (%)	0.0	0.0	0.0	0.0	0.6	99.4	0.0	0.0	0.0	0.0	0.9	99.1
		Patients (#)	1.1	1.2	1.2	1.4	10.2	74.8	3.1	3.2	3.3	3.5	10.2	66.7

Table 3:

Simulation results of the BOIN and COIN designs under the four scenarios in Table 0 when an additional early termination rule (i.e., stopping the trial when 12 patients are treated at a certain dose level) is imposed for both designs.

Scenarios		Cohort size = 1				Cohort size = 3			
		1	2	3	4	1	2	3	4
Total patients (#)	BOIN	20.7	24.6	26.6	22.7	18.4	24.5	29.0	32.9
	COIN	18.4	20.9	22.8	20.6	17.3	21.4	25.2	29.6
Correct selection (%)	BOIN	68.6	58.7	45.5	87.5	77.2	59.9	49.6	84.0
	COIN	63.3	56.6	42.1	88.3	68.0	57.2	45.6	88.0
Correct selection per patient (%)	BOIN	3.3	2.4	1.7	3.9	4.2	2.4	1.7	2.6
	COIN	3.4	2.7	1.8	4.3	3.9	2.7	1.8	3.0
Patients at MTD (#)	BOIN	9.6	9.5	8.1	11.6	11.0	10.0	8.2	10.8
	COIN	8.8	8.6	7.1	11.4	9.7	9.1	8.2	11.4
Correct allocation (%)	BOIN	46.4	38.6	30.5	51.1	59.8	40.8	28.3	32.8
	COIN	47.8	41.1	31.1	55.3	56.1	42.5	32.5	38.5

Note: $\phi_1 = 0.15$ and $\phi_2 = 0.35$ for both designs. The maximum sample size is 36. Correct selection per patient (%) = Correct selection (%) / Total patients (#); Correct allocation (%) = $100 \times$ Patients at MTD (#) / Total patients (#)