Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: www.tandfonline.com/journals/uasa20

Posterior Predictive Design for Phase I Clinical Trials

Chenqi Fu, Shouhao Zhou & J. Jack Lee

To cite this article: Chenqi Fu, Shouhao Zhou & J. Jack Lee (2025) Posterior Predictive Design for Phase I Clinical Trials, Journal of the American Statistical Association, 120:552, 2646-2657, DOI: 10.1080/01621459.2025.2484044

To link to this article: https://doi.org/10.1080/01621459.2025.2484044

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2025, VOL. 120, NO. 552, 2646–2657: Theory and Methods
https://doi.org/10.1080/01621459.2025.2484044

Posterior Predictive Design for Phase I Clinical Trials

Chenqi Fu[a], Shouhao Zhou[a], and J. Jack Lee[b]

[a]Division of Biostatistics and Bioinformatics, Department of Public Health Sciences, Pennsylvania State University, Hershey, PA; [b]Department of Biostatistics, The University of Texas - MD Anderson Cancer Center, Houston, TX

ABSTRACT

Interval-based designs represent cutting-edge adaptive methodologies for phase I clinical trials to identify the maximum tolerated dose (MTD). These designs exhibit robust performance comparable to more intricate, model-based designs, and their pretabulated decision rule enables them to be implemented as simply as the conventional algorithm-based designs. In this paper, we introduce the posterior predictive (PoP) design, a novel interval-based design that leverages advanced Bayesian predictive hypothesis testing techniques for dose escalation and de-escalation. Our work moves beyond the existing model-assisted interval-based designs by achieving global optimality in dose transition. Theoretically, the global optimality ensures that the proposed design can consistently select the true MTD at an impressive convergence rate of $n^{-1/2}$. Through extensive simulation studies, we demonstrate that the PoP design yields substantial improvement in operating characteristics to identify MTD, thereby presenting a valuable upgrade to the popular interval-based designs in practice. Supplementary materials for this article are available online, including a standardized description of the materials available for reproducing the work.

ARTICLE HISTORY
Received October 2023
Accepted March 2025

KEYWORDS
BOIN design; Interval-based designs; Model-assisted design; Predictive Bayes factor; PoP design

1. Introduction

In this work we are interested in the adaptive design of phase I clinical trials to identify the maximum tolerated dose (MTD), which is the highest dose of a treatment that can be safely given. Incorrect determination of MTD can result in either exposing patients to an overly toxic dose or a sub-therapeutic dose. Thus, accurately identifying MTD and treating more patients at the MTD level are key objectives for dose-finding studies to ensure the success of subsequent phase II and III clinical trials.

Popular dose-finding methods can be classified into three major categories: algorithm-based design, model-based designs, and model-assisted designs (Jaki, Clive, and Weir 2013; Iasonos and O'Quigley 2014; van Brummelen et al. 2016; Yuan, Lin, and Lee 2019). Algorithm-based designs use simple rules to guide dose escalation or de-escalation in practice, such as with the "3+3" design, which are easy to implement but have low accuracy in identifying MTD and a lack of flexibility to adjust when encountering variations in real-world situations (Le Tourneau, Lee, and Siu 2009). Addressing more specific dose-finding problems, the model-based methods, such as the continual reassessment method (CRM) (O'Quigley, Pepe, and Fisher 1990) and its extensions (Faries 1994; Braun 2002; Leung and Wang 2002; Yin and Yuan 2009), build upon a parametric model for the dose-toxicity relationship and adaptively evaluate the model based on the observed data. The model-based designs provide more efficient performance of MTD selection (van Brummelen et al. 2016), but they require special software and are challenging to implement in clinical settings.

Model-assisted designs establish a novel model-free paradigm for the easy implementation of phase I clinical trials (Yuan, Lin, and Lee 2022). They benefit from both the simple forms of algorithm-based designs and the superior operating characteristics of model-based designs (Zhou et al. 2018; Yuan, Lin, and Lee 2019). However, model-assisted designs are indeed model-free, and the patient dose assignment is implemented using predetermined intervals to minimize incorrect decisions of dose escalation or de-escalation. Therefore, we refer to the model-assisted designs as interval-based designs (Oron, Azriel, and Hoff 2011) to avoid the misconception of using any specific dose-toxicity model. Some examples of interval-based designs include the modified toxicity probability interval (mTPI) design (Ji and Wang 2013; Guo et al. 2017), the Bayesian optimal interval (BOIN) design (Liu and Yuan 2015), the Keyboard design (Yan, Mandrekar, and Yuan 2017) and their variants and extensions. Interval-based designs can tabulate the decision rules in advance. During the trial, the decision of dose assignment for the next cohort of patients involves only a simple comparison of the observed toxicity rates with a prespecified toxicity tolerance interval. This makes it easy for the clinical team to conduct the trial. Meanwhile, the BOIN and Keyboard designs have demonstrated robust operating characteristics comparable to the CRM design (Zhou et al. 2018; Zhu, Hwang, and Li 2019), highlighting the accuracy, efficiency, and safety of interval-based designs.

CONTACT Shouhao Zhou szhou1@pennstatehealth.psu.edu Division of Biostatistics and Bioinformatics, Department of Public Health Sciences, Pennsylvania State University, Hershey, PA 17033.

Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

Interval-based designs have gained remarkable popularity in dose-finding studies, but they also raise concerns due to the arbitrary choice of the toxicity tolerance interval. The "optimality" property of these designs, which minimizes the risk of incorrect decisions, theoretically holds with respect to a predefined tolerance interval. In reality, the efficiency of these designs is inherently dependent on, and limited by, the choice of this interval. A poorly chosen interval can lead to suboptimal performance, potentially worse than that of conventional algorithm-based designs.

In practice, the tolerance intervals are often pre-specified for each design method and calibrated through simulation to yield reasonable empirical performance. However, there is no guarantee of robust performance across all possible target toxicity rates. The optimization over the tolerance intervals, or equivalently the optimal decision rules, remains largely unknown. Therefore, the theoretical establishment of an optimal tolerance interval for the widely used interval-based designs remains an important open question.

It is also worth noting that even with a large sample size, the popular interval-based designs with well-calibrated tolerance intervals can suffer from another significant limitation. Specifically, these designs may select an arbitrary dose within the tolerance interval, rather than converge to the true MTD (Liu and Yuan 2015; Pan et al. 2020). Shortening the toxicity tolerance interval to exclude undesirable doses will not solve this issue because this may result in overly frequent dose transitions and lead to a loss of efficiency. Therefore, it is critical to develop new interval-based designs that can more accurately and efficiently identify the MTD.

The optimal choice of adaptive decisions for dose transition to minimize wrong decisions in previously published interval-based designs is contingent upon the "local" choice of tolerance interval, which we refer to as the "local optimality". This is in contrast to the "global" optimality in which the optimality is achieved without the specification of a tolerance interval in the dose-finding algorithm. Currently, only local optimality can be claimed in popular interval-based designs such as mTPI, BOIN, Keyboard, and their many variants. However, as we will demonstrate in the Methods section, the specification of a local tolerance interval is arbitrary and unnecessary. If it can be avoided, it will enable the improvement of convergence in probability 1 to identify MTD to reach global optimality.

In this article, we propose a new interval-based approach, the posterior predictive (PoP) design, established on an improved test statistic for decision-making. Our approach applies the predictive Bayes factor (Zhou 2011) as a hypothesis testing tool to determine the optimal action of dose allocation, that is, whether to retain or transit, by regarding each as a candidate hypothesis. In this Bayesian decision framework, the dose escalation or de-escalation boundaries are globally determined by minimizing the risk of incorrect actions. Without the specification of local tolerance interval, the PoP design can reduce the efficiency loss caused by testing of interval hypotheses and converge to the true MTD at a rate of $n^{-1/2}$. The decision rule of the PoP design can be pretabulated, making it as convenient as other interval-based designs.

The remainder of this article is organized as follows. In Section 2, we review the predictive Bayes factor and formulate the dose-finding problem as a model selection problem. Then, we propose the PoP design for dose assignments. In Section 3, we investigate the theoretical properties of the PoP design. In Section 4, we discuss practical implications. In Section 5, we perform simulations to evaluate and compare the operating characteristics of the PoP design. We conclude with a discussion in Section 6.

2. Methods

2.1. Global Hypothesis Testing for Dose Assignment

We assume that there are J prespecified dose levels of the drug of interest. Let $d_1, d_2, \ldots, d_J$ denote these dose levels. The dose-limiting toxicity (DLT) is assessed as a binary outcome for each individual patient/subject, defined by the occurrence of some severe toxicity. The true dose toxicity is assumed monotonically increasing as the dose level increases. Let ϕ be the target toxicity rate, and π_j be the true dose-toxicity of dose level d_j, for $j = 1, 2, \ldots, J$.

To guide the dose finding, we formulate the following *global* hypothesis testing for a given dose level d_j:

$$\begin{aligned} H_{0j} &: \pi_j = \phi \\ H_{1j} &: \pi_j \neq \phi, \end{aligned} \tag{1}$$

where H_{0j} indicates that d_j is MTD and should be retained; H_{1j} reflects the current dose is either below or above MTD, so that we should transit to a different dose level. Thus, the dose assignment decision is well aligned with hypothesis testing of H_{0j} against H_{1j}. When H_{1j} is favored, the direction of dose adjustment is determined by the observed toxicity rate: we either de-escalate the dose if the observed toxicity rate is above the target toxicity rate ϕ, or escalate the dose if the observed toxicity rate is below ϕ.

As patients are enrolled sequentially, the dose assignment problem can be viewed as a series of decisions to update which hypothesis would be favored. However, because the true toxicity rate π_j is unknown, these decisions may not always be accurate. As a result, it is essential to minimize the risk of incorrect decisions to optimize the assignment of patients to MTD. Later in Section 3.1, we will establish the *global optimality* property to minimize the risk of incorrect decisions with respect to the global hypothesis testing in (1), which differs fundamentally from *local optimality* in existing interval-based designs.

2.2. Posterior Predictive (PoP) Design

The Bayes factor is a widely used approach for Bayesian hypothesis testing. However, it may raise issues in conducting the hypothesis testing described in (1) in Section 2.1 (Kass and Raftery 1995). Specifically, even for data sets extremely implausible under the point hypothesis H_{0j}, the Bayes factor may yield confusing results in favor of H_{0j}; this is known as Lindley's paradox (Lindley 1957). It cannot be resolved by increased sample size, as evidence in favor of the point hypothesis H_{0j} may become overwhelming as the sample size increases (Bernardo 1999).

In the dose-finding setting, the straightforward employment of Bayes factors could mistakenly lead to incorrect decisions

of retaining the current dose even if the updated toxicity data suggest a high DLT rate. This issue is particularly concerning, as it can expose patients to higher levels of toxicity than necessary (Berry et al. 2010). To avoid this problem, popular interval-based designs employ *local* hypothesis testing rather than directly testing (1). For example, the prominent BOIN design (Liu and Yuan 2015) adopts

$$\begin{aligned} H_{0j} &: \pi_j = \phi, \\ H^*_{1j} &: \pi_j = \phi_1, \\ H^*_{2j} &: \pi_j = \phi_2, \end{aligned} \tag{2}$$

by specifying two single points ϕ_1 and ϕ_2 ($\phi_1 < \phi < \phi_2$), as the empirical local alternative hypotheses to guide dose escalation and de-escalation, respectively. When the local hypothesis testing strategy takes a detour to replace the global hypothesis testing in (1) in Section 2.1 and solve the target hypothesis testing problem, it is also constrained by the specific choices of ϕ_1 and ϕ_2 for locally guided decision rules in risk minimization of incorrect decisions (i.e., local optimization), potentially diminishing the dose-finding efficiency.

In contrast, the predictive Bayes factor (Zhou 2011) provides an alternative solution to conduct the hypothesis testing problem (1) (see also Appendix A). In the context of dose-finding, we apply the predictive Bayes factor for the problem to determine the dose assignment of the next cohort of subjects. Let $y_j = \sum_{i=1}^{n_j} y_{j,i}$ denote the sum of DLTs among n_j subjects who received dose d_j, for $j = 1, 2, \ldots, J$, following a binomial distribution

$$y_j \sim \text{Bin}(n_j, \pi_j).$$

Under H_1, we assume a non-informative prior $\pi_j \sim \text{Unif}(0, 1)$. Then, the predictive Bayes factor comparing H_{0j} and H_{1j} is given by

$$\text{PrBF}_{0,1} = e\,(n_j + 2)^{n_j} \left(\frac{\phi}{y_j + 1}\right)^{y_j} \left(\frac{1-\phi}{n_j - y_j + 1}\right)^{n_j - y_j}, \tag{3}$$

where e is the Euler number. The smaller the value of $\text{PrBF}_{0,1}$, the stronger the evidence favors H_{1j} for dose transition against H_{0j} for dose retention.

The evidence provided by $\text{PrBF}_{0,1}$ can be interpreted in terms of the observed DLT counts y_j. As a function of y_j, $\text{PrBF}_{0,1}$ in (3) is a strict bell-shaped function, with the tails on both ends favoring H_{1j} for the dose transition. A small $y_j < \phi \cdot n_j$ indicates low toxicity and suggests dose escalation, while a large $y_j > \phi \cdot n_j$ suggests high toxicity and dose de-escalation. Given the observed y_j, there is no ambiguity in deciding whether to escalate or de-escalate when H_{1j} is favored. Therefore, the two-sided interval hypothesis H_{1j} is sufficient for statistical evaluation to guide the dose-finding.

Dose transition. Following the discussion above, we have the following decision rule for dose transition:

- If $\text{PrBF}_{0,1} < C$, the evidence is in favor of H_{1j}. We assign the next cohort of patients to an adjacent dose according to y_j, such as
  - If $y_j < \phi \cdot n_j$, we escalate the dose;
  - If $y_j > \phi \cdot n_j$, we de-escalate the dose.
- Otherwise, we retain the current dose.

The threshold C describes the tolerance in strength of evidence for dose transition. Mathematically, it can be intrinsically determined by specifying a loss function of making incorrect decisions. More technical details will be provided in Section 3.1, which justifies the theoretical foundation for the threshold C.

Dose exclusion. To improve patient safety and trial efficiency, the PoP design also employs a dose exclusion rule in patient allocation. If the predictive Bayes factor based on the observed y_j indicates a dose is above MTD with certain evidence, we exclude the current dose and the doses above it to avoid treating patients at overly toxic doses. Alternatively, if the predictive Bayes factor implies that a dose is substantially below MTD, we exclude the current dose and doses below it to prevent patients from receiving subtherapeutic doses. We present a dose exclusion rule as follows:

- If $\text{PrBF}_{0,1} < E$, the evidence is *strongly* in favor of H_{1j} and:
  - If $y_j < \phi \cdot n_j$, the current dose is deemed as subtherapeutic and we exclude the current dose and lower doses.
  - If $y_j > \phi \cdot n_j$, the current dose is overly toxic and we exclude the current dose and higher doses.

When all the doses are excluded from further investigation, the trial is terminated early. The exclusion threshold E ensures the safety of the patients and efficiency of the design by influencing the early termination rule. See also Section 3.1 for technical details on how to determine E.

MTD selection. After the patient enrollment completes and outcomes are observed, a dose level needs to be selected as MTD. Following the BOIN and keyboard designs (Liu and Yuan 2015; Yan, Mandrekar, and Yuan 2017), the PoP design applies the isotonic regression to the observed DLT rates to ensure they are monotonically no-decreasing and selects MTD based on the isotonic estimates (Brunk et al. 1972). In addition, we consider a dose j to be overly toxic, if

$$P(\pi_j > \phi | y_j, n_j) > 0.95 \text{ and } n_j \geq 3,$$

then we eliminate the dose j and higher doses from the MTD selection process. This dose elimination rule is the same as in the BOIN design. If the lowest dose is eliminated from the MTD selection process, no dose level will be selected as MTD.

3. Statistical Properties

As the first step in evaluating a novel drug in humans, phase I clinical trials require a conscientious and statistically rigorous design. An effective phase I trial should strike a balance of safety, efficiency, and ethical conduct (Le Tourneau, Lee, and Siu 2009). In this section, we present the theoretical properties of the PoP design.

3.1. Global Optimality

We first provide a justification of optimality for the thresholds C and E under the Bayesian decision framework. The objective is to minimize the risk of incorrect decisions for global hypotheses (1). At a current dose level d_j, we have three potential actions: retain the current dose, transit but not exclude the current dose, and transit and exclude the current dose, denoted as $\mathcal{R}$, $\mathcal{T}$, and $\mathcal{E}$, respectively. Here $\mathcal{R}$ is considered the correct decision under "$H_{0j}: d_j$ is MTD". $\mathcal{T}$ and $\mathcal{E}$ are considered as the correct decisions under "$H_{1j}: d_j$ is not MTD", where $\mathcal{E}$ is superior to $\mathcal{T}$. Because the decision rule is identical for various dose levels, hereafter we suppress the subscript j in the notation for theorems and proofs when there is no ambiguity.

	$\mathcal{R}$	$\mathcal{T}$	$\mathcal{E}$
$H_0: \pi_j = \phi$	0	b_1	1
$H_1: \pi_j \neq \phi$	b_2	b_3	0

To assess the risk of incorrect decisions, we define the loss function $L(a(y), H_k)$ for action $a(y)$, and hypothesis H_k with $k = 0, 1$, and loss scores in the minitable above $(b_1, b_2, b_3) \in \{0 < b_1 < b_2(1-b_3), 0 < b_2 < 1, 0 < b_3 < 0.5b_2\}$. We set the upper bound of b_3 to $0.5b_2$ rather than b_2. This is because, in the case of an incorrect dose assignment H_1, the action $\mathcal{T}$ that involves transitioning but not excluding the dose may not be ideal, but is still viable as a suboptimal decision. Thus, the predictive Bayes risk $R(a)$ is given by

$$\begin{aligned} R(a) =& P(H_0(\boldsymbol{D})) \sum L(a(y), H_0(\boldsymbol{D})) \\ &+ P(H_1(\boldsymbol{D})) \sum L(a(y), H_1(\boldsymbol{D})) \\ =& \sum_{y\in\mathcal{E}} P(H_0(\boldsymbol{D})) + \sum_{y\in\mathcal{T}} (b_1 P(H_0(\boldsymbol{D})) + b_3 P(H_1(\boldsymbol{D}))) \\ &+ \sum_{y\in\mathcal{R}} b_2 P(H_1(\boldsymbol{D})) \\ \propto& \sum_{y\in\mathcal{E}} \mathrm{PrBF}_{0,1} + \sum_{y\in\mathcal{T}} (b_1 \mathrm{PrBF}_{0,1} + b_3) + \sum_{y\in\mathcal{R}} b_2 \end{aligned}$$

where $\boldsymbol{D} = \{y, n\}$, and $P(H_k(\boldsymbol{D}))$ denotes the posterior weight of the posterior predictive model from the predictive Bayes factor with $k = 0, 1$ (Appendix A). After some algebra, it can be shown that $R(a)$ is minimized with the following decomposition of the sample space:

$$\begin{aligned} \mathcal{R} &= \left\{ y : \mathrm{PrBF}_{0,1} > \frac{b_2 - b_3}{b_1} \right\}, \\ \mathcal{T} &= \left\{ y : \frac{b_3}{1-b_1} \leq \mathrm{PrBF}_{0,1} \leq \frac{b_2 - b_3}{b_1} \right\}, \\ \mathcal{E} &= \left\{ y : \mathrm{PrBF}_{0,1} < \frac{b_3}{1-b_1} \right\}. \end{aligned}$$

The action sets $\mathcal{R}$, $\mathcal{T}$, and $\mathcal{E}$ are determined by b_1, b_2, and b_3, which characterize the potential loss from incorrect decisions, or equivalently the relative tolerance to risk.

We denote C the $\mathrm{PrBF}_{0,1}$ lower bound for $\mathcal{R}$, and E the $\mathrm{PrBF}_{0,1}$ upper bound for $\mathcal{E}$. It is easy to derive the following property for the PoP design:

Theorem 1. (Global optimality) The PoP design minimizes the risk of the incorrect decision of dose assignment when $C = \frac{b_2 - b_3}{b_1}$ and $E = \frac{b_3}{1-b_1}$.

C and E specify the decision rules and determine the decision boundaries for dose transition and exclusion. In the trial implementation, we set $1 < C < e$. Large C values encourage dose transition. It will be more likely to retain the current dose if C becomes smaller, which poses a stricter rule for dose transition. Similarly, smaller E values suggest a stricter rule for dose exclusion.

Theorem 1 suggests the optimal thresholds for $\mathrm{PrBF}_{0,1}$ in (3) to achieve both the safety and efficiency of the PoP design, as the risk of incorrect dose assignment is minimized at each time point of the adaptive decision-making process. Importantly, Theorem 1 is derived against the interval hypothesis $H_1: \pi_j \neq \phi$, making it unnecessary to specify any point hypotheses for ϕ_1 and ϕ_2 in (2) for dose escalation/de-escalation. To distinguish this optimality result from those specific to point hypotheses, we refer to it as the *global optimality*, in the sense that it provides the best solution across the entire composite alternative space. This global optimality facilitates the PoP design with a unified performance of high efficiency and accuracy in finding MTD.

3.2. Consistency and Convergence Rate

Here we temporarily shift the focus to the large-sample property of the design, which has become increasingly important due to the high failure rate of correctly identifying MTD in dose-finding trials. While interval-based designs such as mTPI, BOIN and Keyboard have been widely used, they suffer from a common drawback: they may fail to distinguish suboptimal doses from the true MTD when the tolerance interval covers multiple dose levels. Subsequently, these designs cannot guarantee the convergence to the true MTD. This is mainly because the lower and upper convergence boundaries of these interval-based designs do not effectively exploit the improving precision of the toxicity rate as the sample size grows. As a result, any dose level within the tolerance interval may trigger a dose-transition stopping, while retaining the current dose with probability 1.

In contrast, the PoP design uses the predictive Bayes factor to overcome this limitation. We demonstrate that the proposed design has the desirable asymptotic dose-finding property, which is shown in the following theorem:

Theorem 2. (Consistency) Dose allocation based on the escalation and de-escalation boundaries in the PoP design converges almost surely to the dose level j^* if $\pi_{j^*} = \phi$. If no dose level achieves the toxicity rate ϕ, the PoP design would eventually oscillate between two dose levels at which the associated toxicity rates straddle ϕ.

The proof is given in Appendix B. Theorem 2 states that the PoP design will continue to seek MTD until it is found. If none of the dose levels have a toxicity rate of ϕ, the oscillation between the two adjacent doses around ϕ could still provide an effective approach to collecting DLT data for isotonic MTD estimation at the end of a trial. Compared to other interval-based designs, the PoP design has the advantage of more accurately identifying

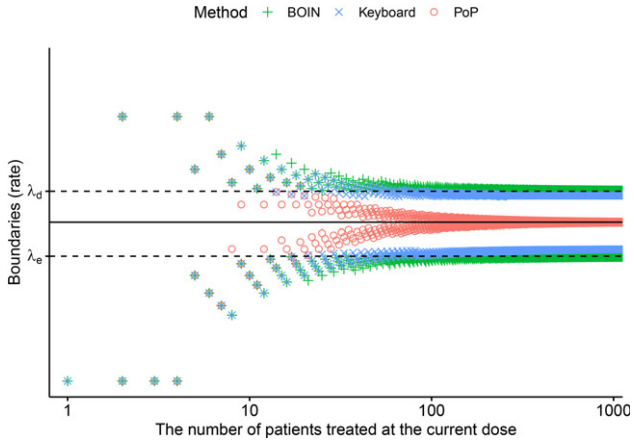

Figure 1. Consistency of interval-based designs. The scatterplot of escalation and de-escalation boundaries of PoP (red dot), Keyboard (blue cross) and BOIN (green plus) design as the number of patients treated at current dose increases. The figure corresponds to $\phi = 0.3$.

MTD and distinguishing it from suboptimal doses, which makes it a better tool for dose-finding trials.

Figure 1 illustrates the consistency property of the PoP design, which is not guaranteed in other interval-based designs (Liu and Yuan 2015; Yan, Mandrekar, and Yuan 2017). The figure shows the escalation and de-escalation boundaries for the PoP design (red), the BOIN design (green) and the Keyboard design (blue) by varying sample sizes at a target toxicity rate of ϕ=0.3. As the sample size increases, the boundaries of the PoP design are roughly monotonic and both upper and lower boundaries converge to the target toxicity rate, indicating that dose-escalation (or de-escalation) may effectively occur when the observed DLT rate deviates from the target toxicity rate. In contrast, the boundaries of the BOIN design and the Keyboard design converge to an interval that may include suboptimal doses higher or lower than MTD.

Furthermore, as the only interval-based design that converges to MTD, we are interested in estimating the convergence rate of the PoP design. We denote the lower and upper transition boundaries with n patients treated at a given dose level as

$$L_n = \frac{\arg\min_y \mathrm{PrBF}_{0,1}(y,n) \geq C}{n} \quad \text{and}$$
$$U_n = \frac{\arg\max_y \mathrm{PrBF}_{0,1}(y,n) \geq C}{n}$$

respectively, where $C > 0$ is an arbitrary threshold. In Appendix C, we prove the following theorem:

Theorem 3. (Convergence rate) Under the condition $C < e$, we have the boundaries of PoP design satisfy $|L_n - \phi| \leq kn^{-1/2}$ and $|U_n - \phi| \leq kn^{-1/2}$ as $n \to \infty$, where $k = \sqrt{2\phi(1-\phi)(1-\log C)}$.

Theorem 3 shows that the lower and upper boundaries of the PoP design converge to the target toxicity rate ϕ at a rate of $n^{-1/2}$, the same as the optimal convergence rate achievable for an unbiased estimator using Cramer-Rao lower bounds. This implies that the PoP design is highly efficient in identifying MTD.

3.3. Coherence

The concept of coherence for trial design was introduced by Cheung (2005), and later extended to long-memory coherence by Liu and Yuan (2015). While the original (also known as short-memory) coherence only concerns the observations from the most recent cohort, a long-memory coherent design prevents counter-intuitive dose escalation (or de-escalation) if the observed toxicity rate at the current dose is larger (smaller) than the target toxicity rate ϕ. It is a principle that aims to prohibit the next patient from receiving overly toxic (subtherapeutic) doses when newly observed data show a reverse signal. Similar to other interval-based designs that use accumulative data to decide dose escalation or de-escalation, the PoP design possesses a long-memory coherence property, but does not strictly adhere to the short-memory coherence principle. We have the following theorem:

Theorem 4. (Long-memory coherence) The PoP design is long-memory coherent in the sense that the probability of dose escalation (or de-escalation) is zero when the observed toxicity rate $\hat{\pi}_j$ at the current dose is higher (or lower) than the target toxicity rate ϕ.

The proof of Theorem 4 is straightforward and is provided in Appendix D. The long-memory coherence is an attractive feature of many interval-based designs, including the PoP, BOIN, and Keyboard designs. It addresses the safety and ethical concerns about dose escalation by ensuring that no subject will be treated at an overly toxic dose when the observed toxicity rate is higher than the target. Moreover, it enhances the efficiency of the trial by ensuring that patients are not allocated to overly safe doses when the observed rate is below the target.

4. Practical Implementation

In practice, implementing the PoP design is simple and straightforward for a clinical research team as dose assignments can be made through the observed DLTs y_j rather than $\mathrm{PrBF}_{0,1}$, with decision boundaries explicitly tabulated for y_j. Table 1 presents the decision boundaries for up to 30 patients with a cohort size of 3 and a target DLT probability of $\phi = 0.25$, where $C = 2.5$ and $E = 5/24$. Figure 2 illustrates the process of the PoP design for conducting a trial starting at the lowest dose.

The values of C and E in Table 1 are determined by minimizing the risk with loss function specified by loss scores $b_1 = 0.2$, $b_2 = 2/3$ and $b_3 = 1/6$. These loss scores $\{b_1, b_2, b_3\}$, similar to the target DLT probability of ϕ, are trial design characteristics that need to be specified by clinical collaborators for any oncology trial given its domain knowledge. In practice, we recommend setting $b_2 \in [0.5, 0.75]$ to explore an alternative dose when the dose seemingly deviates from the true MTD, as is the case here with $b_2 = 2/3$. We recommend setting b_3 as $\frac{1}{4}b_2$ to improve the efficiency of dose transition. To ensure that there exists y_j such that $\mathrm{PrBF}_{0,1}(y_j, n_j) < C$, b_1 must be greater than $b_2/(1+e)$, as the predictive Bayes factor converges to e as $n \to \infty$. Moreover, we believe that b_1 should be greater than b_3, as the loss of transition is likely smaller when the dose is not MTD compared to when it is. Thus, a suggested range of b_1 is

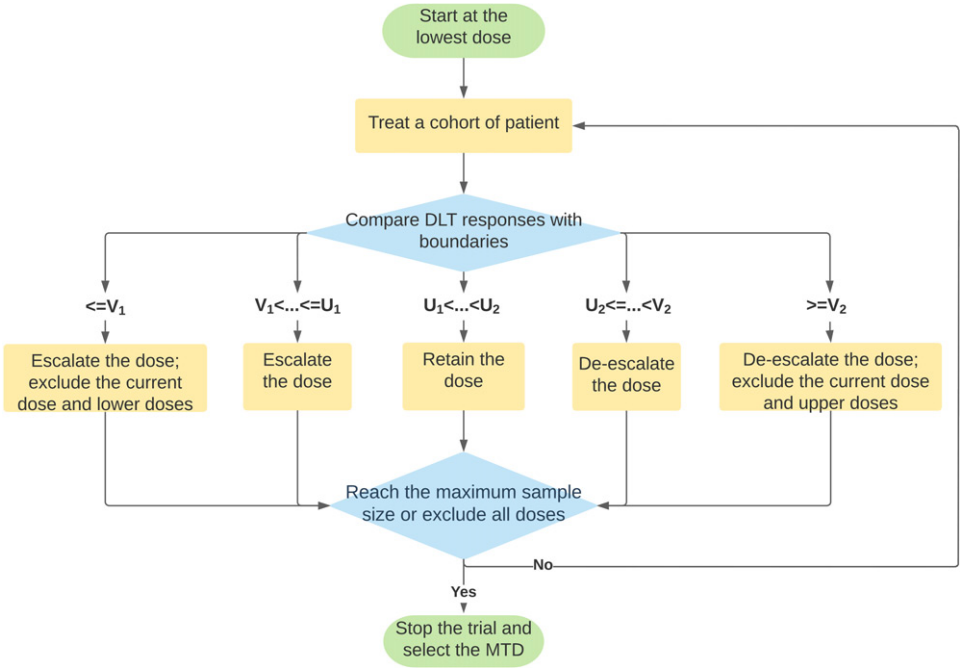

Figure 2. Flowchart of the PoP design.

Table 1. Decision boundaries for y_j in the PoP design with a cutoff value $C = 2.5$, $E = 5/24$, and with a target toxicity $\phi = 0.25$.

		Number of patients treated at the current dose									
	n_j :	3	6	9	12	15	18	21	24	27	30
Escalation if no. of DLT $\leq$	U_1 :	0	0	1	2	2	3	4	4	5	6
De-escalation if no. of DLT $\geq$	U_2 :	2	3	3	4	5	6	7	7	8	9
Overly safe exclusion $\leq$	V_1 :	NA	NA	NA	NA	0	0	1	1	2	2
Overly toxic exclusion $\geq$	V_2 :	3	5	6	7	8	9	11	12	13	14

[0.2, 0.3]. It is demonstrated that the PoP design is not sensitive to the choice of b_1, b_2 and b_3 for practical implementation (see e.g., sensitivity analysis over combinations of loss scores in an appropriate range in Appendix I) with a small sample size. When the sample size is moderate or large, the loss scores will have less impact owing to the guaranteed convergence to MTD.

The PoP design terminates the trial either when the maximum sample size is reached or when all doses are excluded. MTD is then selected according to the rules specified in Section 2.2. The *PoPdesign* package, which can be found on CRAN (Fu, Fang, and Zhou 2022), provides an R implementation for phase I trial design.

5. Simulation

To assess the design operating characteristics, here we present numerical simulation studies to examine and compare the empirical performance of the PoP design with those of the BOIN and Keyboard designs. The simulation scenarios follow Zhou et al. (2018), as described in Section 5.1. The primary focus of the comparison is within the interval-based design class.

5.1. Simulation Settings

We assume there are $K = 4, K = 6$ dose levels and the maximum sample size is $N = 30$ for $K = 4$, and $N = 36$, 60, or 96 for $K = 6$, with cohort of size 1 or 3. The target toxicity rates are ϕ=0.2, 0.25, and 0.3. Under each setting, we generate 10,000 random dose-toxicity scenarios using the pseudo-uniform algorithm proposed by Clertant and O'Quigley (2017) as follows. Given the target toxicity rate ϕ and the number of dose levels K, the algorithm follows these steps:

1. Choose one dose level K_0 from the available set with equal probabilities as MTD.
2. Randomly generate M from Beta$(\max\{K - K_0, 0.5\}, 1)$ and set an upper bound $B = \phi + (1 - \phi)M$ for the toxicity probabilities.
3. Continuously sample toxicity probabilities uniformly from $(0, B)$ for each dose level until a scenario is generated where the selected dose level is MTD.

Under each scenario, we simulate 20,000 trials to control the simulation error.

For the PoP design, we set $b_1 = 0.2$, $b_2 = 2/3$, and $b_3 = 1/6$ as the loss function. Accordingly,

$$C = 2.5 \text{ and } E = 5/24.$$

For the BOIN design, we set the boundaries at recommended 0.6ϕ and 1.4ϕ and assigned the equal prior probability to the hypotheses. For the Keyboard design, we use the recommended default values for the key width of 0.1. In the simulation, all three methods are implemented in combination with accelerated

titration steps in which patients are treated one by one from the starting dose—if no DLT was observed, escalate the dose to the next dose level; otherwise, proceed with the decision table with the specified cohort size (Simon et al. 1997). The dose elimination rule described in the "MTD selection" subsection in Section 2.2 has been applied to all methods in simulation for a fair comparison. Additionally, the specifications of CRM skeletons, based on the indifference-interval based approach of Lee and Cheung (2009), are provided in Appendix E.

5.2. Metrics

We consider the following four metrics to measure the operating characteristics of the designs over 20,000 trials for each scenario:

1. The percentage of correct selection (PCS) of the true MTD, defined as the average proportion of trials that selected the true MTD. It is the primary metric for phase I clinical trials, measuring the accuracy of identifying MTD.
2. The average percentage of correct patients allocated (PCA) to MTD.
3. The risk of overdosing, defined as whether more than a certain percentage of patients (e.g., 70% in Zhou et al. (2018)) were treated as the dose level above MTD.
4. The percentage of overdose selection (POS), defined as the average proportion of the trials that selected a dose above the MTD.
5. The average number of patients enrolled in the trial.

5.3. Results

To illustrate the comparative performance of the PoP design, we focus on its differences from other designs. For example, using BOIN as a benchmark, we reported $\Delta\text{PCS} = \text{PCS}^{PoP} -$

Table 2. Performance metrics of the PoP, BOIN, Keyboard, and CRM designs across 10,000 random scenarios for a cohort size of 1.

K=4, N=30	Target ϕ=0.20				Target ϕ=0.25				Target ϕ=0.30			
	PoP	BOIN	KB	CRM	PoP	BOIN	KB	CRM	PoP	BOIN	KB	CRM
PCS, %	**58.2**	54.0	53.4	55.7	**59.4**	54.2	52.8	55.5	**58.6**	54.2	51.9	55.4
PCA, %	46.3	44.5	44.6	**46.4**	**46.7**	44.5	43.8	46.1	45.6	44.6	43.1	**46.0**
Risk of OD 70%, %	**7.9**	9.3	8.7	8.6	9.1	9.2	9.1	**8.9**	**8.4**	10.1	9.4	9.7
POS, %	13.4	13.5	**12.8**	14.3	15.5	14.5	**13.9**	15.7	16.0	15.8	**14.7**	16.9
Avg. sample size	29.1	28.6	28.5	28.5	29.4	28.8	28.5	28.4	29.0	28.8	28.3	28.3
K=6, N=36	Target ϕ=0.20				Target ϕ=0.25				Target ϕ=0.30			
	PoP	BOIN	KB	CRM	PoP	BOIN	KB	CRM	PoP	BOIN	KB	CRM
PCS, %	**49.6**	46.8	45.9	48.0	**51.5**	47.4	46.2	48.8	**50.4**	47.1	45.0	48.8
PCA, %	37.4	36.5	36.4	**38.0**	38.0	36.6	36.1	**38.2**	37.0	36.6	35.4	**38.2**
Risk of OD 70%, %	**9.7**	11.2	10.6	10.7	**10.4**	10.7	10.5	10.9	**9.6**	11.7	10.9	11.7
POS, %	15.5	16.0	**14.8**	16.3	17.6	16.9	**16.1**	17.4	18.2	18.2	**16.7**	18.9
Avg. sample size	35.0	34.7	34.7	34.6	35.4	34.9	34.7	34.7	34.9	34.9	34.5	34.5
K=6, N=60	Target ϕ=0.20				Target ϕ=0.25				Target ϕ=0.30			
	PoP	BOIN	KB	CRM	PoP	BOIN	KB	CRM	PoP	BOIN	KB	CRM
PCS, %	**55.2**	49.4	48.1	53.5	**57.7**	50.8	49.1	54.7	**56.4**	50.4	47.8	54.9
PCA, %	41.5	40.5	40.1	**43.2**	42.5	40.9	40.1	**43.7**	41.4	40.9	39.0	**43.7**
Risk of OD 70%, %	**7.5**	8.9	8.4	9.5	**8.5**	9.0	8.8	9.8	**8.4**	10.4	9.2	10.7
POS, %	12.7	11.5	**10.7**	13.4	14.7	12.7	**12.0**	14.5	15.5	14.0	**12.7**	15.9
Avg. sample size	56.9	57.4	57.3	57.3	57.8	57.8	57.4	57.4	56.9	57.8	57.1	57.1
K=6, N=96	Target ϕ=0.20				Target ϕ=0.25				Target ϕ=0.30			
	PoP	BOIN	KB	CRM	PoP	BOIN	KB	CRM	PoP	BOIN	KB	CRM
PCS, %	**59.7**	50.7	49.1	58.2	**62.6**	52.5	50.6	59.9	**61.2**	52.2	49.3	60.1
PCA, %	44.9	43.5	42.9	**48.0**	46.4	44.4	43.2	**48.9**	45.1	44.4	42.0	**48.9**
Risk of OD 70%, %	**6.0**	7.0	6.8	8.5	**6.9**	7.6	7.3	8.9	**7.1**	9.0	7.8	9.7
POS, %	10.1	8.3	**7.8**	11.0	11.9	9.4	**8.9**	12.0	12.7	10.6	**9.5**	13.3
Avg. sample size	88.3	91.1	91.0	91.0	90.1	91.8	91.2	91.2	88.3	91.8	90.7	90.7

Abbv: KB: Keyboard; PCS: percentage of correct selection; PCA: percentage of correct allocation; POS: percentage of overdose selection; OD: overdosing. Methods that exhibit superior performance are highlighted in bold.

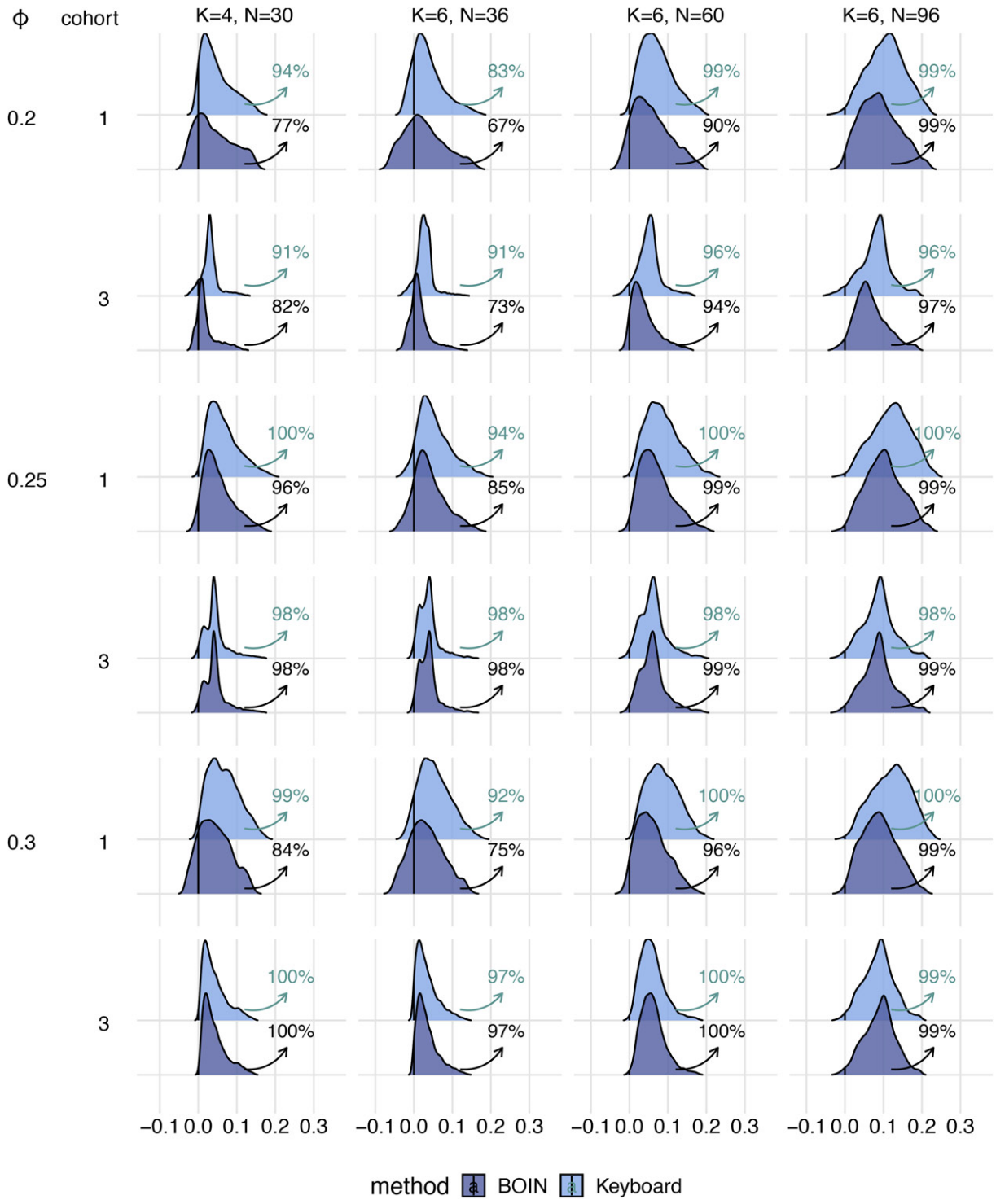

Figure 3. Plots of difference in the percentage of correct selection (ΔPCS) of the maximum tolerated dose (MTD) for PoP versus BOIN and PoP versus Keyboard across 10,000 scenarios. In each scenario, we calculated ΔPCS with 20,000 simulated trials. Arrowed numbers represent the percentage of scenarios in which ΔPCS was positive. A value $>$ 50% indicates the PoP design outperformed the design under comparison (top, Keyboard design, in light blue; bottom, BOIN design, in dark blue).

PCS^{BOIN}, where a positive ΔPCS value indicates superior performance of the PoP design in accuracy of identifying MTD. Table 2 presents the average of each metric described above for each design.

MTD selection. Figure 3 demonstrates the performance of the PoP design with respect to MTD selection based on PCS under various settings. Generally, the PoP design yielded higher PCS than the BOIN and Keyboard designs. In each panel of Figure 3, we highlighted the percentage of scenarios in which $PCS^{PoP} - PCS^{\dagger}$ was positive († for either Keyboard design (top, in light blue) or BOIN design (bottom, in dark blue), respectively), indicating that PoP design consistently outperformed the BOIN and Keyboard designs to select the correct MTD. When the cohort size was 1, the percentage of positive difference in PCS was larger than that of when the cohort size was 3, which indicated that the PoP design had even higher accuracy in selecting

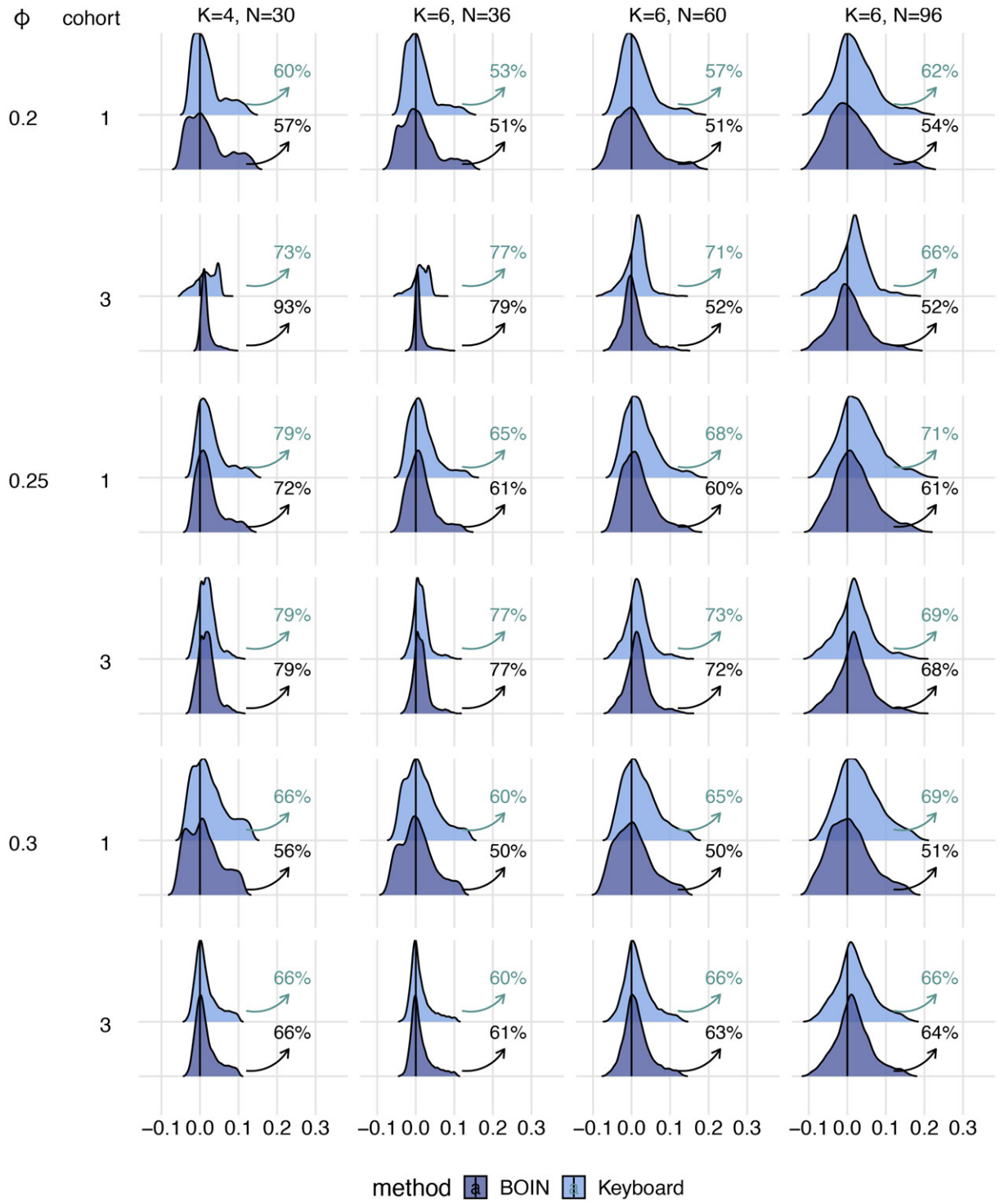

Figure 4. Density plots of difference in the percentage of correct allocation (ΔPCA) of patients to the maximum tolerated dose (MTD) for PoP versus BOIN and PoP versus Keyboard across 10,000 scenarios. In each scenario, we averaged ΔPCA with 20,000 simulated trials. Arrowed numbers represent the percentage of scenarios in which ΔPCA was positive. A value $>$ 50% indicates the PoP design outperformed the design under comparison (top, Keyboard design, in light blue; bottom, BOIN design, in dark blue).

MTD for continuous interim assessment. When there were more patients relative to the dose numbers (e.g., with $K = 6, N = 60$ and $K = 6, N = 96$), the PoP design outperformed the BOIN and Keyboard designs in more dose-toxicity scenarios, which aligned with the consistency property discussed in Section 3.2.

Patient allocation. Figure 4 depicts the comparative performance of each design with respect to assigning patients to MTD according to the average percentage of correct patient allocation (PCA). In most of the panels, the majority parts of the density were positive, indicating that more patients in the PoP design would be treated at the true MTD than in the BOIN and Keyboard designs. When the cohort size was 3, the percentage of positive difference in PCA was often larger than that of when the cohort size was 1.

Overdose control. The risk of overdosing at least 70% of patients and the percentage of overdose allocation are shown in

Table 3. Performance metrics of the PoP, BOIN, Keyboard, and CRM designs under four prespecified dose-toxicity scenarios for cohort sizes of 1. The target toxicity rate is 0.25. $N = 36$.

Design		Target ϕ=0.25							
		Dose level						Not Choosing Any Dose	Risk of Over-dosing
		1	2	3	4	5	6		
Scenario 1	Pr(toxicity)	0.25	0.35	0.5	0.6	0.7	0.8		
PoP	Selection(%)	**63.7**	26.1	1.9	0.1	0.0	0.0	8.1	19.5
	# Patients	**20.4**	10.2	3.0	0.7	0.2	0.0		
BOIN	Selection(%)	54.4	23.2	1.8	0.1	0.0	0.0	20.5	20.1
	# Patients	17.9	9.2	2.9	0.9	0.3	0.1		
Keyboard	Selection(%)	53.4	22.3	1.6	0.1	0.0	0.0	22.5	19.5
	# Patients	17.6	8.9	2.8	0.8	0.2	0.1		
CRM	Selection(%)	53.0	22.5	1.6	0.0	0.0	0.0	22.8	17.7
	# Patients	18.0	8.9	2.6	0.6	0.2	0.0		
Scenario 2	Pr(toxicity)	0.1	0.25	0.4	0.6	0.7	0.8		
PoP	Selection(%)	15.0	**68.1**	16.5	0.3	0.0	0.0	0.2	8.6
	# Patients	8.7	16.6	8.3	1.6	0.3	0.1		
BOIN	Selection(%)	23.0	60.6	14.7	0.5	0.1	0.0	1.0	9.1
	# Patients	10.2	15.8	7.4	1.8	0.5	0.1		
Keyboard	Selection(%)	24.5	59.3	13.9	0.5	0.1	0.0	1.7	8.6
	# Patients	10.6	15.6	7.0	1.7	0.4	0.1		
CRM	Selection(%)	14.4	66.3	17.4	0.3	0.0	0.0	1.7	7.5
	# Patients	8.6	**17.3**	7.8	1.4	0.3	0.1		
Scenario 3	Pr(toxicity)	0.05	0.1	0.25	0.32	0.5	0.6		
PoP	Selection(%)	0.2	15.7	50.7	30.3	3.0	0.1	0.0	18.5
	# Patients	2.2	7.7	13.0	9.1	3.1	0.7		
BOIN	Selection(%)	1.1	21.8	45.2	28.5	3.1	0.2	0.1	18.6
	# Patients	2.4	8.8	11.9	8.9	3.1	0.9		
Keyboard	Selection(%)	1.8	23.2	44.4	27.4	2.8	0.1	0.3	18.4
	# Patients	2.6	9.1	11.8	8.7	2.9	0.8		
CRM	Selection(%)	0.1	12.7	**55.2**	29.3	2.4	0.0	0.3	19.2
	# Patients	2.1	7.0	**14.4**	9.3	2.5	0.6		
Scenario 4	Pr(toxicity)	0.01	0.02	0.03	0.04	0.05	0.25		
PoP	Selection(%)	0.0	0.0	0.0	0.1	12.5	87.4	0.0	0.0
	# Patients	1.1	1.3	1.4	1.7	7.7	21.9		
BOIN	Selection(%)	0.0	0.0	0.0	0.2	22.2	77.5	0.0	0.0
	# Patients	1.1	1.2	1.3	1.6	9.5	21.1		
Keyboard	Selection(%)	0.0	0.1	0.1	0.4	24.1	75.1	0.0	0.0
	# Patients	1.1	1.3	1.4	1.7	10.0	20.6		
CRM	Selection(%)	0.0	0.0	0.0	0.2	10.8	**88.9**	0.0	0.0
	# Patients	1.1	1.1	1.3	1.7	6.6	**24.1**		

Table 2. Because it is critical to explore the boundary between the overly toxic doses and MTD in a dose-finding trial, rather than completely avoiding the overly toxic assignment, an attractive design feature is to minimize the risk of exposing too many patients to overly toxic doses. Compared with the BOIN and Keyboard designs, the PoP design had the lowest risk of overdose

allocation. Meanwhile, PoP design showed robustness in dose selection, with POS slightly higher than BOIN and Keyboard but marginally lower than CRM. Coupled with its significant advantage in PCS over BOIN and Keyboard, the simulation studies confirm that the PoP design achieved superior empirical performance among interval-based designs for dose selection.

Table 3 presents four prespecified scenarios for oncology trials in Liu and Yuan (2015). We evaluated these scenarios for $\phi = 0.25$ (Table 3) to illustrate cases where a specified dose matches the target toxicity rate and $\phi = 0.20$ (Table A1 in Appendix F) where specified doses oscillate around the MTD. The results showed that the PoP design outperformed the CRM design in Scenarios 1 and 2. The CRM design performed best in Scenarios 3 and 4, with the PoP design being the second best. However, isolated scenarios across different design categories should be interpreted with caution, as performance can substantially vary by scenario. In general, among the interval-based designs, the PoP design yielded the highest PCS and PCA in all scenarios.

The distributions of MTD recommendation by different designs were shown in Appendix G. A numerical example in Appendix H demonstrated the consistency property of the PoP design. Finally, a sensitivity analysis that evaluates the performance of the PoP design with respect to various sets of loss functions was included in Appendix I.

6. Discussion

We have proposed the PoP design, a cutting-edge interval-based design for phase I dose-finding clinical trials. Given a binary primary endpoint over multiple dose levels at a small sample size, a key challenge is to maintain algorithm simplicity and safety while improving the efficiency of correct MTD selection. The PoP design shares common advantages with other interval-based designs, such as its straightforward implementation, facilitated by pretabulated decision rules, and its long-memory coherence. Remarkably, the PoP design has demonstrated outstanding performance in dose selection compared with existing phase I designs. As suggested by the theoretical and empirical results, this research endeavor may, while keeping the trial implementation simple, help to better identify MTD and deliver safer and more effective phase I clinical studies.

The proposed method is established based on a robust predictive Bayesian hypothesis testing approach to guide dose transition. The decision rules in the PoP design can achieve global optimality by minimizing the predictive risk of the incorrect decision of dose assignment during the trial, with the potential of customized loss functions to meet different clinical needs. One unique feature of the PoP design is that for a large sample size, it is consistent by distinguishing the true MTD from the suboptimal doses, therefore ameliorating the inconsistency concern of phase I interval-based trial designs to select the suboptimal MTD. As more modern phase I trials start to recruit patients in moderate sample sizes, this property will be increasingly important and appealing for dose-finding problems.

We note that the optimality property discussed in Section 3.1 differs from the optimality property in popular interval-based designs (Liu and Yuan 2015; Guo et al. 2017; Yan, Mandrekar, and Yuan 2017), rooted from different alternative hypotheses to guide dose transitions. For example, the BOIN design has the *local* optimality for a prespecified set of (λ_d, λ_e), which respectively corresponds to two local alternative point hypotheses for dose de-escalation and escalation. Given a different set of point hypotheses, the optimal decision boundaries must be recalibrated. Similarly, the optimality in the Keyboard design is dependent on the preselected length of the key. In contrast, the PoP design achieves the global optimality for dose transition based on Bayesian testing against the global interval alternative hypothesis in (1) when the predictive Bayes factor assures the sound asymptotic and finite properties without suffering from Lindley's paradox. Subsequently, the global optimality allows the PoP design to exploit the accumulated data more effectively and adaptively, which explains why the PoP design is both more efficient in selecting MTD and safer with a lower risk of assigning patients to overly toxic or subtherapeutic doses.

At the early stage of a trial with few accrued patients, the PoP design tends to be more conservative in dose escalation and de-escalation than the other interval-based designs. It may require more patients treated at a given dose for enough evidence to make correct decisions for dose transition. For example, the dose might only escalate 0/6 in Table 1 rather than the conventional 1/6 in mTPI, BOIN, keyboard or "3+3" designs. A possible extension is to calibrate the loss function to encourage dose transition when n_j is small. Practically speaking, the losses b_1, b_2, and b_3 need not be constant and how best to specify them deserves further study, as the sample size-dependent loss function is fully reflected in the decision boundary tabulation, so that the trial can be the same easily implemented from user's end, while theoretically, all the attractive properties like optimality, coherence, and consistency are still maintained. Nevertheless, an accelerated titration step is helpful and recommended to reduce the overall number of undertreated subjects.

Like many conventional phase I trial designs, the PoP design relies on rapidly ascertaining toxicity relative to patient accrual, requiring the toxicity outcomes of previously treated patient cohorts to be fully assessable by the time a new cohort is ready for dose assignment (Cheung and Chappell 2000). Extending the PoP design for time-to-event outcomes to effectively manage delayed toxicities remains an area for future research. Another limitation of the PoP design is that it only focuses on toxicity outcomes without efficacy evaluation. An interesting future direction is to extend the PoP design to phase I/II trials for simultaneous efficacy and toxicity monitoring. In fact, the US Food and Drug Administration launched the initiative of "Project Optimus" in 2021 to reform the dose optimization and dose selection paradigm in oncology drug development with the goal of finding safe and efficacious doses (FDA 2022; Fourie Zirkelbach et al. 2022; Korn, Moscow, and Freidlin 2023; Thall et al. 2024; Yuan, Zhou, and Liu 2024). As demonstrated in Table 2, the underlying algorithm becomes more effective when the sample size increases. With the relatively larger sample size in the phase I/II dose-finding trials, it will be attractive to apply the posterior predictive probability approach in pursuing global optimality as a key design feature when evaluating both toxicity and efficacy together.

Supplementary Materials

The supplemental materials include a brief introduction to the predictive Bayes factor (Section A); proofs of Theorems 2, 3, and 4 (Sections B, C, and

D, respectively); the CRM simulation settings (Section E); and additional simulation results (Sections F, G, H, and I).

Acknowledgments

The authors are grateful to the handling editor, associate editor, and reviewers for their thoughtful comments and suggestions, which have significantly improved the clarity and rigor of this article. We are also grateful to Ms. Xinying Fang, who assisted in the development of the R package *PoPdesign*.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

Zhou is partially supported by the Pennsylvania State University Early Career Research Award and the Pennsylvania Department of Health TSF Cure Program. Lee is partially supported by the grant P30CA016672 from the National Cancer Institute.

ORCID

Chenqi Fu http://orcid.org/0000-0001-6387-6358
Shouhao Zhou http://orcid.org/0000-0002-8124-5047

References

Bernardo, J. M. (1999), "Nested Hypothesis Testing: The Bayesian Reference Criterion," *Bayesian Statistics*, 6, 101–130. [2647]

Berry, S. M., Carlin, B. P., Lee, J. J., and Muller, P. (2010), *Bayesian Adaptive Methods for Clinical Trials*, Boca Raton, FL: CRC Press. [2648]

Braun, T. M. (2002), "The Bivariate Continual Reassessment Method: Extending the CRM to Phase I Trials of Two Competing Outcomes," *Controlled Clinical Trials*, 23, 240–256. [2646]

Brunk, H., Barlow, R. E., Bartholomew, D. J., and Bremner, J. M. (1972), "Statistical Inference Under Order Restrictions. (The Theory and Application of Isotonic Regression)," Technical Report, Missouri Univ Columbia Dept of Statistics. [2648]

Cheung, Y. K. (2005), "Coherence Principles in Dose-Finding Studies," *Biometrika*, 92, 863–873. [2650]

Cheung, Y. K., and Chappell, R. (2000), "Sequential Designs for Phase I Clinical Trials with Late-Onset Toxicities," *Biometrics*, 56, 1177–1182. [2656]

Clertant, M., and O'Quigley, J. (2017), "Semiparametric Dose Finding Methods," *Journal of the Royal Statistical Society*, Series B, 79, 1487–1508. [2651]

Faries, D. (1994), "Practical Modifications of the Continual Reassessment Method for Phase I Cancer Clinical Trials," *Journal of Biopharmaceutical Statistics*, 4, 147–164. [2646]

FDA. (2022), "Project Optimus," *https://www.fda.gov/about-fda/oncology-center-excellence/project-optimus*. [2656]

Fourie Zirkelbach, J., Shah, M., Vallejo, J., Cheng, J., Ayyoub, A., Liu, J., Hudson, R., Sridhara, R., Ison, G., Amiri-Kordestani, L., et al. (2022), "Improving Dose-Optimization Processes Used in Oncology Drug Development to Minimize Toxicity and Maximize Benefit to Patients," *Journal of Clinical Oncology*, 40, 3489–3500. [2656]

Fu, C., Fang, X., and Zhou, S. (2022), "PoPdesign: Posterior Predictive (PoP) Design for Phase I Clinical Trials," available at *https://cran.rstudio.com/web/packages/PoPdesign/index.html* [2651]

Guo, W., Wang, S.-J., Yang, S., Lynn, H., and Ji, Y. (2017), "A Bayesian Interval Dose-Finding Design Addressingockham's Razor: mtpi-2," *Contemporary Clinical Trials*, 58, 23–33. [2646,2656]

Iasonos, A., and O'Quigley, J. (2014), "Adaptive Dose-Finding Studies: A Review of Model-Guided Phase I Clinical Trials," *Journal of Clinical Oncology*, 32, 2505. [2646]

Jaki, T., Clive, S., and Weir, C. J. (2013), "Principles of Dose Finding Studies in Cancer: A Comparison of Trial Designs," *Cancer Chemotherapy and Pharmacology*, 71, 1107–1114. [2646]

Ji, Y., and Wang, S. J. (2013), "Modified Toxicity Probability Interval Design: A Safer and More Reliable Method than the 3 + 3 Design for Practical Phase I Trials," *Journal of Clinical Oncology*, 31, 1785–1791. [2646]

Kass, R. E., and Raftery, A. E. (1995), "Bayes Factors," *Journal of the American Statistical Association*, 90, 773–795. [2647]

Korn, E. L., Moscow, J. A., and Freidlin, B. (2023), "Dose Optimization During Drug Development: Whether and When to Optimize," *Journal of the National Cancer Institute*, 115, 492–497. [2656]

Le Tourneau, C., Lee, J. J., and Siu, L. L. (2009), "Dose Escalation Methods in Phase I Cancer Clinical Trials," *JNCI: Journal of the National Cancer Institute*, 101, 708–720. [2646,2648]

Lee, S. M., and Cheung, Y. K. (2009), "Model Calibration in the Continual Reassessment Method," *Clinical Trials*, 6, 227–238. [2652]

Leung, D. H.-Y., and Wang, Y.-G. (2002), "An Extension of the Continual Reassessment Method Using Decision Theory," *Statistics in Medicine*, 21, 51–63. [2646]

Lindley, D. V. (1957), "A Statistical Paradox," *Biometrika*, 44, 187. [2647]

Liu, S., and Yuan, Y. (2015), "Bayesian Optimal Interval Designs for Phase I Clinical Trials," *Journal of the Royal Statistical Society*, Series C, 64, 507–523. [2646,2647,2648,2650,2656]

O'Quigley, J., Pepe, M., and Fisher, L. (1990), "Continual Reassessment Method: A Practical Design for Phase 1 Clinical Trials in Cancer," *Biometrics*, 46, 33–48. [2646]

Oron, A. P., Azriel, D., and Hoff, P. D. (2011), "Dose-Finding Designs: The Role of Convergence Properties," *International Journal of Biostatistics*, 7, 39. [2646]

Pan, H., Lin, R., Zhou, Y., and Yuan, Y. (2020), "Keyboard Design for Phase I Drug-Combination Trials," *Contemporary Clinical Trials*, 92, 105972. [2647]

Simon, R., Rubinstein, L., Arbuck, S. G., Christian, M. C., Freidlin, B., and Collins, J. (1997), "Accelerated Titration Designs for Phase I Clinical Trials in Oncology," *Journal of the National Cancer Institute*, 89, 1138–1147. [2652]

Thall, P. F., Garrett-Mayer, E., Wages, N. A., Halabi, S., and Cheung, Y. K. (2024), "Current Issues in Dose-Finding Designs: A Response to the US Food and Drug Adminstrations's Oncology Center of Excellence Project Optimus," *Clinical Trials*, 21, 267–272. [2656]

van Brummelen, E. M., Huitema, A. D., van Werkhoven, E., Beijnen, J. H., and Schellens, J. H. (2016), "The Performance of Model-based versus Rule-based Phase I Clinical Trials in Oncology," *Journal of Pharmacokinetics and Pharmacodynamics*, 43, 235–242. [2646]

Yan, F., Mandrekar, S. J., and Yuan, Y. (2017), "Keyboard: A Novel Bayesian Toxicity Probability Interval Design for Phase I Clinical Trials," *Clinical Cancer Research*, 23, 3994–4003. [2646,2648,2650,2656]

Yin, G., and Yuan, Y. (2009), "Bayesian Model Averaging Continual Reassessment Method in Phase I Clinical Trials," *Journal of the American Statistical Association*, 104, 954–968. [2646]

Yuan, Y., Lee, J. J., and Hilsenbeck, S. G. (2019), "Model-Assisted Designs for Early-Phase Clinical Trials: Simplicity Meets Superiority," *JCO Precision Oncology*, 3, 1–12. [2646]

Yuan, Y., Lin, R., and Lee, J. J. (2022), *Model-Assisted Bayesian Designs for Dose Finding and Optimization: Methods and Applications*, Boca Raton, FL: CRC Press. [2646]

Yuan, Y., Zhou, H., and Liu, S. (2024), "Statistical and Practical Considerations in Planning and Conduct of Dose-Optimization Trials," *Clinical Trials*, 21, 273–286. [2656]

Zhou, H., Murray, T. A., Pan, H., and Yuan, Y. (2018), "Comparative Review of Novel Model-Assisted Designs for Phase I Clinical Trials," *Statistics in Medicine*, 37, 2208–2222. [2646,2651,2652]

Zhou, S. (2011), "Bayesian Model Selection in Terms of Kullback-Leibler Discrepancy," PhD thesis, Columbia University. [2647,2648]

Zhu, Y., Hwang, W.-T., and Li, Y. (2019), "Evaluating the Effects of Design Parameters on the Performances of Phase I Trial Designs," *Contemporary Clinical Trials Communications*, 15, 100379. [2646]