Received: 3 November 2022 | Revised: 2 April 2023 | Accepted: 6 April 2023

DOI: 10.1002/pst.2304

WILEY

MAIN PAPER

Time-to-event calibration-free odds design: A robust efficient design for phase I trials with late-onset outcomes

Huaqing Jin[1] | **Guosheng Yin**[2]

[1]Department of Radiology and Biomedical Imaging, University of California, San Francisco, California, USA

[2]Department of Mathematics, Imperial College London, London, UK

Correspondence
Guosheng Yin, Department of Mathematics, Imperial College London, London, UK.
Email: guosheng.yin@imperial.ac.uk

Abstract

Compared with most of the existing phase I designs, the recently proposed calibration-free odds (CFO) design has been demonstrated to be robust, model-free, and easy to use in practice. However, the original CFO design cannot handle late-onset toxicities, which have been commonly encountered in phase I oncology dose-finding trials with targeted agents or immunotherapies. To account for late-onset outcomes, we extend the CFO design to its time-to-event (TITE) version, which inherits the calibration-free and model-free properties. One salient feature of CFO-type designs is to adopt game theory by competing three doses at a time, including the current dose and the two neighboring doses, while interval-based designs only use the data at the current dose and is thus less efficient. We conduct comprehensive numerical studies for the TITE-CFO design under both fixed and randomly generated scenarios. TITE-CFO shows robust and efficient performances compared with interval-based and model-based counterparts. As a conclusion, the TITE-CFO design provides robust, efficient, and easy-to-use alternatives for phase I trials when the toxicity outcome is late-onset.

KEYWORDS

Bayesian design, dose-finding trial, late-onset toxicity, oncology, phase I trial

1 | INTRODUCTION

As the first-in-human study, the phase I trial is a crucial step during the development of new treatment, which directly affects the subsequent phase II or III trials. The main target of a phase I oncology trial is to determine the maximum tolerated dose (MTD) which is typically defined as the dose with the dose-limiting toxicity (DLT) probability closest to a pre-determined target toxicity rate.[1] The major barrier for conducting a phase I trial is the limited number of available subjects,[2] which makes the identification of MTD rather challenging.

Currently, depending on whether to adopt a model assumption on the dose–toxicity curve or not, the phase I designs have two mainstreams: the algorithm-based (model-free) and the model-based approaches. Examples of the algorithm-based methods include the $3+3$ design,[3] which is the most commonly used design for phase I oncology trials, the cumulative cohort design (CCD),[4] the modified toxicity probability interval (mTPI) design[5] and the Bayesian optimal interval (BOIN) design.[6] Recently, more algorithm-based designs have emerged. For example, the keyboard design was developed[7] by partitioning the toxicity probability scale into more and shorter intervals. Inspired by the uniformly most powerful Bayesian test,[8] Lin and Yin[9] proposed the uniformly most powerful Bayesian interval (UMPBI) design for phase I dose-finding trials. The algorithm-based methods typically only consider the data at the current dose level for the next dose assignment with no regard to data at other dose levels, thus are generally less efficient compared

with the model-based designs, which consider accrued information at all dose levels. The most popular model-based design is the continual reassessment method (CRM)[10,11] and it often uses a single unknown parameter to link the true toxicity rates with the prespecified toxicity rates at different dose levels. Due to the popularity of the CRM, it has many extentions[12–15] aiming for enhancing its practical performance. Another model-based design, called the escalation with overdose control (EWOC),[16] pays more attention to the safety aspect in a phase I trial. Lin and Yin[17] devised the nonparametric overdose control design, enhancing model robustness with little sacrifice on trial efficiency. The approximate Bayesian computation (ABC) design[18] is another method, which utilizes the ABC to draw the posterior samples of DLT rates.

For the aforementioned designs, the DLT outcome is assumed to be ascertainable immediately after trial participants receive the treatment. Nevertheless, there has been a trend that more and more clinical trials investigate noncytotoxic therapies such as molecularly targeted therapies and immunotherapies whose toxicity outcomes are often late-onset.[19–21] Such late-onset toxicity causes new challenges for dose finding because the DLT data required for dose assignment may not be available at the decision-making time and thus become missing data. Several phase I designs are proposed for the late-onset DLTs by extending the original design to accommodate time information of delayed outcomes, including the time-to-event CRM (TITE-CRM) design,[22] TITE-CCD design,[4] TITE-EWOC design[23] and TITE-BOIN design.[19] The fractional design is another family of approaches to imputing the unobserved toxicity data by fractionizing the binary outcome with the well-known Kaplan–Meier estimator,[24] which does not need to assume any parametric survival distribution. Chapple and Thall[25] demonstrated a Bayesian design for precision dose finding based on time-to-toxicity in a phase I clinical trial with two or more patient subgroups. Through a novel formulation and approximation of the likelihood of the observed data, Lin and Yuan[26] proposed a general methodology for model-assisted designs to handle pending DLT outcomes. Further, a time-to-event Bayesian optimal interval design[27] was developed to accelerate the dose-finding process by utilizing toxicity grades based on both cumulative and pending toxicity outcomes.

For the existing phase I designs, a salient issue for their implementation is the design parameter calibration. No matter whether a design is model-based or algorithm-based, calibrating the design parameters is typically required in order to achieve good operating characteristics. The calibration step may increase the difficulty and burden for conducting a phase I trial. Moreover, as there is very limited information on the treatment prior to a phase I trial, the calibration step may not be reliable that guarantees good performance of the trial. Therefore, both algorithm-based and model-based methods could be at risk of using inappropriate parameters of the design, which leads to compromised dose assignment and incorrect MTD identification.

To address the calibration problem and ease the implementation of a phase I design, a novel calibration-free odds (CFO) design has been proposed.[28] As indicated by the name, the CFO design does not require to calibrate any essential design parameter. To make it calibration-free, the CFO design adopts the game competition idea, which compares the evidence supporting dose de-escalation and escalation at the current dose level and its two neighboring doses. Similar to a two-player game, one tries to push the dose up and the other tries to push it down. The MTD can be reached when this competition game achieves an equilibrium.

To accommodate late-onset toxicity outcomes, we extend the CFO design to the time-to-event CFO (TITE-CFO) design. The TITE-CFO design inherits the model-free and calibration-free properties of the original CFO, which makes it also easy to use in practice. The TITE-CFO design shows satisfactory performances in the extensive simulation studies in comparison with other competitors. Especially, CFO-type designs are robust, efficient, and safe under different settings and dose-toxicity scenarios. The dose escalation and de-escalation rules can be pre-tabled for practical use. Therefore, they can be regarded as useful and easy-to-use alternatives for conducting phase I trials.

The remainder of the article is organized as follows. In Section 2, we introduce the CFO and TITE-CFO designs, for which a trial example is used for illustration. In Section 3, we give the evaluation methods and compare the TITE-CFO design with other phase I designs through extensive simulation studies. Section 4 concludes this paper with some discussions.

2 | METHODS

2.1 | The CFO design

Borrowing the idea from game theory, the CFO design assigns the dose level for the next cohort of subjects by comparing the evidence from the current dose level and its two neighboring dose levels.[28] More specifically, after enrolling n cohorts of patients, we can calculate the odds of the true DLT rates being greater than the target DLT rate for the current dose level and its two neighboring (left and right) dose levels, denoted as (O_L, O_C, O_R). For $k = L, C, R$, let p_k denote

the toxicity probability at dose level k, and x_k and m_k are the corresponding number of DLTs and number of patients, respectively. The odds of $p_k > \phi$ is calculated as

$$O_k = \frac{\Pr(p_k > \phi | x_k, m_k)}{\Pr(p_k \leq \phi | x_k, m_k)}$$

for $k = L, C, R$ corresponding to left, current and right doses. Similarly, the odds of the true DLT rates being smaller than the target DLT rate, $(\overline{O}_L, \overline{O}_C, \overline{O}_R)$, can also be obtained for the three dose levels. That is, the reciprocal $\overline{O}_k = 1/O_k$ represents the odds of $p_k \leq \phi$.

Intuitively, the odds O_C represents the evidence of the current dose level being overly toxic. If the value of O_C is large, we should consider dose de-escalation. The odds $\overline{O}_L$ measures the evidence of the dose level on the left side of the current dose level being overly tolerable. When $\overline{O}_L$ is large, de-escalation is undesirable. Similar to a two-player game, we compare the evidence from O_C and $\overline{O}_L$. If the evidence from odds O_C is stronger, we prefer dose de-escalation, while if the evidence from odds $\overline{O}_L$ is stronger, it is more reasonable to stay at the current dose level. By competing odds O_C with odds $\overline{O}_L$, we can obtain a vote between the decision on dose de-escalation and dose remaining unchanged.

The competition between two odds can be summarized in the ratio $O_C/\overline{O}_L$, for which a larger value suggests dose de-escalation to be the desirable decision. Because we adopt the posterior distribution of the DLT rate to calculate the odds values, the ratio $O_C/\overline{O}_L$ implicitly takes the sample size into consideration. When the same size increases, the variance of the posterior distribution diminishes, yielding more reliable evidence for dose assignment decisions.[28] To obtain a suitable threshold value γ_L for deciding whether the evidence is strong enough (i.e., whether the odds ratio $O_C/\overline{O}_L$ is large enough to trigger de-escalation), we minimize the probability of making a wrong vote, that is, the probability that the vote is against the true situation.

Analogously, we can obtain a vote between staying at the current dose and escalation by competing odds $\overline{O}_C$ against O_R with respect to the corresponding threshold value γ_R.

By aggregating the two votes, we obtain the decision on which dose level to be assigned for the next cohort as shown in Table 1. The mathematical details on how to calculate the odds and threshold values are provided in Appendix A.1.

When the current dose level is the lowest or highest (i.e., at the boundary), there is only one neighboring dose. In such boundary cases, we only need to calculate the odds ratio and the decision rule on one side and make the corresponding decision based on one vote as shown in Table 2.

As an illustration, we consider a phase I trial with the target DLT rate 0.3. Suppose that at a certain point of the trial, 2 out of 6 patients experienced toxicities at the current dose level. At the left neighboring dose level, there was no DLT outcome for 3 treated patients. At the right neighboring dose level, 2 toxicity outcomes were observed among 3 treated patients. The two threshold values in the CFO design are $(\gamma_L, \gamma_R) = (0.348, 0.318)$. The odds ratios are $(O_C/\overline{O}_L, \overline{O}_C/O_R) = (0.046, 0.027)$ and both are smaller than the corresponding threshold values. By the rules in Table 1, the trial should stay at the current dose level. Hypothetically, if at the current dose level and its right neighboring dose level, both numbers of the observed DLT outcomes are reduced to 1, the odds ratios would change to $(O_C/\overline{O}_L, \overline{O}_C/O_R) = (0.004, 2.952)$ while the threshold values are still the same $(\gamma_L, \gamma_R) = (0.348, 0.318)$, leading to the decision of dose escalation for the next cohort.

TABLE 1 Dose escalation and de-escalation rules of the CFO design.

		$O_C/\overline{O}_L > \gamma_L$	
		Yes (De-escalation)	**No (Stay)**
$\overline{O}_C/O_R > \gamma_R$	Yes (Escalation)	Stay	Escalation
	No (Stay)	De-escalation	Stay

TABLE 2 Dose escalation and de-escalation rules of the CFO design when the current dose level is the lowest or highest.

Current dose	Rule	Yes	No
Lowest dose	$\overline{O}_C/O_R > \gamma_R$	Escalation	Stay
Highest dose	$O_C/\overline{O}_L > \gamma_L$	De-escalation	Stay

According to the illustrative example, the CFO design does not require input of any essential design parameters for its implementation. We only need the target DLT rate and the observed data from the current dose level and its two neighboring dose levels for making the decision among de-escalation, staying at the current dose or escalation. This feature distinguishes the CFO design from most of other adaptive phase I designs. For example, the CRM requires a parametric model skeleton, which can affect the performance of the phase I trial significantly. These parameters may undermine the robustness of the design and hamper its use in practice.

Our CFO design is developed based on the observation that dose level skipping is typically not allowed during the implementation of a phase I trial. Therefore, the decision rule of the CFO design utilizes the information from three dose levels by mimicking a two-player game: the left player of the current dose tries to push the dose up while the right player tries to push it down. This feature makes the CFO design more efficient than most of the algorithm-based methods, because the decision rule of an algorithm-based design is typically based on the data from the current dose level only. During a phase I trial, most of the patients are assigned to the dose levels around the true MTD, so ignoring information from boundary dose levels at the far end would not incur much information loss. Such compromise helps the CFO design to avoid adopting any parametric model assumption, but still deliver performances close to model-based designs. In particular, the large-scale numerical studies demonstrate that the CFO design yields comparable performance to the CRM,[28] yet its robustness advantage over the CRM is also noted.

Given the target DLT rate, the threshold values (γ_L, γ_R) do not depend on the DLT outcomes. As a result, these threshold values can be calculated beforehand, which greatly eases the implementation of the CFO design. In Figure A.3 of Appendix C.2, we show the threshold values when the target DLT rates are 0.2, 0.25, and 0.3, respectively. In general, the threshold values for γ_R are larger than those for γ_L, potentially indicating the CFO design has a more formidable obstacle to escalate than to de-escalate (i.e., a conservative dose escalation scheme). When the treated number of patients increases, the threshold values exhibit a corresponding ascending pattern, signifying the CFO design requires stronger evidence to move the dose as more information is accrued.

2.2 | The TITE-CFO design

The original CFO design requires the DLT outcome can be ascertained immediately after the treatment is administrated, so that the dose level for the next cohort of subjects can be timely decided. In reality, some phase I trials may require a long assessment window, say 90 days, to determine the DLT outcome. In such cases, if the accrual rate is fast, for example two patients per week, the implementation of the CFO design would face the logistic difficulty, due to pending outcomes upon arrival of a new cohort. If we ignore the pending DLT data, the efficiency of the design is compromised because the follow-up data are not fully utilized. To incorporate delayed outcomes, we propose the TITE-CFO design by taking the late-onset outcomes into consideration.

Originally, the CFO design takes the binary outcome 0 or 1 as input to make the decision. However, the binomial likelihood can also accept a decimal value between $(0,1)$ as the input. For the TITE-CFO design, the problem boils down how to impute the pending DLT data with a decimal value between $(0,1)$. Intuitively, at the decision time, if an outcome-pending subject is less likely to experience DLT, the imputed outcome should be small; otherwise, it should be large. As noted in TITE-CRM and TITE-BOIN,[19,22] the follow-up time for the pending data contain rich information about the DLT outcome. For example, when the assessment window is 90 days, an outcome-pending subject whose follow-up time is 80 days is much more likely to have a non-DLT outcome compared with another outcome-pending subject whose follow-up time is only 10 days.

We propose to impute the pending outcome by the expectation of the potential outcome given the follow-up time. At a certain point of a trial, let y denote the binary DLT outcome for a patient treated at dose level k, with $y=1$ if the patient experiences DLT, and $y=0$ if no DLT is observed. The assessment window is τ, and suppose the subject's DLT outcome is still pending. The follow-up time of the outcome-pending patient is t $(t<\tau)$ and the time to DLT is denoted by T which has not been observed yet. Following Yuan et al.[19] and Cheung and Chappell,[22] we assume the time to DLT follows a uniform distribution over $[0,\tau]$. The pending y is imputed as the expectation,

$$\mathbb{E}(y|T>t)=\frac{\Pr(y=1)\Pr(T>t|y=1)}{\Pr(y=1)\Pr(T>t|y=1)+\Pr(y=0)\Pr(T>t|y=0)}=\frac{p_k(1-t/\tau)}{p_k(1-t/\tau)+(1-p_k)},$$

where p_k is the true DLT rate at dose level k. As p_k is unknown, we estimate p_k using the Bayesian posterior mean $\widetilde{p}_k$ based on the observed data. Given the prior $p_k \sim \text{Beta}(\alpha,\beta)$, we obtain the posterior mean as

$$\widetilde{p}_k = \frac{x_{ko} + \alpha}{m_{ko} + \alpha + \beta},$$

where (x_{ko}, m_{ko}) are the number of DLT outcomes and the number of patients who have completed the DLT assessment at dose level k. By default, we choose $\alpha = \phi$ and $\beta = 1 - \phi$,[28] where ϕ is the target DLT rate.

In this way, we can utilize the follow-up time information for dose assignment which increases the efficiency of our design. For example, given a 90-day assessment window and a target DLT rate of 0.3, suppose there is 1 DLT outcome out of 3 ascertained subjects at the current dose. The follow-up times for the two pending subjects at the current dose are 10 and 80 days, respectively. Using our imputation method, the imputed outcomes for the two pending outcomes are 0.264 and 0.043, respectively. Such results are consistent with our intuition, that is, the pending subject with the 10-day follow-up has a larger chance to experience DLT than the one with the 80-day follow-up. The mathematical details on calculating the imputed value are provided in Appendix A.2.

After the imputation, we can calculate two odds ratios $\left(O_C/\overline{O}_L, \overline{O}_C/O_R\right)$ with both the complete and imputed outcomes. The decision is further made by comparing the odds ratios with the threshold values (γ_L, γ_R) as in Table 1. Because the threshold values do not depend on DLT outcomes, the TITE-CFO design uses the same threshold values as the CFO design.

The procedure of the TITE-CFO design is detailed as follows:

1. Treat the first cohort of patients at the lowest or prespecified starting dose.
2. Based on the follow-up times of treated patients, impute the pending data for the current dose level and its two neighboring dose levels.
3. Calculate the odds ratios with both complete and imputed outcomes and make the escalation or de-escalation decision based on Table 1.
4. Repeat steps 2–3 until the prespecified maximum sample size is reached and select the MTD using the isotonic regression.[29]

As a common practice in phase I designs,[17,19,28] we impose an overdose control rule to guard patient safety. When the data suggest the current dose level is over-toxic with high probability, we eliminate the current dose level as well as those higher dose levels. If the lowest dose level is eliminated due to over toxicity, the trial is terminated immediately.

As an extension of CFO, the TITE-CFO design inherits the desirable properties that it does not require calibration of any essential parameters, which eases the trial implementation in practice. At the same time, TITE-CFO shows comparable performance to the model-based TITE-CRM design as demonstrated in our extensive numerical studies.

2.3 | Trial example

We provide a concrete trial example to illustrate the implementation of the TITE-CFO design. To mimic the real situation, we add an accrual suspension rule.[19] Specifically, if at the current dose or its two adjacent doses, over 50% of the patients' outcomes remain pending, the accrual process is temporarily halted until further data become available. Consider a phase I trial with a target DLT rate 0.3 and five dose levels under investigation. The assessment window is 3 months and the accrual rate is two patients per month. There are a total of 10 cohorts with cohort size 3. We assume a uniform distribution for the arrival time of each patient.

As shown in Figure 1, the first cohort of subjects was treated at dose level 1, and there was one DLT outcome out of the three subjects. However, at the time when the first patient of the second cohort arrived, all the three patients in the first cohort were still pending for DLT outcomes. Based on the suspension rule, the trial is suspended until day 117 when two of the three patients in the first cohort completed the follow-up with non-DLT outcomes. After imputation, their outcomes were $(0,0,0.005)$, and the odds ratio was $\overline{O}_C/O_R = 6.39$ which was greater than the threshold value $\gamma_R = 0.09$. By the rules in Table 2, the trial escalated the dose to level 2. At the decision time upon arrival of cohort 3, none of the patients in the second cohort had the full outcome, leading to a suspension until day 234. Up to day 234, the first cohort had complete DLT outcomes $(0,0,1)$, while there was still one pending patient in the second cohort.

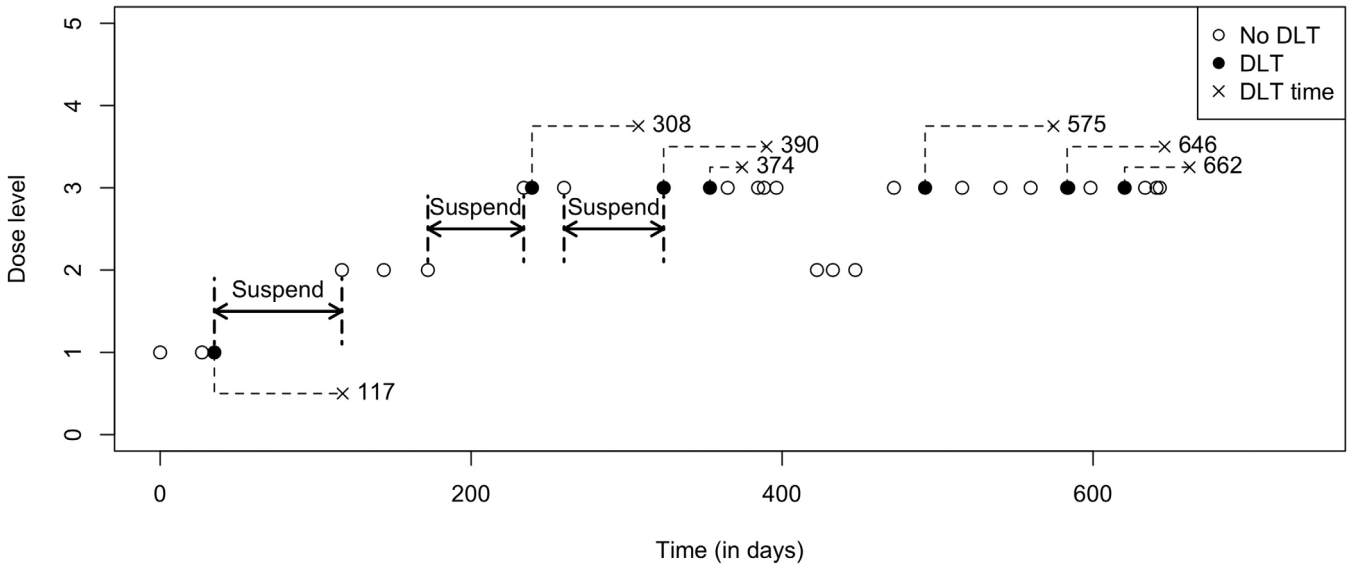

FIGURE 1 A hypothetical phase I clinical trial using the TITE-CFO design with the suspension rule. Patients are treated in a cohort size of 3, and the number on the right of the cross × indicates the time when the DLT event occurred.

The odds ratios at the current dose level 2 were $\left(O_C/\overline{O}_L, \overline{O}_C/O_R\right) = (0.12, 5.95)$ and the threshold values are $(\gamma_L, \gamma_R) = (0.81, 0.09)$. Based on our TITE-CFO decision rules, dose level 3 was selected for cohort 3. The subsequent cohorts 4 and 5 were treated at dose level 3 as well. On day 422 when the first patient of the sixth cohort arrived, four patients had complete outcomes with one DLT while the other four patients' outcomes were still pending. With imputed data, the odds ratios at the current dose level 3 were $\left(O_C/\overline{O}_L, \overline{O}_C/O_R\right) = (0.56, 0.00)$ which were compared with the threshold values $(\gamma_L, \gamma_R) = (0.30, 0.24)$. Based on the rules in Table 1, the dose was de-escalated to level 2. The remaining 4 cohorts were all treated at dose level 3. On day 733, all patients completed their follow-ups and had full DLT data. Dose level 3 was eventually selected as the MTD, at which 6 out of 21 patients suffered DLT with the estimated DLT rate 0.29 and 95% confidence interval (0.12,0.49). Under the TITE-CFO design, the trial took around 733 days to complete. If we adopt the phase I design which requires full DLT assessment for each cohort of patients, it would take 900 days to finish the trial on average. If we do not adopt the accrual suspension rule with 50% data fully observed, the TITE-CFO design could finish the trial even sooner.

3 | SIMULATION STUDIES

Simulation configuration: We conduct extensive numerical studies to assess the operating characteristics of the TITE-CFO design via comparing it with TITE-BOIN[19] and TITE-CRM.[22] For the TITE-BOIN method, we adopt the default parameters suggested in Liu and Yuan.[6] Following Lin and Yin,[17] the TITE-CRM takes the power model where the skeleton is chosen by the model calibration method[30] with a half width of the indifference interval of 0.05. In order to more thoroughly evaluate the impact of imputation on reducing the trial duration and influencing design precision, we incorporate the original CFO design,[28] for comparison. To streamline the comparison, none of the three late-onset designs employ a suspension rule. The detailed settings of the methods in comparison are given in Appendix B.2.

We explore eight representative fixed scenarios with the target DLT rate 0.3. To avoid cherry-picking, the eight fixed scenarios are adopted from the paper of the TITE-BOIN design.[19] The detailed information on the eight fixed scenarios is provided in Appendix B.3. To make our numerical studies more comprehensive, we also compare the three TITE-designs under random scenarios, which are generated following the method of Paoletti et al.[31] as detailed in Appendix B.4. The average probability difference around the target is controlled at 0.05, 0.07, 0.1, and 0.15 respectively. Under both fixed and random scenarios, we consider seven dose levels and a total of 12 cohorts with cohort size 3. For the simulated trials with late-onset toxicity, the DLT assessment window is 3 months and the accrual rate is two patients per month. The patient arrival time follows a uniform distribution. The time to DLT is sampled from a Weibull distribution, with 50% of DLT events occurring in the second half of the assessment window. All simulation studies are

repeated for 5000 times. We also explore the case with the target DLT rate 0.2 under both fixed and random scenarios and the results are displayed in Figures A.1 and A.2 of Appendix C.1.

Performance metrics: We use six performance metrics to evaluate the results for comparing different designs:

1. Percentage of correct selection of the MTD (MTD selection);
2. Percentage of patients allocated to the MTD (MTD allocation);
3. Percentage of selecting a dose above the MTD (overdose selection);
4. Percentage of allocating patients at dose levels above the MTD (overdose allocation);
5. Percentage of the patients suffering DLT (average DLT rate);
6. Average trial duration.

Among the six performance metrics, the first two metrics are the main measurements, reflecting the accuracy and efficiency of a design, for which the higher the better. Metrics 3–5 quantify the safety aspects of a trial, for which the

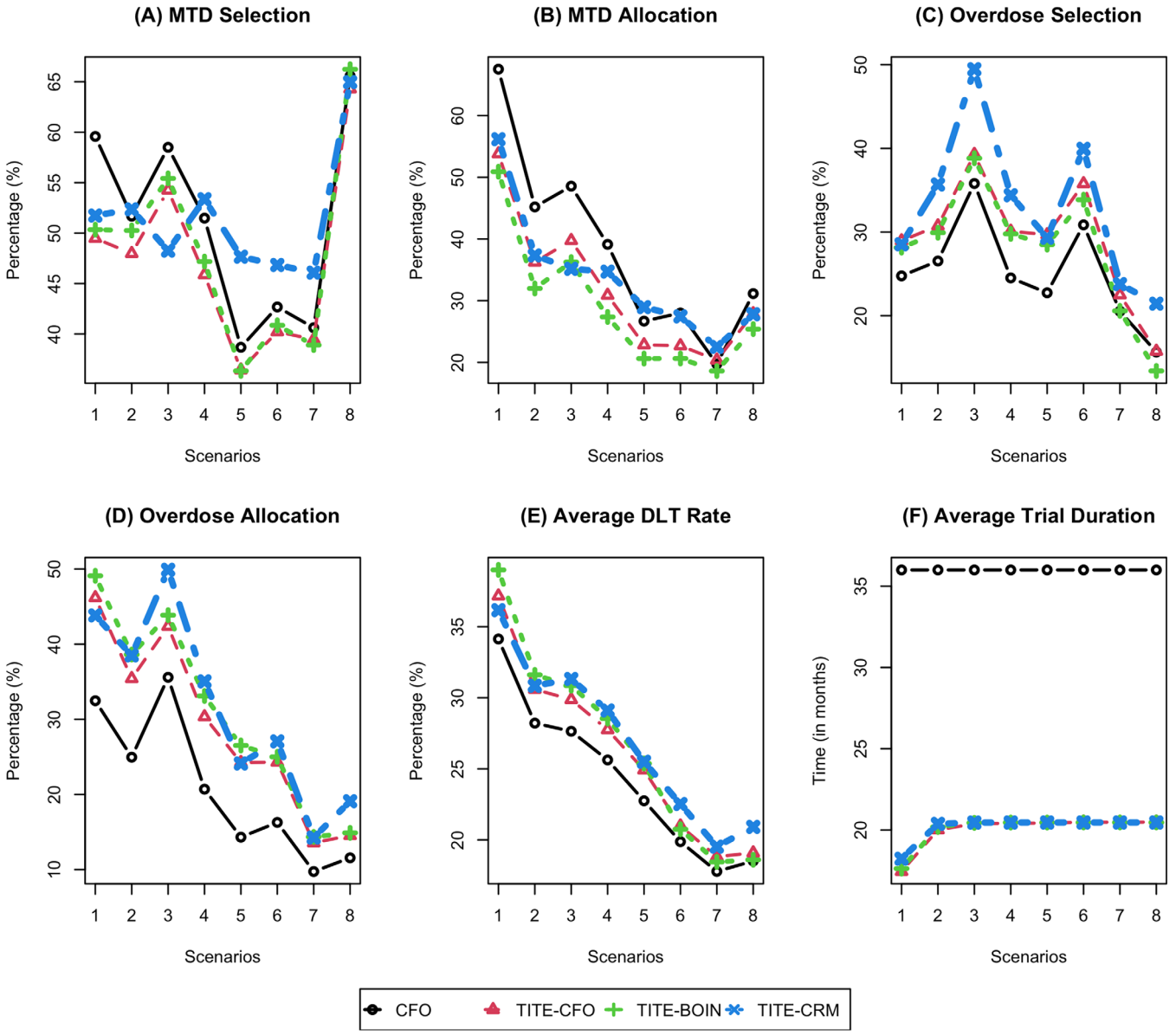

FIGURE 2 Simulation results of the CFO, TITE-CFO, TITE-BOIN and TITE-CRM designs with the target DLT rate 0.3 and sample size 36 under eight fixed scenarios. For MTD selection and allocation, a higher value is preferred. For overdose selection and allocation, the average DLT rate and average trial duration, a lower value is preferred.

lower the better. The last metric is the average duration time of a trial with late-onset outcomes, for which the shorter the better.

Accuracy and efficiency evaluation: The results of CFO, TITE-CFO, TITE-BOIN and TITE-CRM are exhibited in panels A and B of Figures 2 and 3. In both fixed and random scenarios concerning the MTD selection percentage, the TITE-CRM design yields the overall best performance among three TITE-designs, while the TITE-CFO and TITE-BOIN deliver comparable results. However, the TITE-CRM's preeminence is not uniformly evident across all instances. In scenario 3, its performance is considerably inferior to the other two contenders, because the CRM is inherently sensitive to model assumptions due to its model-based nature. With regard to the MTD allocation percentage, the TITE-CRM still performs the best except in scenario 3, and TITE-CFO delivers consistently better results than TITE-BOIN. While the CFO design utilizes richer information for deciding the next dose level, it does not show much superiority in the MTD selection percentage. Nevertheless, it performs the best in terms of the MTD allocation percentage. The results with a target DLT rate of 0.2 in Appendix C.1 demonstrate different phenomena. The three methods perform similarly in

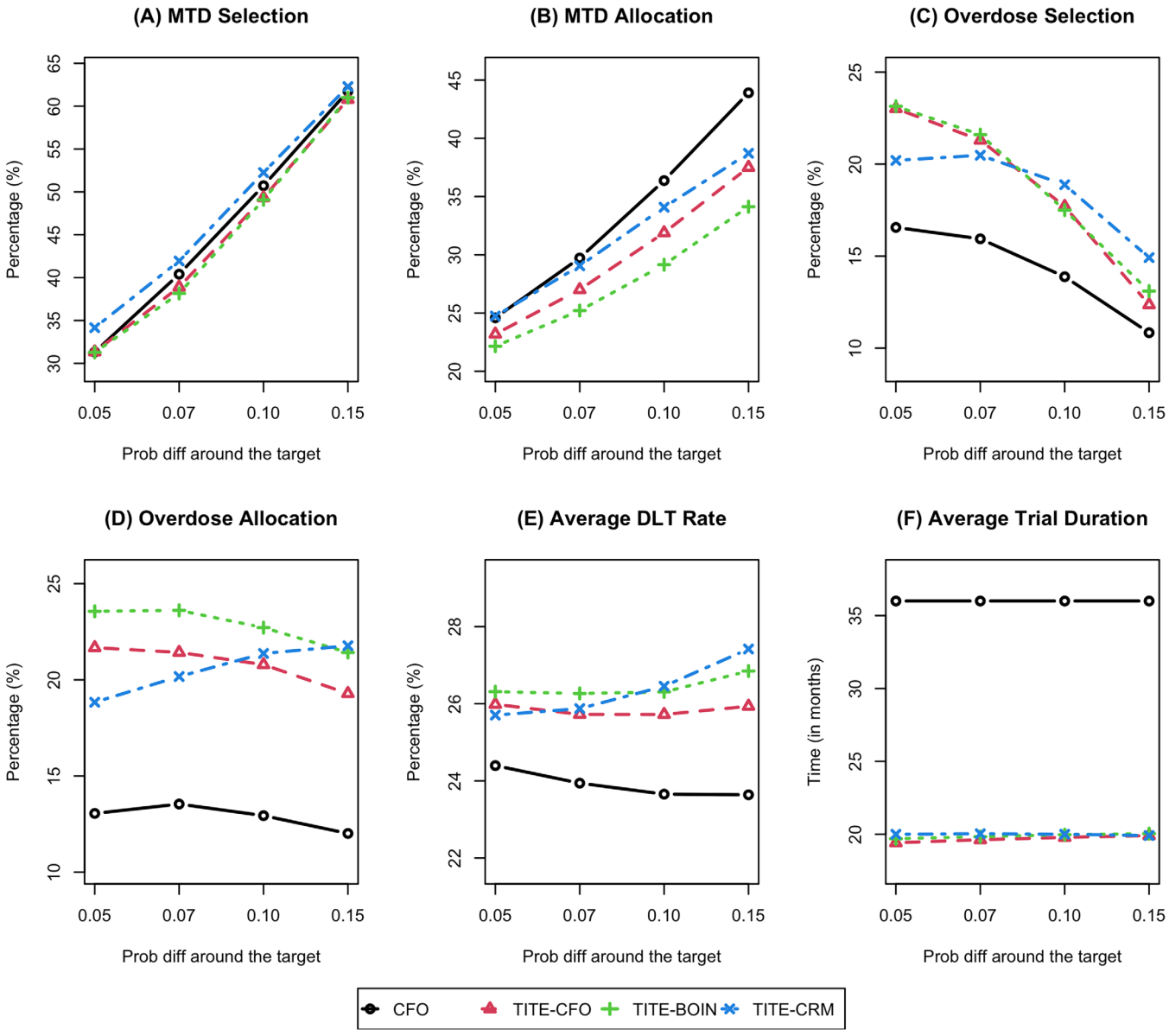

FIGURE 3 Simulation results of the CFO, TITE-CFO, TITE-BOIN and TITE-CRM designs with the target DLT rate 0.3 and sample size 36 under random scenarios when the average probability difference around the target DLT rate varies from 0.05 to 0.15. For MTD selection and allocation, a higher value is preferred. For overdose selection and allocation, the average DLT rate and average trial duration, a lower value is preferred.

terms of the MTD selection percentage, while the TITE-BOIN marginally underperforms the other two TITE-designs for the MTD allocation percentage.

Safety evaluation: Based on panels C–E of Figures 2 and 3, the TITE-CRM design shows under the fixed scenarios notably higher percentages of overdose selection and allocation than the other two TITE-designs, which indicates TITE-CRM is more aggressive in dose escalation. The TITE-CFO and TITE-BOIN perform similarly in the aspect of the safety metrics. Under the random scenarios, TITE-CFO shows consistently superior performances than TITE-BOIN for different configurations. When the average probability difference around the target DLT rate is small, TITE-CRM performs better than TITE-CFO, but when the difference becomes larger, the performance of TITE-CFO surpasses TITE-CRM in the safety aspects. The superiority of the CFO design is unequivocally evident in terms of these safety metrics. Under all configurations, the CFO design is uniformly safer than other three TITE-designs. Under the target DLT rate of 0.2 as shown in Appendix C.1, TITE-BOIN becomes the most aggressive method, while TITE-CFO remains to be conservative and safe.

Duration time: Discernible from panel F of Figures 2 and 3, the three TITE-designs exhibit negligible disparities in the average trial duration. Indeed, for a prototypical design addressing late-onset outcomes, the duration hinges predominantly upon the patient accrual rate and the overdose control rule, yielding only minute differences among the three designs. In contrast, the CFO design, devoid of an imputation mechanism, requires full DLT assessment prior to dose assignment, resulting in an average trial completion time of 36 months. The TITE-CFO, with an approximate duration of 20 months, significantly reduces the time required in comparison to the original CFO design. This phenomenon is similarly observed when the target DLT rate is 0.2, as illustrated in Appendix C.1.

4 | DISCUSSION

Based on the calibration-free and well-performing CFO design,[28] for phase I clinical trials, we further take late-onset outcomes into consideration, by developing the TITE-CFO design. Unlike most of the phase I designs, which use either "local data" at the current dose or the accumulated data at all doses, both CFO and TITE-CFO borrow the game competition idea, which utilizes the information from the current, and its two neighboring dose levels. By comparing the evidence from the three dose levels at a time, the CFO-type designs take a middle-ground approach between using information from one dose level alone or all dose levels.

Because they are model-free and calibration-free, CFO and TITE-CFO are robust and easy-to-use.[28] Through our comprehensive numerical studies, TITE-CFO demonstrates high robustness in comparison to both the rule-based TITE-BOIN and model-based TITE-CRM designs. Due to the model-based nature, the TITE-CRM design is sensitive to the selection of the model skeleton, which barricades its usage in practice. As demonstrated by our simulation studies, the efficiency of the TITE-CFO design is close to that of the TITE-CRM design. This phenomenon is due to the fact that in a phase I trial most of the subjects are assigned to the MTD or its neighboring dose levels. Therefore, ignoring data from the boundary dose levels far away from the MTD does not incur much information loss.

In the TITE-CFO, we assume that the time to DLT follows a uniform distribution, which is also shared by the TITE-CRM and TITE-BOIN designs. Like TITE-CRM[22] and TITE-BOIN,[19] the TITE-CFO is insensitive to the choice of the time-to-DLT distribution. In practice, the uniform distribution can be used as the default setting for general purposes. If there is some prior information on the time-to-DLT distribution, the TITE-CFO can be adapted accordingly with minor modifications on the imputation method.

Our discussions focus on the phase I trials with a single agent by considering the toxicity only. As a possible future work, the CFO design can be extended to handle the drug combination trials by comparing the evidence from multiple drugs. A further avenue of exploration entails its application to the phase I/II trial setting,[1] where both toxicity and efficacy outcomes are late-onset. Alternatively, in a more intricate scenario, the two outcomes are competing risk, that is, the occurrence of the first outcome terminates the other one.[32] To facilitate the practical use of the CFO and TITE-CFO designs, we provide the R code at https://github.com/JINhuaqing/CFOs.

ACKNOWLEDGMENTS

We thank the two referees, Associate Editor and Editor for their many constructive and insightful comments that have led to significant improvements in the article.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data derived from public domain resources.

ORCID

Huaqing Jin https://orcid.org/0000-0002-1932-1336
Guosheng Yin https://orcid.org/0000-0003-3276-1392

REFERENCES

1. Yin G. *Clinical Trial Design: Bayesian and Frequentist Adaptive Methods*. 876. John Wiley & Sons; 2012.
2. Iasonos A, O'Quigley J. Adaptive dose-finding studies: a review of model-guided phase I clinical trials. *J Clin Oncol*. 2014;32(23):2505-2511.
3. Storer BE. Design and analysis of phase I clinical trials. *Biometrics*. 1989;45(3):925-937.
4. Ivanova A, Flournoy N, Chung Y. Cumulative cohort design for dose-finding. *J Stat Plan Infer*. 2007;137(7):2316-2327.
5. Ji Y, Liu P, Li Y, Nebiyou BB. A modified toxicity probability interval method for dose-finding trials. *Clin Trials*. 2010;7(6):653-663.
6. Liu S, Yuan Y. Bayesian optimal interval designs for phase I clinical trials. *J R Stat Soc Ser C Appl Stat*. 2015;64(3):507-523.
7. Yan F, Mandrekar SJ, Yuan Y. Keyboard: a novel Bayesian toxicity probability interval design for phase I clinical trials. *Clin Cancer Res*. 2017;23(15):3994-4003.
8. Johnson VE. Uniformly most powerful Bayesian tests. *Ann Stat*. 2013;41(4):1716-1741.
9. Lin R, Yin G. Uniformly most powerful Bayesian interval design for phase I dose-finding trials. *Pharm Stat*. 2018;17(6):710-724.
10. O'Quigley J, Pepe M, Fisher L. Continual reassessment method: a practical design for phase 1 clinical trials in cancer. *Biometrics*. 1990;46(1):33-48.
11. O'Quigley J, Shen LZ. Continual reassessment method: a likelihood approach. *Biometrics*. 1996;52(2):673-684.
12. Braun TM. The bivariate continual reassessment method: extending the CRM to phase I trials of two competing outcomes. *Control Clin Trials*. 2002;23(3):240-256.
13. Yin G, Yuan Y. Bayesian model averaging continual reassessment method in phase I clinical trials. *J Am Stat Assoc*. 2009;104(487):954-968.
14. Yuan Z, Chappell R, Bailey H. The continual reassessment method for multiple toxicity grades: a Bayesian quasi-likelihood approach. *Biometrics*. 2007;63(1):173-179.
15. Wages NA, Conaway MR, O'Quigley J. Continual reassessment method for partial ordering. *Biometrics*. 2011;67(4):1555-1563.
16. Babb J, Rogatko A, Zacks S. Cancer phase I clinical trials: efficient dose escalation with overdose control. *Stat Med*. 1998;17(10):1103-1120.
17. Lin R, Yin G. Nonparametric overdose control with late-onset toxicity in phase I clinical trials. *Biostatistics*. 2017;18(1):180-194.
18. Jin H, Du W, Yin G. Approximate Bayesian computation design for phase I clinical trials. *Stat Methods Med Res*. 2022;31(12):2310-2322.
19. Yuan Y, Lin R, Li D, Nie L, Warren KE. Time-to-event Bayesian optimal interval design to accelerate phase I trials. *Clin Cancer Res*. 2018;24(20):4921-4930.
20. Postel-Vinay S, Gomez-Roca C, Molife LR, et al. Phase I trials of molecularly targeted agents: should we pay more attention to late toxicities? *J Clin Oncol*. 2011;29(13):1728-1735.
21. June CH, Warshauer JT, Bluestone JA. Is autoimmunity the Achilles' heel of cancer immunotherapy? *Nat Med*. 2017;23(5):540-547.
22. Cheung YK, Chappell R. Sequential designs for phase I clinical trials with late-onset toxicities. *Biometrics*. 2000;56(4):1177-1182.
23. Mauguen A, Deley M, Zohar S. Dose-finding approach for dose escalation with overdose control considering incomplete observations. *Stat Med*. 2011;30(13):1584-1594.
24. Yin G, Zheng S, Xu J. Fractional dose-finding methods with late-onset toxicity in phase I clinical trials. *J Biopharm Stat*. 2013;23(4):856-870.
25. Chapple AG, Thall PF. Subgroup-specific dose finding in phase I clinical trials based on time to toxicity allowing adaptive subgroup combination. *Pharm Stat*. 2018;17(6):734-749.
26. Lin R, Yuan Y. Time-to-event model-assisted designs for dose-finding trials with delayed toxicity. *Biostatistics*. 2020;21(4):807-824.
27. Takeda K, Xia Q, Liu S, Rong A. TITE-gBOIN: time-to-event Bayesian optimal interval design to accelerate dose-finding accounting for toxicity grades. *Pharm Stat*. 2022;21(2):496-506.
28. Jin H, Yin G. CFO: calibration-free odds design for phase I/II clinical trials. *Stat Methods Med Res*. 2022: 09622802221079353;31:1066.
29. Bril G, Dykstra R, Pillers C, Robertson T. Algorithm AS 206: isotonic regression in two independent variables. *J R Stat Soc Ser C Appl Stat*. 1984;33(3):352-357.
30. Lee SM, Cheung YK. Model calibration in the continual reassessment method. *Clin Trials*. 2009;6(3):227-238.
31. Paoletti X, O'Quigley J, Maccario J. Design efficiency in dose finding studies. *Comput Stat Data Anal*. 2004;45(2):197-214.
32. Zhang Y, Cao S, Zhang C, Jin IH, Zang Y. A Bayesian adaptive phase I/II clinical trial design with late-onset competing risk outcomes. *Biometrics*. 2021;77(3):796-808.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Jin H, Yin G. Time-to-event calibration-free odds design: A robust efficient design for phase I trials with late-onset outcomes. *Pharmaceutical Statistics*. 2023;22(5):773-783. doi:10.1002/pst.2304