Supporting information for "CFO: Calibration-Free Odds Design for Phase I/II Clinical Trials" by Huaqing Jin and Guosheng Yin

A Simulation Details

A.1 Compared designs for phase I trials

The detailed settings of the BOIN (Liu and Yuan, 2015) and CRM (O'Quigley et al., 1990) used in the simulation studies are described as follows.

- **BOIN:** Following Lin and Yin (2017b) and Liu and Yuan (2015), we set $\phi_1 = 0.6\phi$ and $\phi_2 = 1.4\phi$. We eliminate dose level k and all dose levels above from the trial if $\Pr(p_k > \phi | x_k, m_k \geq 3) > 0.95$. If the posterior probability of the first dose level satisfies $\Pr(p_1 > \phi | x_1, m_1 \geq 3) > 0.95$, then we terminate the entire trial for safety.

- **CRM:** We adopt the power model $p_j = \pi_j^{\exp(\alpha)}$ with the model skeleton selected by the method of Lee and Cheung (2009). We choose an initial guess of the MTD at dose level $\lceil K/2 \rceil$ while the halfwidth of the indifference interval is set as 0.05 following Lin and Yin (2017b, 2018). The early stopping rule terminates the trial if $\Pr(p_1 > \phi | \text{data}) > 0.95$.

A.2 Compared deigns for phase I/II trials

The detailed settings of the STEIN (Lin and Yin, 2017a), MADA (Xu et al., 2016) and WT (Wages and Tait, 2015) methods used in the simulation studies are described as follows.

- **STEIN:** Following Section 2.4 of Lin and Yin (2017a), we use $\phi_1 = 0.75\phi$ and $\phi_2 = 1.25\phi$. For the efficacy rates, we set $\psi_1 = \psi = 0.3$ and $\psi_2 = 0.65$. For a fair comparison, the utility function is defined as $U(p_k, q_k) = q_k - 100I(p_k > \phi)$ for $k = 1, \ldots, K$, where $I(\cdot)$ is the indicator function.

- **MADA:** We use $(\theta_T, \theta_L, \theta_U) = (0.95, 0.4, 0.7)$ and there are 24 subjects for the first stage (Xu et al., 2016).

- **WT:** Following Wages and Tait (2015), the sample size for the adaptive randomization phase is 25% of the total sample size. The prior model skeletons for the dose–efficacy relationship adopt the skeletons in Section 4.1 of Wages and Tait (2015). The prior skeleton for the dose–toxicity relationship follows the CRM setting in Section A.1.

B Random Scenario Generation

B.1 Scheme for phase I trials

We generate random scenarios to assess the performance of the phase I designs in Sections 3.1 and 3.2 following Paoletti et al. (2004). Specifically, the procedure is detailed as follows.

1. We randomly select, with equal probabilities, one of the K dose levels as the MTD and denote that dose level as k_{MTD}.

2. Let Φ be the cumulative density function (CDF) of the standard normal distribution. The probability of the MTD is $p_{k_{\mathrm{MTD}}} = \Phi(\epsilon_{\mathrm{MTD}})$ with $\epsilon_{\mathrm{MTD}} \sim N(\Phi^{-1}(\phi), \sigma_0^2)$, where ϕ is the target toxicity probability.

3. For $\{p_k\}_{k=1}^{k_{\mathrm{MTD}}-1}$, we generate

$$p_{k-1} = \Phi\left[\Phi^{-1}(p_k) - \left\{\Phi^{-1}(p_k) - \Phi^{-1}(2\phi - p_k)\right\} I\left\{\Phi^{-1}(p_k) > \Phi^{-1}(\phi)\right\} - \epsilon_{k-1}^2\right],$$

where $I(\cdot)$ is the indicator function and $\epsilon_{k-1} \sim N(\mu_1, \sigma_1^2)$.

4. For $\{p_k\}_{k=k_{\mathrm{MTD}}+1}^{K}$, we generate

$$p_{k+1} = \Phi\left[\Phi^{-1}(p_k) + \left\{\Phi^{-1}(2\phi - p_k) - \Phi^{-1}(p_k)\right\} I\left\{\Phi^{-1}(p_k) < \Phi^{-1}(\phi)\right\} + \epsilon_{k+1}^2\right],$$

where $\epsilon_{k+1} \sim N(\mu_2, \sigma_2^2)$.

Following Liu and Yuan (2015), we choose $\sigma_0 = 0.05$ and $\sigma_1 = \sigma_2 = 0.35$, and tune the parameters $\mu_1 = \mu_2$ to achieve the desirable Δ, i.e., the average probability difference around the target.

B.2 Scheme for phase I/II trials

As there is no monotone assumption on $\{q_k\}_{k=1}^K$, we divide the dose–efficacy curve into two types, the plateau-shape curve and the umbrella-shape curve. The monotone increasing curve is a special case of the plateau-shape or umbrella-shape curve when the highest efficacy rate is achieved at dose level K.

The detailed procedure to generate the scenarios for phase I/II trials is described as follows.

1. We randomly select, with equal probabilities, one of the K dose levels as the OBD and denote that dose level as k_{OBD}.

2. We randomly select, with equal probabilities, one dose level as the MTD from $\{k_{\mathrm{OBD}}, k_{\mathrm{OBD}} + 1, \ldots, K\}$ and denote that dose level as k_{MTD}.

3. Given k_{MTD}, we generate the DLT rates $\{p_k\}_{k=1}^K$ following the procedure in Section B.1.

4. Let ψ and ψ_U be the lowest acceptable efficacy rate and the upper bound of the efficacy rate, we generate $q_{\mathrm{OBD}} \sim \mathrm{Uniform}(\psi, \psi_U)$. We choose $\psi_U = 0.7$ throughout our experiments.

5. To generate plateau-shape curves, we first generate $k_{\mathrm{OBD}} - 1$ values from $\mathrm{Uniform}(0, q_{\mathrm{OBD}})$, and then sort them in an ascending order to obtain $\{q_k\}_{k=1}^{k_{\mathrm{OBD}}-1}$. For $k = k_{\mathrm{OBD}} + 1, \ldots, K$, we let $q_k = q_{\mathrm{OBD}}$.

6. To generate umbrella-shape curves, we first generate $k_{\mathrm{OBD}} - 1$ values from $\mathrm{Uniform}(0, q_{\mathrm{OBD}})$, and then sort them in an ascending order to obtain $\{q_k\}_{k=1}^{k_{\mathrm{OBD}}-1}$. Similarly, we generate

3

$K - k_{\mathrm{OBD}}$ values from Uniform$(0, q_{\mathrm{OBD}})$, and then sort them in a descending order to obtain $\{q_k\}_{k=k_{\mathrm{OBD}}+1}^{K}$.

C Additional simulation results for toxicity evaluation

C.1 Toxicity evaluation by ANOVA

As determination of the MTD is an essential part of the CFO design, we conduct extensive simulation studies in the context of identifying the MTD. For comparison, we also implement the CRM (O'Quigley et al., 1990) and Bayesian optimal interval design (BOIN) (Liu and Yuan, 2015) to assess their operating characteristics for monitoring toxicity only. The detailed settings of the compared designs are given in Section A.1.

We first investigate the influential factors that affect the result of the dose-finding trial in terms of the percentage of MTD selection via the analysis of variance (ANOVA) method used by Cangul et al. (2009). To avoid cherry-picking cases, we randomly generate dose–toxicity scenarios following the approach of Paoletti et al. (2004). The major factors affecting the result of the phase I trial are listed in Table A.1, which shows a total of $3 \times 2 \times 4 \times 4 \times 4 = 384$ factorial settings, and under each cell we compare the BOIN, CFO and CRM designs via 1000 randomly generated scenarios.

After obtaining the percentage of MTD selection for each cell in the factorial design, we perform ANOVA with regard to these percentages using the simulation factors including all the pairwise interactions in Table A.1. The factors are ordered by the mean squared error (MSE) in the ANOVA, and the three most influential factors on the MTD selection are the average probability difference around the target ϕ, the sample size, and the number of dose levels K. These three factors account for 89.7% of the MTD selection percentage variance. As expected, the statistical design has rather minor influence on the MTD selection. In practice, a real trial would be conducted only once and thus it is of paramount importance to apply the most suitable design to pin down the MTD accurately.

The percentage of MTD selection and the percentage of patients treated at the MTD across the three main factors are reported in Table A.2. We also present the percentage of overdose selection and the percentage of patients treated at the over-toxic doses across the same factors in Table A.3. The four metrics under the CFO, BOIN and CRM designs vary dramatically when the three dominant factors change. The CRM has an overall highest percentage of MTD selection in comparison with the other two methods. It is because the CRM design is a model-based approach that can borrow information across all doses, while the other two methods can only utilize local information for dose escalation. The CFO design shows a slightly higher percentage of MTD selection compared with BOIN, while in term of the percentage of the MTD allocation, the CFO and CRM designs have similar performances and both are better than BOIN. For the two safety measures, CFO yields the best performance and CRM is clearly worse than the other two methods. Such a trade-off is common in the dose-finding task, when a method tends to be more efficient by exploring more untried and potentially risky doses, it would typically sacrifice on the safety aspect. As a summary based on the results, the CFO design strikes a balance between efficiency and safety.

Table A.1: Simulation factors affecting the dose-finding performance of the phase I trial and the results of ANOVA in terms of the percentage of MTD selection. The ANOVA also includes all the pairwise interactions between the simulation factors.

Factor	Levels of factor	DF	MSE
Average probability difference around ϕ	$\{0.05, 0.07, 0.10, 0.15\}$	3	3.558
Sample size	$\{21, 30, 48, 60\}$	3	0.467
Number of dose levels K	$\{3, 5, 7, 9\}$	3	0.203
Statistical design	$\{$BOIN, CFO, CRM$\}$	2	0.122
Target toxicity probability ϕ	$\{0.25, 0.30, 0.33\}$	2	0.076
Cohort size	$\{1, 3\}$	1	0.007

DF: degree of freedom; MSE: mean squared error

Table A.2: The percentage of MTD selection (the percentage of patients treated at the MTD) under the CFO, BOIN and CRM designs across the three dominant factors: the probability difference around ϕ, sample size, and number of doses.

Design	Factor Level				
Prob diff (ϕ)	0.05	0.07	0.10	0.15	Average
CFO	36.0 (31.6)	43.1 (36.6)	51.2 (42.7)	62.1 (51.0)	48.1 (40.5)
BOIN	36.0 (31.1)	43.1 (35.6)	50.7 (40.6)	61.0 (47.4)	47.7 (38.7)
CRM	38.2 (32.8)	46.1 (38.0)	54.7 (43.9)	64.9 (51.3)	51.0 (41.5)
Sample size	21	30	48	60	
CFO	43.2 (34.6)	47.1 (38.5)	50.4 (43.4)	51.6 (45.6)	48.1 (40.5)
BOIN	43.0 (33.3)	47.1 (36.7)	49.9 (41.5)	50.6 (43.2)	47.7 (38.7)
CRM	45.0 (34.7)	48.9 (38.9)	54.1 (44.9)	55.8 (47.4)	51.0 (41.5)
Number of doses	3	5	7	9	
CFO	52.4 (50.4)	48.6 (41.6)	46.4 (36.6)	45.0 (33.4)	48.1 (40.5)
BOIN	51.2 (48.8)	48.3 (39.8)	46.6 (34.8)	44.6 (31.3)	47.7 (38.7)
CRM	53.5 (51.2)	51.4 (42.6)	50.0 (37.8)	48.9 (34.4)	51.0 (41.5)

Table A.3: The percentage of overdose selection (the percentage of patients treated at over-toxic doses) under the CFO, BOIN and CRM designs across the three factors: the probability difference around ϕ, sample size, and number of doses.

Design	Factor Level				
Prob diff (ϕ)	0.05	0.07	0.10	0.15	Average
CFO	19.3 (19.0)	17.6 (18.2)	14.7 (16.5)	10.4 (13.7)	15.5 (16.8)
BOIN	20.3 (20.2)	17.9 (19.0)	15.1 (17.3)	10.1 (14.4)	15.8 (17.7)
CRM	21.7 (21.3)	20.4 (20.6)	18.4 (19.4)	14.5 (16.9)	18.7 (19.6)
Sample size	21	30	48	60	
CFO	16.8 (16.4)	16.3 (17.0)	15.1 (17.3)	13.8 (16.7)	15.5 (16.8)
BOIN	17.3 (17.4)	17.0 (18.2)	15.1 (17.9)	14.0 (17.5)	15.8 (17.7)
CRM	22.3 (19.2)	19.0 (20.0)	16.9 (19.7)	16.1 (19.4)	18.7 (19.6)
Number of doses	3	5	7	9	
CFO	15.5 (18.4)	15.8 (17.6)	15.6 (16.2)	15.2 (15.2)	15.5 (16.8)
BOIN	15.3 (19.0)	16.1 (18.7)	16.0 (17.2)	15.9 (16.2)	15.8 (17.7)
CRM	17.1 (19.9)	18.7 (20.3)	19.3 (19.5)	19.9 (18.5)	18.7 (19.6)

C.2 CRM results under the optimal halfwidth

In our simulation studies, we choose the halfwidth of the CRM design as 0.05 consistently following Lin and Yin (2017b, 2018). In fact, we also try to select the optimal halfwidth for each setting based on the algorithm of Lee and Cheung (2009). However, as shown in Table A.4, the optimal halfwidth does not yield better results compared with those by fixing the halfwidth at 0.05. Thus, we fix the halfwidth at 0.05 throughout the simulation studies for simplicity and consistency.

Table A.4: The overall results of the CRM when the halfwidth is fixed as 0.05 and when the optimal halfwidth is selected by the method of Lee and Cheung (2009) in ANOVA.

	MTD Selection	MTD Allocation	Overdose Selection	Overdose Allocation
CRM (optimal)	50.4	40.8	20.4	21.1
CRM (0.05)	51.0	41.5	18.7	19.6

C.3 Implementation of CFO at the beginning of a trial

In Table A.5, we show the design statistics at the beginning of a trial when the cohort size is 1 and $\phi = 0.3$. If the cohort size is 3, the table is very long to enumerate all the possible outcomes.

Table A.5: The design statistics when CFO is implemented at the beginning of a trial.

(x_L, x_C, x_R)	$(O_C/\bar{O}_L, \bar{O}_C/O_R)$	$\mathcal{A}$	(y_L, y_C, y_R)	$(\Pr(q_k = \max_{j=\mathcal{A}}\{q_j\}))_k$
$(m_L, m_C, m_R) = (\text{NA}, 1, 0)$, $(\gamma_L, \gamma_R) = (\text{NA}, 0.003)$				
$(\text{NA}, 0, 0)$	$(\text{NA}, 1.546)$	$\{1, 2\}$	$(\text{NA}, 0, 0)$	$(0.296, 0.704)$
			$(\text{NA}, 1, 0)$	$(0.710, 0.290)$
$(\text{NA}, 1, 0)$	$(\text{NA}, 0.003)$	$\{1\}$	$(\text{NA}, 0, 0)$	(1)
			$(\text{NA}, 1, 0)$	(1)
$(m_L, m_C, m_R) = (1, 1, 0)$, $(\gamma_L, \gamma_R) = (0.048, 0.003)$				
$(0, 0, 0)$	$(0.048, 1.546)$	$\{1, 2, 3\}$	$(0, 0, 0)$	$(0.217, 0.202, 0.581)$
			$(0, 1, 0)$	$(0.056, 0.670, 0.274)$
$(0, 1, 0)$	$(6.244, 0.003)$	$\{1\}$	$(0, 0, 0)$	(1)
			$(0, 1, 0)$	(1)

References

M. Cangul, Y. R. Chretien, R. Gutman, and D. B. Rubin. Testing treatment effects in unconfounded studies under model misspecification: Logistic regression, discretization, and their combination. *Statistics in medicine*, 28(20):2531–2551, 2009.

S. M. Lee and Y. K. Cheung. Model calibration in the continual reassessment method. *Clinical Trials*, 6(3):227–238, 2009.

R. Lin and G. Yin. STEIN: A simple toxicity and efficacy interval design for seamless phase I/II clinical trials. *Statistics in Medicine*, 36(26):4106–4120, 2017a.

R. Lin and G. Yin. Nonparametric overdose control with late-onset toxicity in phase I clinical trials. *Biostatistics*, 18(1):180–194, 2017b.

R. Lin and G. Yin. Uniformly most powerful Bayesian interval design for phase I dose-finding trials. *Pharmaceutical Statistics*, 17(6):710–724, 2018.

S. Liu and Y. Yuan. Bayesian optimal interval designs for phase I clinical trials. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 64(3):507–523, 2015.

J. O'Quigley, M. Pepe, and L. Fisher. Continual reassessment method: a practical design for phase 1 clinical trials in cancer. *Biometrics*, pages 33–48, 1990.

X. Paoletti, J. O'Quigley, and J. Maccario. Design efficiency in dose finding studies. *Computational Statistics & Data Analysis*, 45(2):197–214, 2004.

N. A. Wages and C. Tait. Seamless phase I/II adaptive design for oncology trials of molecularly targeted agents. *Journal of biopharmaceutical statistics*, 25(5):903–920, 2015.

J. Xu, G. Yin, D. Ohlssen, and F. Bretz. Bayesian two-stage dose finding for cytostatic agents via model adaptation. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 65(3):465–482, 2016.