Appendix for "Time-to-event Bayesian Optimal Interval Design to Accelerate Phase I Trials" by Yuan, Lin, Li, Nie and Warren.

A. Statistical Principles of the TITE-BOIN Design

Let y_i denote the binary DLT outcome, with y_i =1 indicating that the *i*th patient experiences DLT, and y_i = 0 indicating that the *i*th patient does not experience DLT. Suppose that at a certain moment in the trial, a total of n patients have been enrolled at the current dose, among which r patients have completed the DLT assessment and their DLT data y_i are known, and $c = n - r$ patients have not completed the DLT assessment and their DLT data y_i are pending. Let O denote the set of patients whose DLT data are known (i.e., observed), and M denote the set of patients whose DLT data are pending (i.e., missing). Let p denote the true DLT rate of the current dose level. Under the BOIN design, the dose escalation and de-escalation decision is determined by comparing the estimate of p, given by

$$\hat{p} = \frac{\sum_{i\in O} y_i + \sum_{i\in M} y_i}{n}, \qquad (1)$$

with a pair of fixed, predetermined escalation and de-escalation boundaries, λ_e and λ_d. If $\hat{p} \le \lambda_e$, escalate the dose to the next higher level; if $\hat{p} \ge \lambda_d$, de-escalate the dose to the next lower level; otherwise stay at the current dose. Given the target DLT rate of ϕ, and assuming the noninformative prior that *a priori* the current dose is equally likely to be below, equal to or above the MTD, the optimal escalation and de-escalation boundaries λ_e and λ_d that minimize the incorrect decision of dose escalation and de-escalation are given by

$$\lambda_e = \log\frac{1-\phi_1}{1-\phi} \Big/ \log\frac{\phi(1-\phi_1)}{\phi_1(1-\phi)}$$

$$\lambda_d = \log\frac{1-\phi}{1-\phi_2} \Big/ \log\frac{\phi_2(1-\phi)}{\phi(1-\phi_2)}$$

where ϕ_1 is the highest toxicity probability that is deemed subtherapeutic (i.e., below the MTD) such that dose escalation should be made, and ϕ_2 is the lowest toxicity probability that is deemed overly toxic such that dose de-escalation is required. Liu and Yuan (2016) recommend default values $\phi_1 = 0.6\phi$ and $\phi_2 = 1.4\phi$ for general use. Because of using the non-informative prior, the decision rule of the BOIN has an appearance of the classical frequentist design and only involves the observed DLT rate. Actually, the BOIN can also be derived as a frequentist design, and Liu and Yuan (2016) show that its decision rule is equivalent to using the likelihood ratio test to determine dose escalation/de-escalation. Having both Bayesian and frequentist interpretations is a strength of the BOIN, making it appealing to wider audiences.

In the presence of late-onset toxicity or fast accrual, the difficulty is that y_i is unknown (or missing) for patients whose DLT data are pending, i.e., $i \in M$. We handle this missing data problem using imputation. Let T denote the pre-specified DLT assessment window, X_i denote the time to DLT, and t_i ($< T$) denote the follow-up time for the patient whose DLT data are pending, i.e., $i \in M$. Assuming that the

time to DLT follows a uniform distribution over $[0, T]$, the expected value of y_i, $i \in M$, for a patient treated at the current dose j with follow-up time t_i is given by

$$
\begin{aligned}
E(y_i | X_i > t_i) &= \Pr(y_i = 1 \mid X_i > t_i) \\
&= \frac{\Pr(y_i = 1)\Pr(X_i > t_i \mid y_i = 1)}{\Pr(y_i = 1)\Pr(X_i > t_i \mid y_i = 1) + \Pr(y_i = 0)\Pr(X_i > t_i \mid y_i = 0)} \\
&= \frac{p\left(1 - \frac{t_i}{T}\right)}{p\left(1 - \frac{t_i}{T}\right) + (1 - p)} \\
&\approx \frac{p\left(1 - \frac{t_i}{T}\right)}{(1 - p)}
\end{aligned} \tag{2}
$$

A major concern for late-onset toxicity is that it may lead to aggressive dose escalation, thus we take a conservative approach by adopting the approximation as the last equation. This approximation slightly inflates the expected value of y_i to reduce the chance of aggressive dose escalation, but is sufficiently accurate to yield superior operating characteristics (see simulation study) because in practice p and thus $p\left(1 - \frac{t_i}{T}\right)$ are often small, compared to $1 - p$. The assumption that the time to DLT is uniformly distributed over $[0, T]$ seems strong and restrictive, but our numerical study shows that the performance of the design is remarkably robust to the violation of this assumption. This result is consistent with that reported by Cheung and Chappell (2000) for the TITE-CRM, which also uses the assumption that the time to DLT follows a uniform distribution. In Section A2, we describe an approach that does not assume the uniform distribution.

In equation (1), replacing the unknown values of y_i, $i \in M$, with its expected value $\hat{y}_i$, and let $s = \sum_{i \in O} y_i$ denote the number of patients experienced DLT, we have

$$
\begin{aligned}
\hat{p} &= \frac{\sum_{i \in O} y_i + \sum_{i \in M} \hat{y}_i}{n} \\
&= \frac{s + \sum_{i \in M} \frac{p\left(1 - \frac{t_i}{T}\right)}{(1 - p)}}{n} \\
&= \frac{s + \frac{p}{1 - p}\left(c - \frac{1}{T}\sum_{i \in M} t_i\right)}{n} \\
&= \frac{s + \frac{p}{1 - p}(c - \mathrm{STFT})}{n}
\end{aligned} \tag{3}
$$

where STFT $= \sum_{i \in M} t_i / T$ is the standardized total follow-up time (STFT) for pending patients at the current dose, and s is the number of patients who experienced DLT among the r patients whose DLT data are observed at the current dose. In the statistical literature, the above approach is known as single mean imputation (Little and Rubin, 2012). One drawback of single mean imputation is that although it provides an unbiased and consistent point estimate, the resulting variance estimate is biased because of ignoring the imputation uncertainty. In our case, this is not a concern as the decision rules of the BOIN only rely on

the point estimate of p. After the single mean imputation, $\hat{p}$ is a valid, unbiased point estimate of p (Little and Rubin, 2012).

Equation (3) involves an unknown value p. We replaced it with its Bayesian posterior mean estimate $\tilde{p}$ based on the observed data. Assuming the beta-binomial model with the prior $p \sim Beta(\alpha, \beta)$, the Bayesian posterior mean estimate is given by $\tilde{p} = (s + \alpha)/(r + \alpha + \beta)$. We use a vague beta prior with $\alpha = 0.5\phi$, and $\beta = 1 - \alpha$ such that the prior corresponds to an effective prior sample size of 1 with prior mean $\phi/2$.

One important property of equation (3) is that $\hat{p}$ is a monotonically decreasing function of STFT. This makes it possible to tabulate the decision rule before trial conduct. After some algebra and imposing the long-memory coherence property to be consistent with the BOIN, it can be shown that $\hat{p}$ crosses the BOIN escalation boundary λ_e (i.e., $\hat{p}_j \leq \lambda_e$) when STFT $\geq \pi_e$; and $\hat{p}$ crosses the BOIN de-escalation boundary λ_d (i.e., $\hat{p}_j \geq \lambda_d$) when STFT $\leq \pi_d$, where π_e and π_d are given by

$$\pi_e = \left\{c - \frac{1-\tilde{p}}{\tilde{p}}(n\lambda_e - s)\right\} I\left\{\frac{s}{n} < \phi\right\} + \infty I\left\{\frac{s}{n} \geq \phi\right\} \qquad (4)$$

$$\pi_d = \left\{c - \frac{1-\tilde{p}}{\tilde{p}}(n\lambda_d - s)\right\} I\left\{\frac{s}{n} > \phi\right\} - \infty I\left\{\frac{s}{n} \leq \phi\right\} \qquad (5).$$

In other words, if the STFT for c pending patients $\geq \pi_e$, escalate the dose to the next higher level; and if the STFT for c pending patients $\leq \pi_d$, de-escalate the dose to the next lower level; otherwise, stay at the current dose level. The TITE-BOIN decision tables, such as Table 1 (in the main text; with a target DLT rate of 0.2) and Table S1 (with a target DLT rate of 0.3), are generated by this procedure. In equations (4) and (5), the indicator function $I\left\{\frac{s}{n} < \phi\right\}$ and $I\left\{\frac{s}{n} \geq \phi\right\}$ are imposed to ensure that the TITE-BOIN has the similar long-memory coherence property as the BOIN (i.e., the dose is never escalated/deescalated if the observed DLT rate *s/n* is greater/smaller than the target DLT rate ϕ).

One remarkable feature of the TITE-BOIN is that its decision rule is invariant to T. This means that given a target DLT rate, the same decision table, such as Table 1, can be used to guide dose escalation and de-escalation, regardless of the length of the assessment window. For example, Table 1 can be used for any trial with the target DLT rate = 0.2, regardless of its assessment window. This is practically appealing and greatly simplifies trial protocol preparation because in practice what often varies across trials is the assessment window, while the target DLT rate is often 0.2, 0.25 or 0.3.

During trial conduct, we impose the following overdose control / safety stopping rule: if $\Pr(p > \phi | s, n) > 0.95$ and $n \geq 3$, eliminate the current and higher doses from the trial; if the lowest dose is eliminated, terminate the trial early for safety. When the current dose is eliminated, the dose is de-escalated to the next lower level. The posterior probability $\Pr(p > \phi | s, n)$ is evaluated based on the following beta-binomial model:

$$s|p \sim Binom(p)$$

$$p \sim Beta(1, 1).$$

That is, $\Pr(p > \phi | s, n) = Beta(\phi;\ 1 + s, n - s + 1)$, where $Beta(\phi;\ 1 + s, n - s + 1)$ is the cumulative density function of a beta distribution evaluated at ϕ with parameters $1 + s$ and $n - s + 1$.

B. Simulation Settings

To simulate the toxicity outcome for a patient treated at a specific dose level *j*, we generated the time to DLT X_i for the *i*th patient, $i = 1, \cdots, N$, from a Weibull distribution $Weibull(\xi_j, \lambda_j)$. If $X_i < T$, we set y_i =1 to indicate that the *i*th patient will experience DLT; otherwise, we set y_i =0 to indicate that the *i*th patient will not experience DLT. We set $\xi_j = \log\left(-\log\left(\frac{1-p_j}{\log 2}\right)\right)/\log 2$ and $\lambda_j = \log(1 - p_j)\,/T^{\xi_j}$ such that 50% of DLT occurred in the latter half of the DLT assessment window, where p_j is the true DLT rate for dose level *j*, $j = 1, \cdots, J$. For the standard 3+3 method and cohort expansion, the trial enrolls the next cohort of new patients only when the DLT data of previously enrolled patients have been cleared.

The TITE-CRM method is based on the following power model

$$p_j = a_j^{\exp(\alpha)}$$

where a_j is the skeleton (i.e., prior estimate of DLT rate for dose level *j*), and α is the model parameter. When the target toxicity probability is 0.2, the skeleton was equal to scenario 7 in Table 1, which is consistent with the scenario reported by Normolle and Lawrence (2006). When the target toxicity probability is 0.3, the skeleton was based on the calibration method of Lee and Cheung (2009), with the prior MTD being dose level 4 and the halfwidth being 0.04, leading to skeleton = (0.10, 0.15, 0.22, 0.30, 0.38, 0.46, 0.53). The prior distribution for α is $N(0, 1.34)$. The uniform weighting scheme was used for the patients with DLT data pending. For the TITE-BOIN design, we took the default setting for the design parameters $\phi_1 = 0.6\phi$ and $\phi_2 = 1.4\phi$ to determine the escalation and de-escalation boundaries, as recommended by Liu and Yuan (2015). For fair comparison, the same accrual suspension rule (described in "TITE-BOIN

C. Sensitivity Analysis

Different late-onset profiles

We considered more severe late-onset DLTs by simulating the time to DLT from a Weibull distribution, with 70% of the DLTs occurring in the latter half of the assessment window. The results of the TITE-BOIN, TITE-CRM and R6 design (see Figures S1-S3) are similar to those reported in the main text, where 50% of DLTs occurred in the latter half of the assessment window, suggesting that the TITE-BOIN is robust to the late-onset toxicity profile.

Different accrual rates

In addition to the accrual rate of 2 patients per month, we considered a slower accrual rate of 1 patient/ month and a faster accrual rate of 3 patients/month. Simulation results (see Figures S4-S6 for 1 patient/month and S7-S9 for 3 patients/month) demonstrated that the performance of TITE-BOIN is not sensitive to different accrual rates and consistently better than that of the R6 design.

Different dose-toxicity scenarios

To confirm that our comparison results based on the 16 dose–toxicity scenarios (see Table S2) are generally applicable, we conducted a much larger scale simulation study that compared the performance of the 3+3, R6, TITE-BOIN and TITE-CRM designs based on 50,000 dose–toxicity scenarios, randomly generated using the method of Clertant and O'Quigley (2017). Figure S10 shows the 50 randomly generated dose–-toxicity scenarios with the target DLT rate of 0.2. For each of the 50,000 dose–toxicity scenarios, 10,000 trials were simulated. We considered the patient accrual rate of 1, 2 or 3 patients/month, with the DLT assessment window of 3 months. The skeleton of TITE-CRM was chosen based on the model calibration method of Lee and Cheung (2009), with the third dose level as the initial MTD guess and a halfwidth of 0.05, leading to skeleton = (0.05, 0.11, 0.20, 0.31, 0.42, 0.43) for the target DLT rate of 0.2 and skeleton = (0.12, 0.20, 0.30, 0.40, 0.50, 0.59) for the target DLT rate of 0.3. The remaining simulation configurations for the four designs were the same as those in the main simulation study with 16 dose–toxicity scenarios. Table S3 show the average performance of the four designs over 50,000 dose–toxicity scenarios. The results (based on 50,000 randomly generated scenarios) are consistent with those reported in the main text based on 16 scenarios. Specifically, in terms of MTD identification accuracy and trial duration, the TITE-BOIN design is comparable to the model-based TITE-CRM design, and performs uniformly better than the 3+3 and R6 designs. When the target DLT rate is 0.2, TITE-BOIN is as safe as the 3+3 and R6 designs and is slightly more conservative than TITE-CRM. When the target DLT rate is 0.3, the TITE-BOIN has substantially better overdose control than the TITE-CRM.

D. Incorporating prior information on the time to DLT

In some trials, prior information is available on the distribution of X_i (i.e., time to DLT). For example, for a certain drug, we may know *a priori* that the DLT is more likely to occur in the later part of the DLT assessment window $[0, T]$. Such prior information can be easily incorporated into the TITE-BOIN by specifying a piecewise uniform prior distribution for X_i, which partitions $[0, T]$ into several intervals and assumes a uniform distribution within each interval. By increasing the number of partitions, the piecewise uniform distribution can approximate any shape of the time-to-toxicity distribution. For ease of exposition, we consider three partitions, $[0, T/3]$, $(T/3, 2T/3]$ and $(2T/3, T]$, that are often adequate for practical use. Let $(\omega_1, \omega_2, \omega_3)$ be the prior probability that the DLT would occur at the three intervals, where $\omega_1 + \omega_2 + \omega_3 = 1$. Define $h_0 = 0$, $h_1 = T/3$, $h_2 = 2T/3$, $h_3 = T$, respectively, and define

$$\tilde{t}_{ik} = \begin{cases} 1, & t_i > h_k \\ \dfrac{3(t_i - h_{k-1})}{T}, & t_i \in (h_{k-1}, h_k], \\ 0, & otherwise. \end{cases} \quad k = 1, 2, 3.$$

Then, the conditional probability $\Pr(X_i > t_i \mid Y_i = 1)$ is given by

$$\Pr(X_i > t_i \mid Y_i = 1) = 1 - \sum_{k=1}^{3} \omega_k \tilde{t}_{ik},$$

Plugging this into equation (2) and going through the similar algebra and approximation, we have

$$\hat{p} = \frac{s + \dfrac{p}{1-p}(c - \text{WSTFT})}{n} \qquad (6)$$

where WSTFT is the weighted STFT, given by

$$\text{WSTFT} = \sum_{i \in M} \sum_{k=1}^{3} \omega_k \tilde{t}_{ik}$$

with M indicates the set of patients with pending DLT data. When $\omega_1 = \omega_2 = \omega_3 = 1/3$ (i.e., assigning an equal weight over the DLT assessment window), the WSTFT reduces to STFT. In other words, equation (3) using the uniform weight is a special case of equation (6). Following the same derivation procedure described previously, the same dose escalation and de-escalation boundaries given by equations (4) and (5) are obtained. The only difference is that we now compare WSTFT, rather than STFT, with the boundaries to determine dose escalation and de-escalation. More specifically, if the WSTFT for c pending patients $\geq \pi_e$, escalate the dose to the

next higher level; and if the WSTFT for c pending patients $\leq \pi_d$, de-escalate the dose to the next lower level; otherwise, stay at the current dose level.

Table S1. Dose escalation and de-escalation boundaries for TITE-BOIN with a target DLT rate of 0.3 and cohort size of 3.

No. treated	No. DLTs	No. data pending	STFT			No. treated	No. DLTs	No. data pending	STFT		
			Escalate	Stay	De-escalate				Escalate	Stay	De-escalate
3	0	≤1	Y			12	2	5	≥2.72	<2.72	
3	0	≥2	Suspend accrual			12	2	6	≥4.11	<4.11	
3	1	0		Y		12	2	≥7	Suspend accrual		
3	1	1		>0.88	≤0.88	12	3	≤6		Y	
3	1	≥2	Suspend accrual			12	3	≥7	Suspend accrual		
3	2	≤1			Y	12	4	0		Y	
3	3	0			Y&Elim	12	4	1		>0.43	≤0.43
6	0	≤3	Y			12	4	2		>1.50	≤1.50
6	0	≥4	Suspend accrual			12	4	3		>2.57	≤2.57
6	1	≤1	Y			12	4	4		>3.65	≤3.65
6	1	2	≥0.60	<0.60		12	4	5		>4.72	≤4.72
6	1	3	≥1.96	<1.96		12	4	6		>5.79	≤5.79
6	1	≥4	Suspend accrual			12	4	≥7	Suspend accrual		
6	2	0		Y		12	5, 6	≤7			Y
6	2	1		>0.73	≤0.73	12	≥7	≤5			Y&Elim
6	2	2		>1.80	≤1.80	15	0	≤7	Y		
6	2	3		>2.87	≤2.87	15	0	≥8	Suspend accrual		
6	2	≥4	Suspend accrual			15	1	≤7	Y		
6	3	≤3			Y	15	1	≥8	Suspend accrual		
6	≥4	≤2			Y&Elim	15	2	≤5	Y		
9	0	≤4	Y			15	2	6	≥0.35	<0.35	
9	0	≥5	Suspend accrual			15	2	7	≥2.07	<2.07	
9	1	≤4	Y			15	2	≥8	Suspend accrual		
9	1	≥5	Suspend accrual			15	3	≤1	Y		
9	2	0	Y			15	3	2	≥0.11	<0.11	
9	2	1	≥0.59	<0.59		15	3	3	≥1.29	<1.29	
9	2	2	≥1.65	<1.65		15	3	4	≥2.46	<2.46	
9	2	3	≥2.71	<2.71		15	3	5	≥3.64	<3.64	

9	2	4	≥3.77	<3.77		15	3	6	≥4.81	<4.81	
9	2	≥5	Suspend accrual			15	3	7	≥5.98	<5.98	
9	3	0	Y			15	3	≥8	Suspend accrual		
9	3	1		>0.58	≤0.58	15	4	≤7	Y		
9	3	2		>1.65	≤1.65	15	4	≥8	Suspend accrual		
9	3	3		>2.72	≤2.72	15	5	0	Y		
9	3	4		>3.79	≤3.79	15	5	1		>0.28	≤0.28
9	3	≥5	Suspend accrual			15	5	2		>1.35	≤1.35
9	4	≤5			Y	15	5	3		>2.42	≤2.42
9	≥5	≤4			Y&Elim	15	5	4		>3.50	≤3.50
12	0	≤6	Y			15	5	5		>4.57	≤4.57
12	0	≥7	Suspend accrual			15	5	6		>5.64	≤5.64
12	1	≤6	Y			15	5	7		>6.72	≤6.72
12	1	≥7	Suspend accrual			15	5	≥8	Suspend accrual		
12	2	≤3	Y			15	6, 7	≤9			Y
12	2	4	≥1.33	<1.33		15	≥8	≤7			Y&Elim

Note: "No. treated" is the total number of patients treated at the current dose level, "No. DLTs" is the number of patients who experienced DLT at the current dose level, "No. with data pending" denotes that number of patients whose DLT data are pending at the current dose level, "STFT" is the standardized total follow-up time (in months) for the patients with data pending, defined as the total follow-up time for the patients with data pending divided by the length of the DLT assessment window. "Y" represents "Yes", and "Y&Elim" represents "Yes & Eliminate". When a dose is eliminated, all higher doses should also be eliminated.

Table S2. Sixteen true toxicity scenarios with target DLT rates of 0.2 and 0.3. The target DLT rate is 0.2 in scenarios 1-8 and 0.3 in scenarios 9-16. The MTD is in bold face. Scenarios 3, 4, 5 and 7 were previously considered by Normolle and Lawrence (2006; JCO).

Scenario	Dose level						
	1	2	3	4	5	6	7
			Target DLT rate is 0.2				
1	0.05	**0.20**	0.46	0.50	0.60	0.70	0.80
2	0.02	0.05	**0.20**	0.28	0.34	0.40	0.44
3	0.01	0.05	0.10	**0.20**	0.32	0.50	0.70
4	0.01	0.04	0.07	**0.10**	0.50	0.70	0.90
5	0.01	0.05	0.10	0.14	**0.20**	0.26	0.34
6	0.01	0.02	0.03	0.05	**0.20**	0.40	0.50
7	0.01	0.04	0.07	0.10	0.15	**0.20**	0.25
8	0.01	0.02	0.03	0.04	0.05	**0.20**	0.45
			Target DLT rate is 0.3				
9	**0.30**	0.40	0.50	0.60	0.70	0.80	0.90
10	0.14	**0.30**	0.39	0.48	0.56	0.64	0.70
11	0.07	**0.23**	0.41	0.49	0.62	0.68	0.73
12	0.05	0.15	**0.30**	0.40	0.50	0.60	0.70
13	0.05	0.12	0.20	**0.30**	0.38	0.49	0.56
14	0.01	0.04	0.08	0.15	**0.30**	0.36	0.43
15	0.02	0.04	0.08	0.10	0.20	**0.30**	0.40
16	0.01	0.03	0.05	0.07	0.09	**0.30**	0.50

Table S3. Simulation results averaged over 50,000 randomly generated dose–toxicity scenarios with six dose levels. Under each of the random scenarios, 10,000 trials were simulated.

	Target DLT probability is 0.2				**Target DLT probability is 0.3**			
	Method				Method			
Performance characteristic	3+3	R6	TITE-CRM	TITE-BOIN	3+3	R6	TITE-CRM	TITE-BOIN
Accrual rate = 1 patient/month								
Correct selection %	31.0	38.1	46.0	44.2	21.4	24.0	49.5	46.9
Overdose selection %	12.4	13.5	13.8	11.7	4.9	3.8	17.4	13.1
Correct allocation %	27.6	24.6	32.7	30.8	21.3	19.1	34.5	32.0
Overdose allocation %	15.6	12.2	14.6	12.3	7.7	5.3	18.6	11.9
Regretful trials %	0.0	35.0	13.5	3.1	0.0	36.6	35.2	5.9
Average duration	56.8	49.3	40.5	41.7	54.5	46.8	40.6	43.0
Accrual rate = 2 patients/month								
Correct selection %	30.8	38.4	44.7	44.8	21.7	23.5	48.0	46.7
Overdose selection %	12.0	13.4	13.4	11.8	4.4	4.1	17.9	11.8
Correct allocation %	27.5	24.7	32.0	31.1	21.5	19.1	34.0	31.6
Overdose allocation %	15.2	12.4	14.5	12.6	7.5	5.7	17.9	11.0
Regretful trials %	0.0	36.0	18.6	2.9	0.0	38.8	45.3	5.7
Average duration	44.5	36.1	25.6	27.8	42.6	34.7	25.7	28.7
Accrual rate = 3 patients/month								
Correct selection %	31.6	38.2	43.6	45.3	21.8	23.8	47.7	46.5
Overdose selection %	12.0	13.3	13.0	12.1	4.4	4.2	17.4	11.5
Correct allocation %	28.0	24.5	31.0	31.0	21.5	19.0	32.6	31.8
Overdose allocation %	15.4	12.4	14.4	13.3	7.4	5.9	18.0	11.2
Regretful trials %	0.0	35.7	20.8	2.9	0.0	38.3	46.6	6.2
Average duration	40.2	31.6	21.0	23.6	38.7	30.7	21.0	24.5

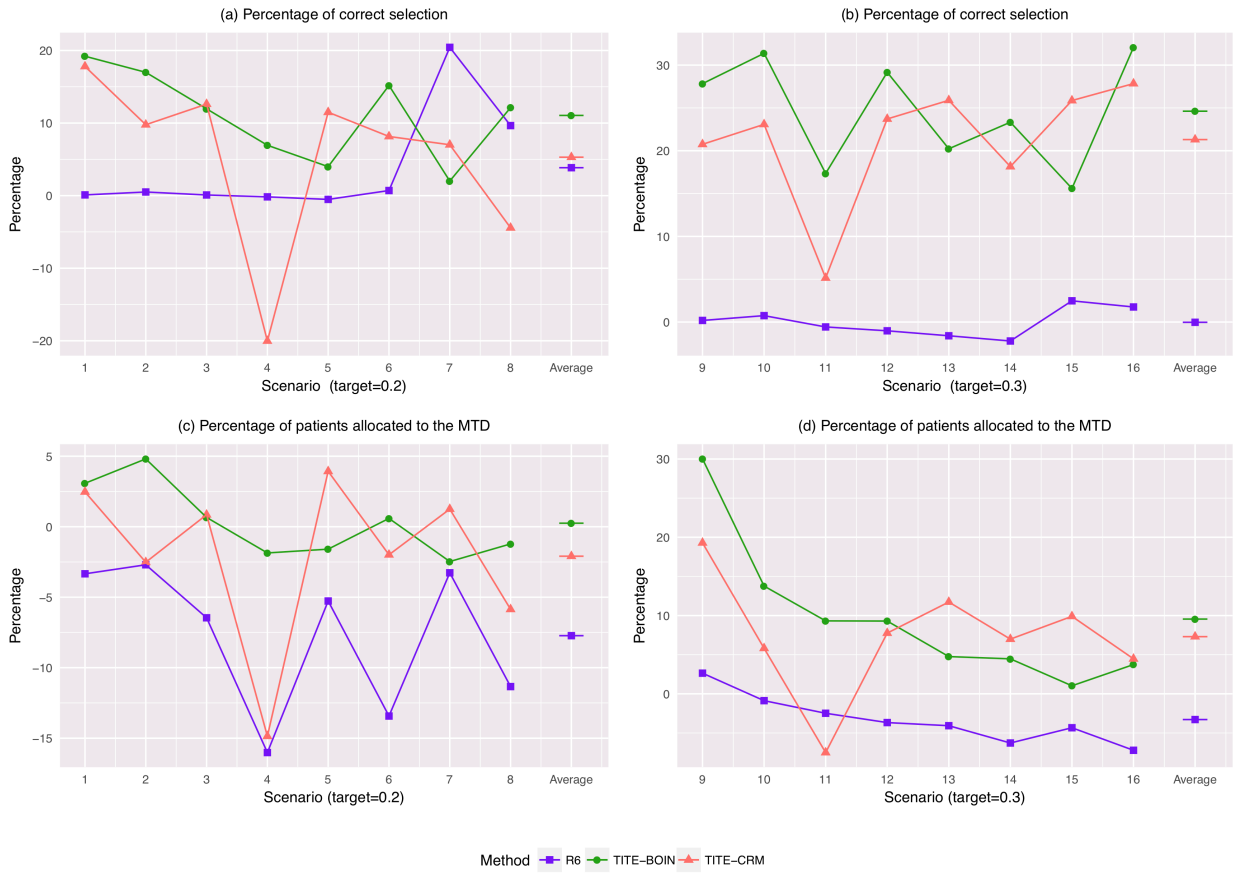

Figure S1. The relative percentages of correct selection of the MTD and the relative percentages of patients assigned to the correct MTD based on the time-to-event Bayesian optimal interval (TITE-BOIN), the rolling six (R6), and the time-to-event continual reassessment method (TITE-CRM) designs, with respect to the 3+3 design, when 70% of the toxicity events occur in the latter half of the DLT assessment window. The target DLT rate in scenarios 1-8 is 0.2, while that in scenarios 9-16 is 0.3. The accrual rate is 2 patients/month.

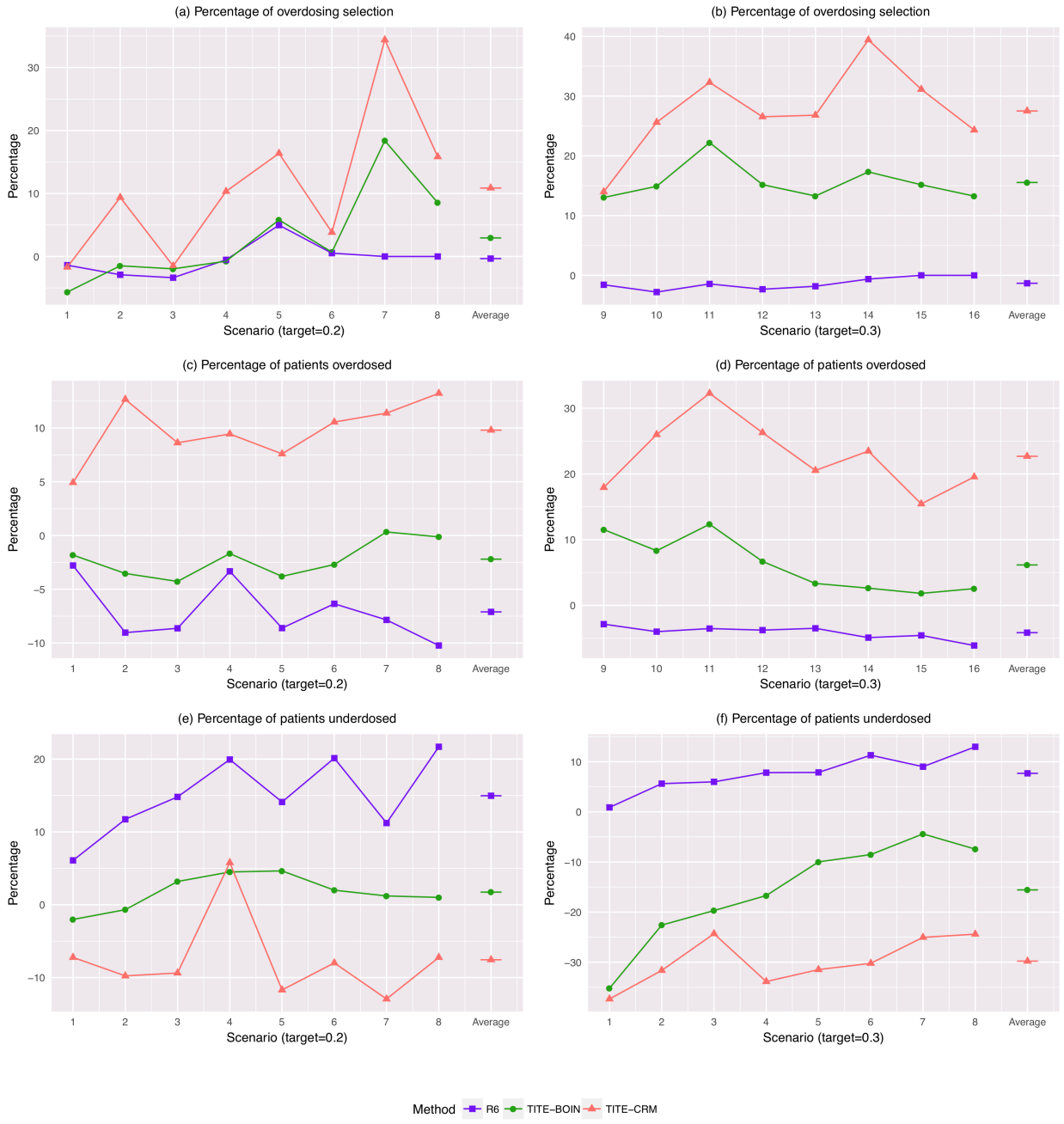

Figure S2. The relative percentages of overdose selection (a and b), the relative percentages of patients assigned to the doses above the MTD (c and d) and below the MTD (e and f), based on the time-to-event Bayesian optimal interval (TITE-BOIN), the rolling six (R6), and the time-to-event continual reassessment method (TITE-CRM) designs, with respect to the 3+3 design, when 70% of the toxicity events occur in the latter half of the DLT assessment window. The target DLT rate in scenarios 1-8 is 0.2, while that in scenarios 9-16 is 0.3. The accrual rate is 2 patients/month.

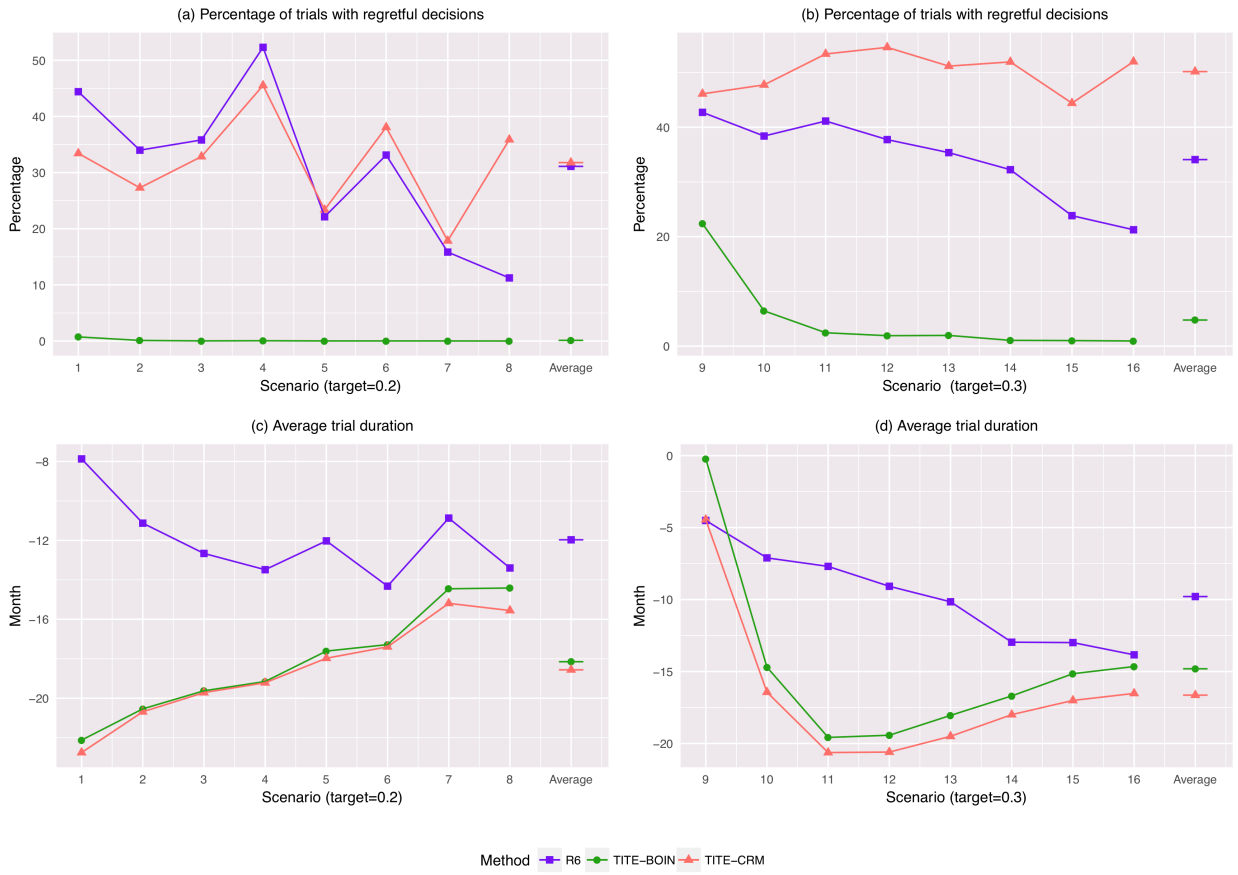

Figure S3. The relative percentages of "regretful" trials and the relative average trial duration based on the time-to-event Bayesian optimal interval (TITE-BOIN), the rolling six (R6), and the time-to-event continual reassessment method (TITE-CRM) designs, with respect to the 3+3 design, when 70% of the toxicity events occur in the latter half of the DLT assessment window. The target DLT rate in scenarios 1-8 is 0.2, while that in scenarios 9-16 is 0.3. The accrual rate is 2 patients/month.

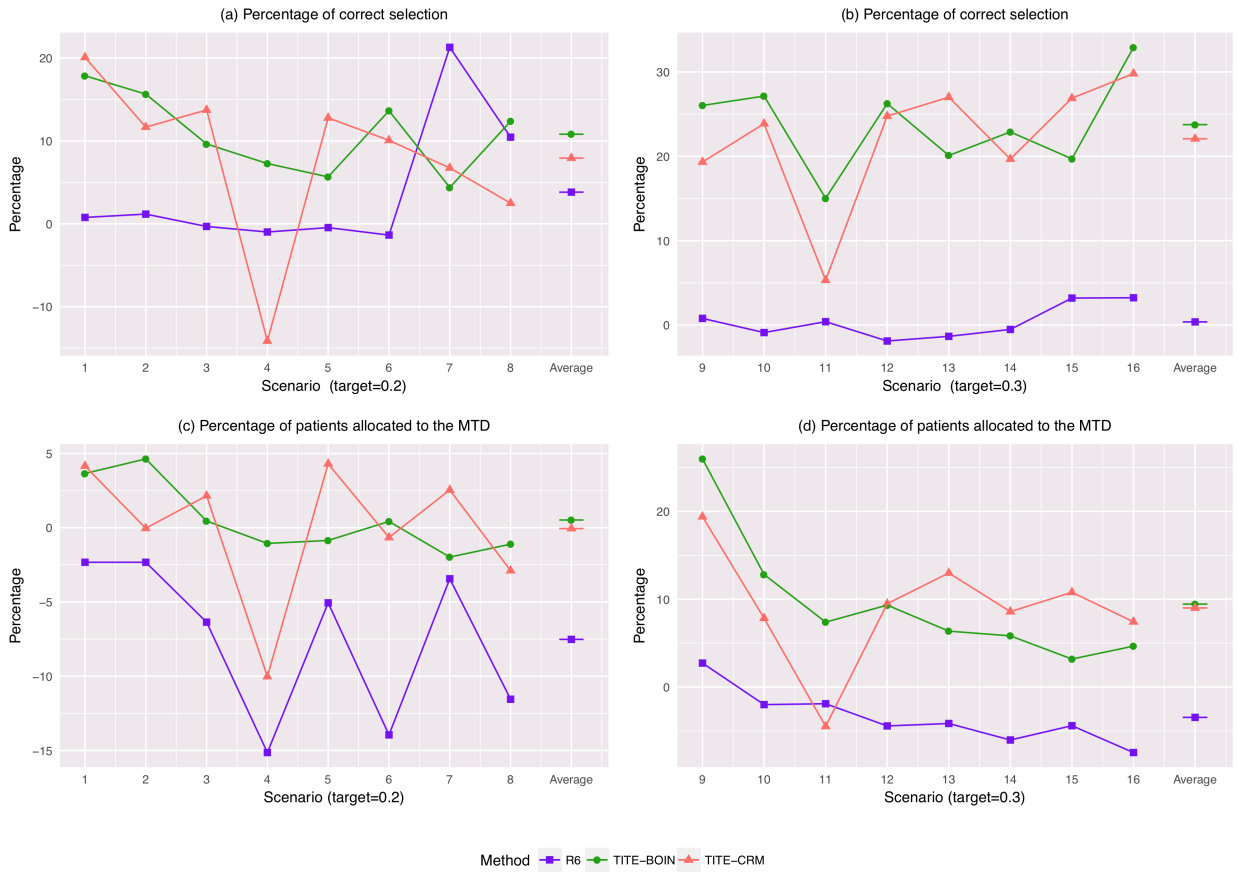

Figure S4. The relative percentages of correct selection of the MTD and the relative percentages of patients assigned to the correct MTD based on the time-to-event Bayesian optimal interval (TITE-BOIN), the rolling six (R6), and the time-to-event continual reassessment method (TITE-CRM) designs, with respect to the 3+3 design, when 50% of the toxicity events occur in the latter half of the DLT assessment window. The target DLT rate in scenarios 1-8 is 0.2, while that in scenarios 9-16 is 0.3. The accrual rate is 1 patient/month.

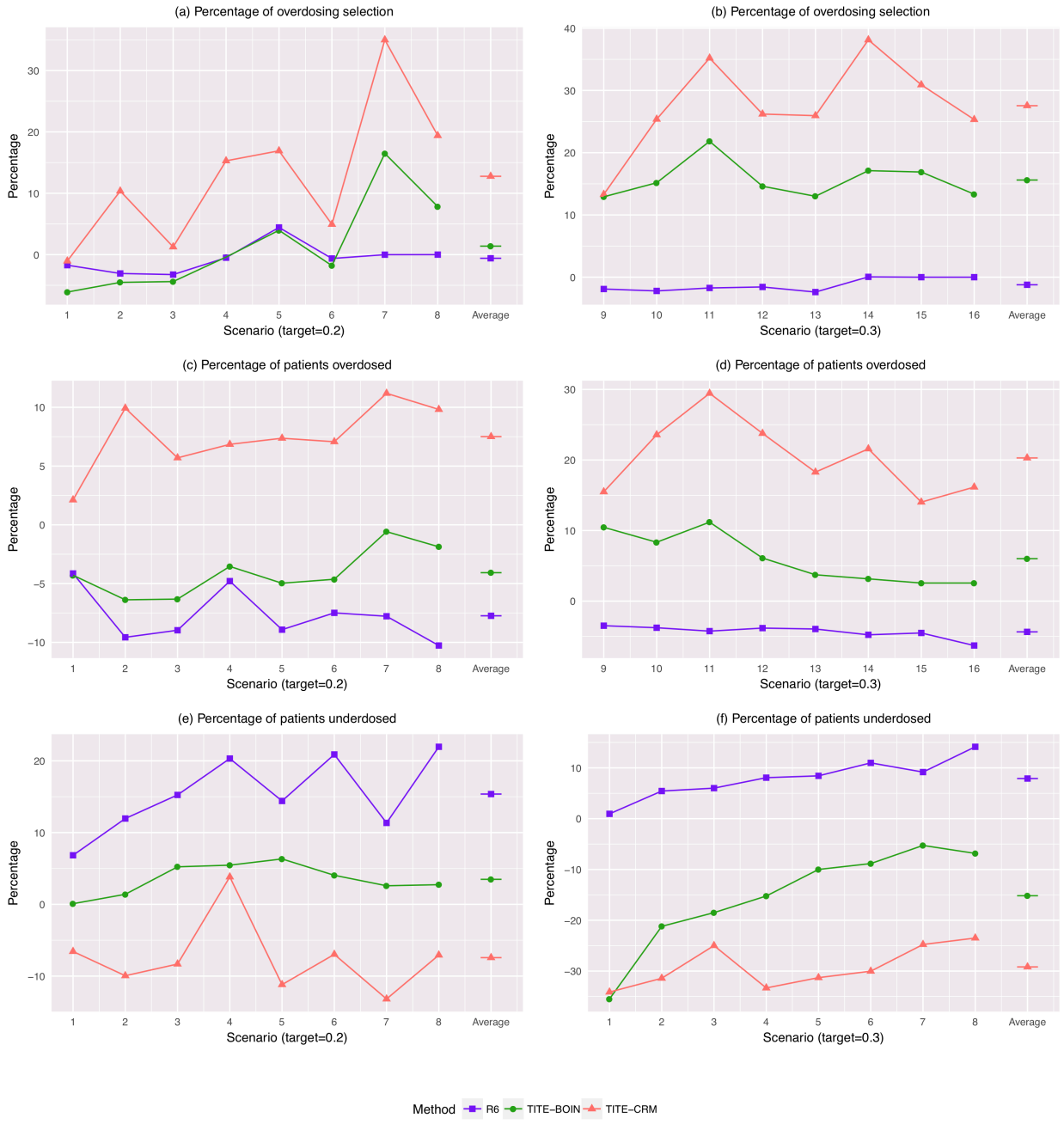

Figure S5. The relative percentages of overdose selection (a and b), and the relative percentages of patients assigned to the doses above the MTD (c and d) and below the MTD (e and f), based on the time-to-event Bayesian optimal interval (TITE-BOIN), the rolling six (R6), and the time-to-event continual reassessment method (TITE-CRM) designs, with respect to the 3+3 design, when 50% of the toxicity events occur in the latter half of the DLT assessment window. The target DLT rate in scenarios 1-8 is 0.2, while that in scenarios 9-16 is 0.3. The accrual rate is 1 patient/month.

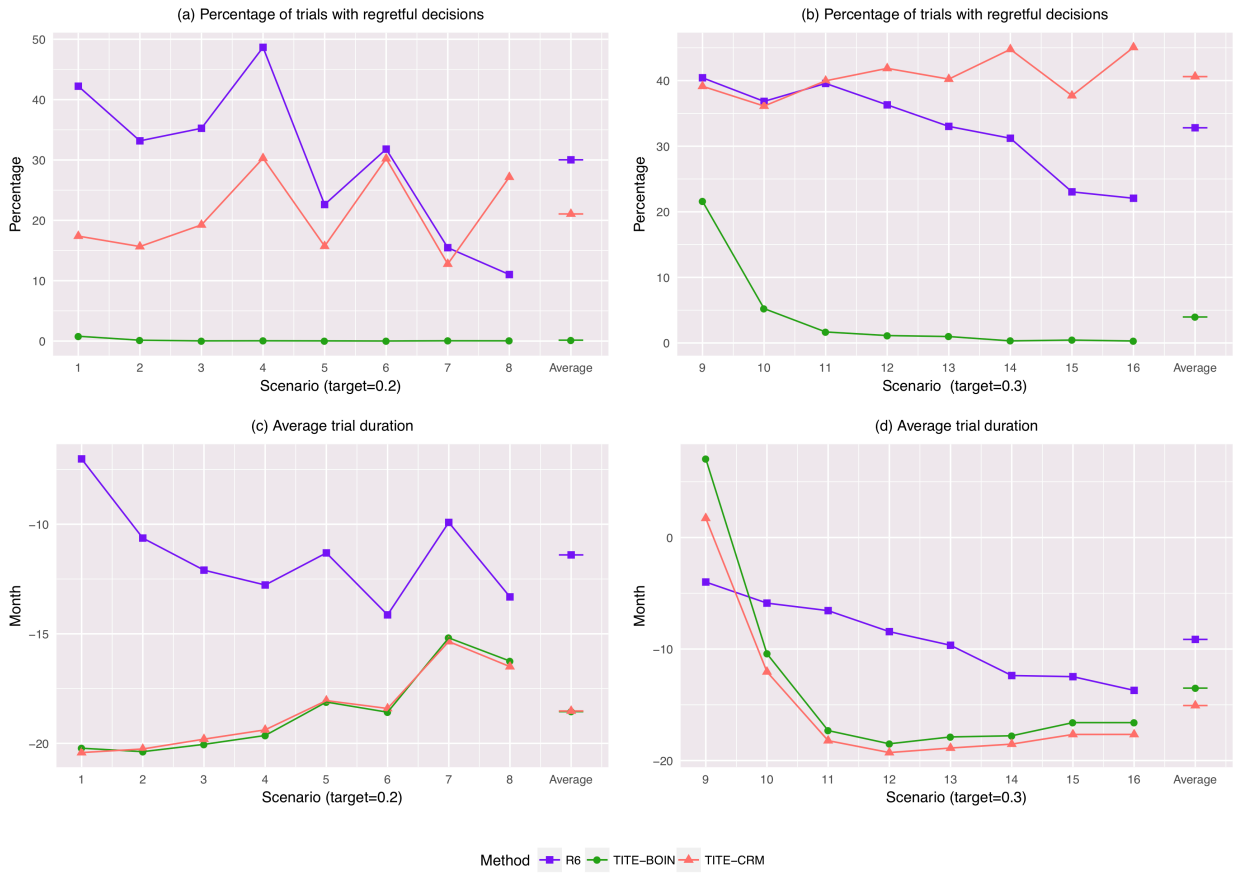

Figure S6. The relative percentages of "regretful" trials and the relative average trial duration based on the time-to-event Bayesian optimal interval (TITE-BOIN), the rolling six (R6), and the time-to-event continual reassessment method (TITE-CRM) designs, with respect to the 3+3 design, when 50% of the toxicity events occur in the latter half of the DLT assessment window. The target DLT rate in scenarios 1-8 is 0.2, while that in scenarios 9-16 is 0.3. The accrual rate is 1 patient/month.

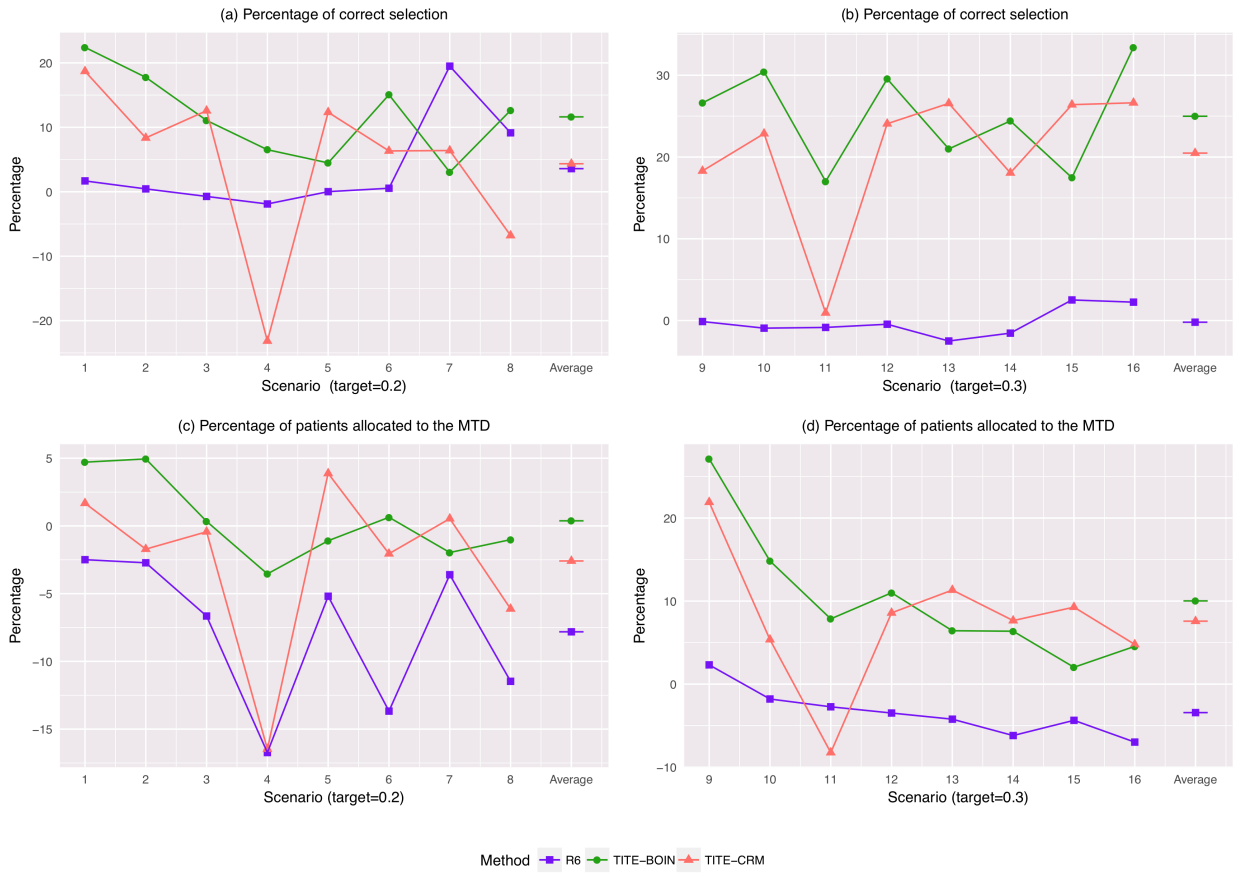

Figure S7. The relative percentages of correct selection of the MTD and the relative percentages of patients assigned to the correct MTD based on the time-to-event Bayesian optimal interval (TITE-BOIN), the rolling six (R6), and the time-to-event continual reassessment method (TITE-CRM) designs, with respect to the 3+3 design, when 50% of the toxicity events occur in the latter half of the DLT assessment window. The target DLT rate in scenarios 1-8 is 0.2, while that in scenarios 9-16 is 0.3. The accrual rate is 3 patients/month.

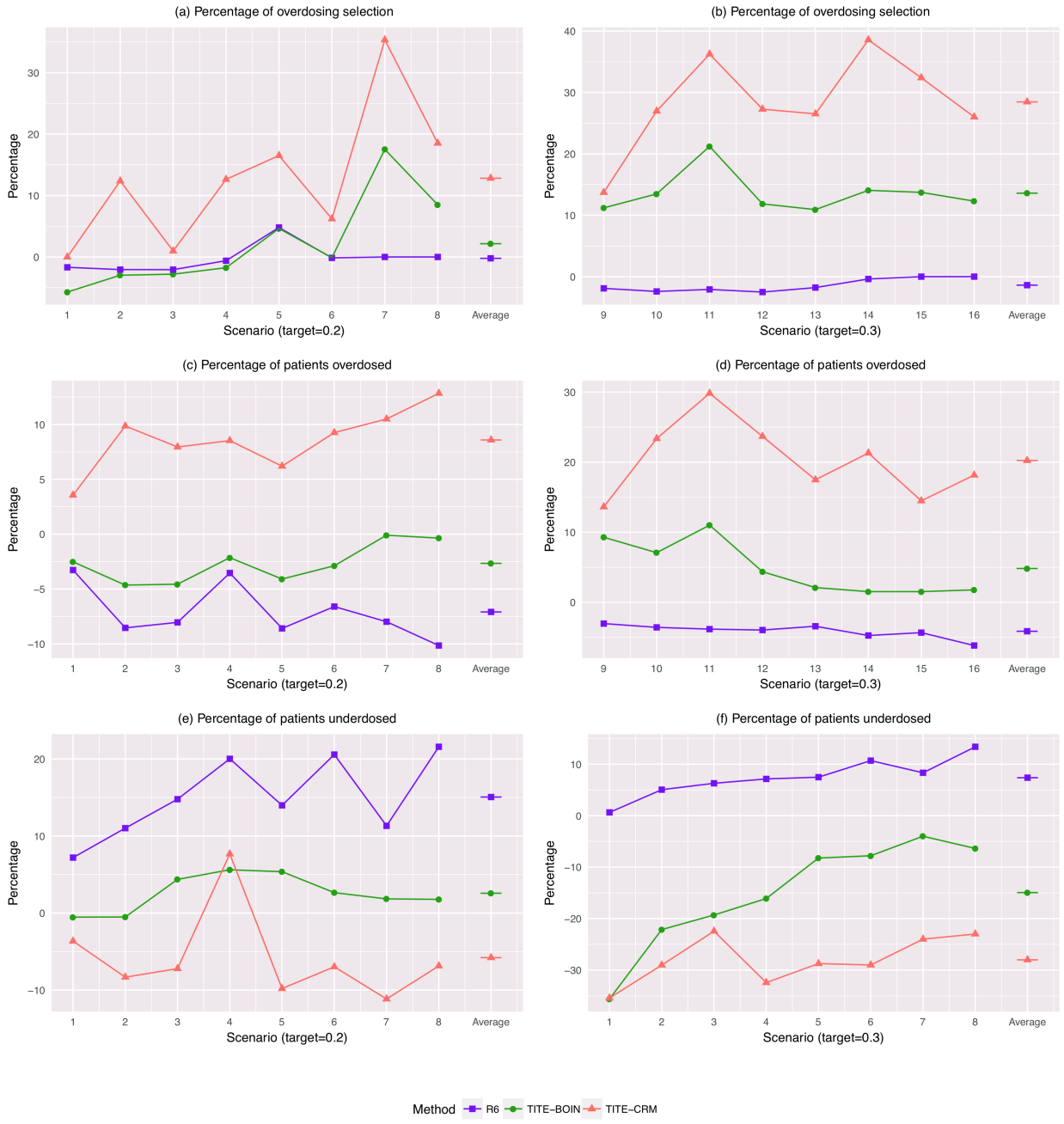

Figure S8. The relative percentages of overdose selection (a and b), and the relative percentages of patients assigned to the doses above the MTD (c and d) and below the MTD (e and f), based on the time-to-event Bayesian optimal interval (TITE-BOIN), the rolling six (R6), and the time-to-event continual reassessment method (TITE-CRM) designs, with respect to the 3+3 design, when 50% of the toxicity events occur in the latter half of the DLT assessment window. The target DLT rate in scenarios 1-8 is 0.2, while that in scenarios 9-16 is 0.3. The accrual rate is 3 patients/month.

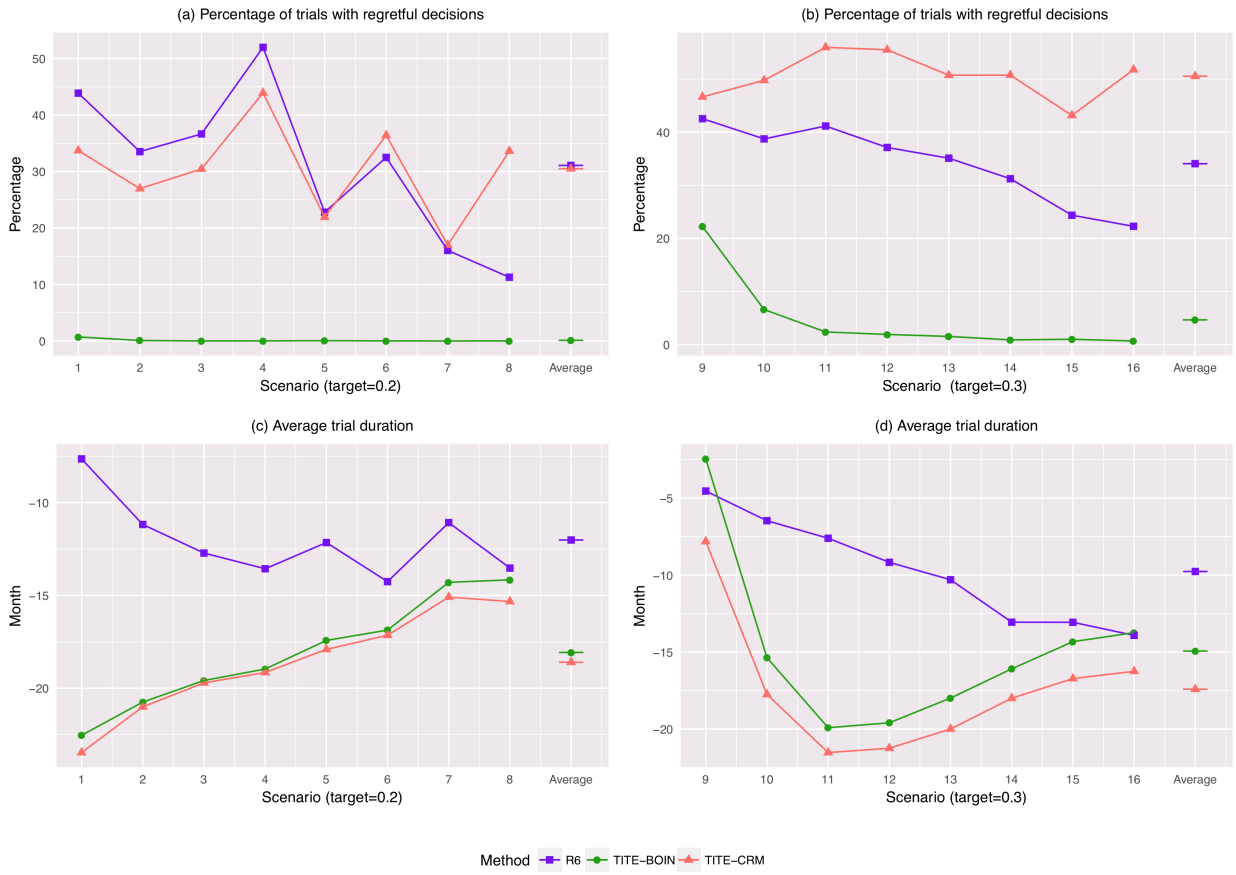

Figure S9. The relative percentages of "regretful" trials and the relative average trial duration based on the time-to-event Bayesian optimal interval (TITE-BOIN), the rolling six (R6), and the time-to-event continual reassessment method (TITE-CRM) designs, with respect to the 3+3 design, when 50% of the toxicity events occur in the latter half of the DLT assessment window. The target DLT rate in scenarios 1-8 is 0.2, while that in scenarios 9-16 is 0.3. The accrual rate is 3 patients/month.

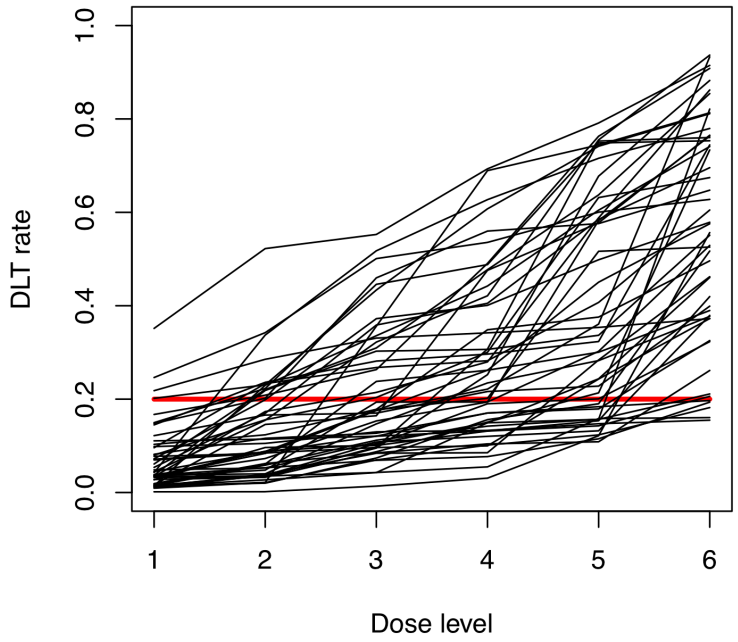

Figure S10. Fifty randomly generated dose–toxicity curves with the target DLT rate of 0.2.

References

1. Liu S, Yuan Y: Bayesian optimal interval designs for phase I clinical trials. J R Stat Soc Ser C Appl Stat 64: 507-523, 2015
2. Little RJ, Rubin DB: Statistical Analysis with Missing Data. John Wiley & Sons, 2004
3. Normolle D, Lawrence T: Designing dose-escalation trials with late-onset toxicities using the time-to-event continual reassessment method. J Clin Oncol 24: 4426-4433, 2006
4. Lee SM, Cheung YK: Model calibration in the continual reassessment method. Clin Trials 6:227-238, 2009
5. Clertant M, O'Quigley J: Semiparametric dose finding methods. J R Stat Soc Series B Stat Methodol 79: 1487-1508, 2017