Online Supplementary Data

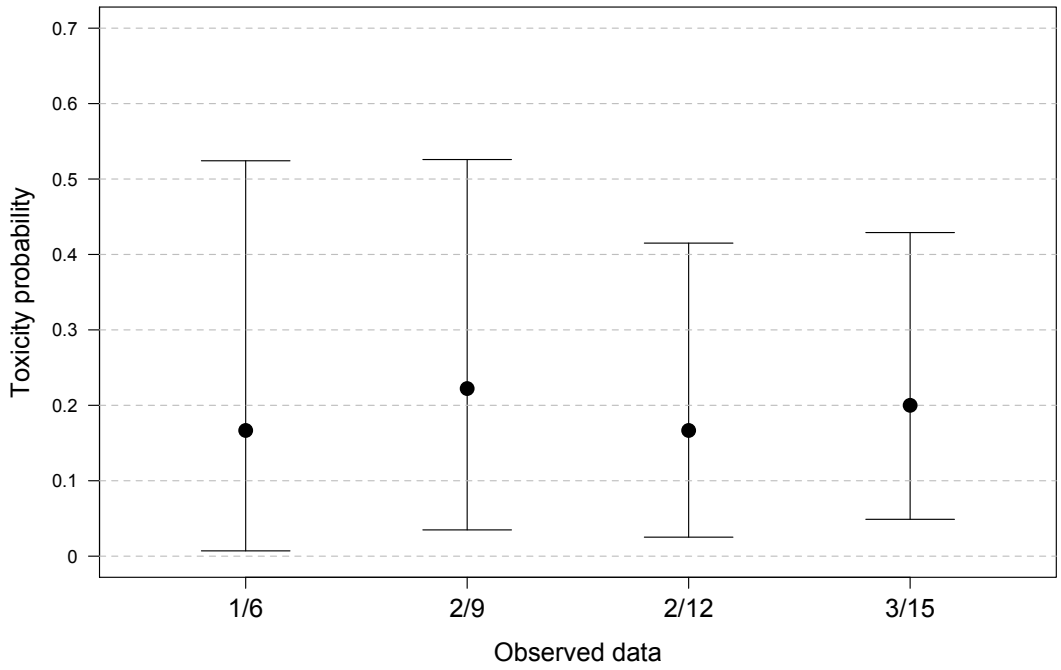

Figure S1. 95% posterior credible intervals for probability of toxicity at the MTD, corresponding to each of four phase I trials in which 1/6, 2/9, 2/12 and 3/15 patients experienced a DLT, respectively. The dots indicate the observed toxicity rate at the MTD.

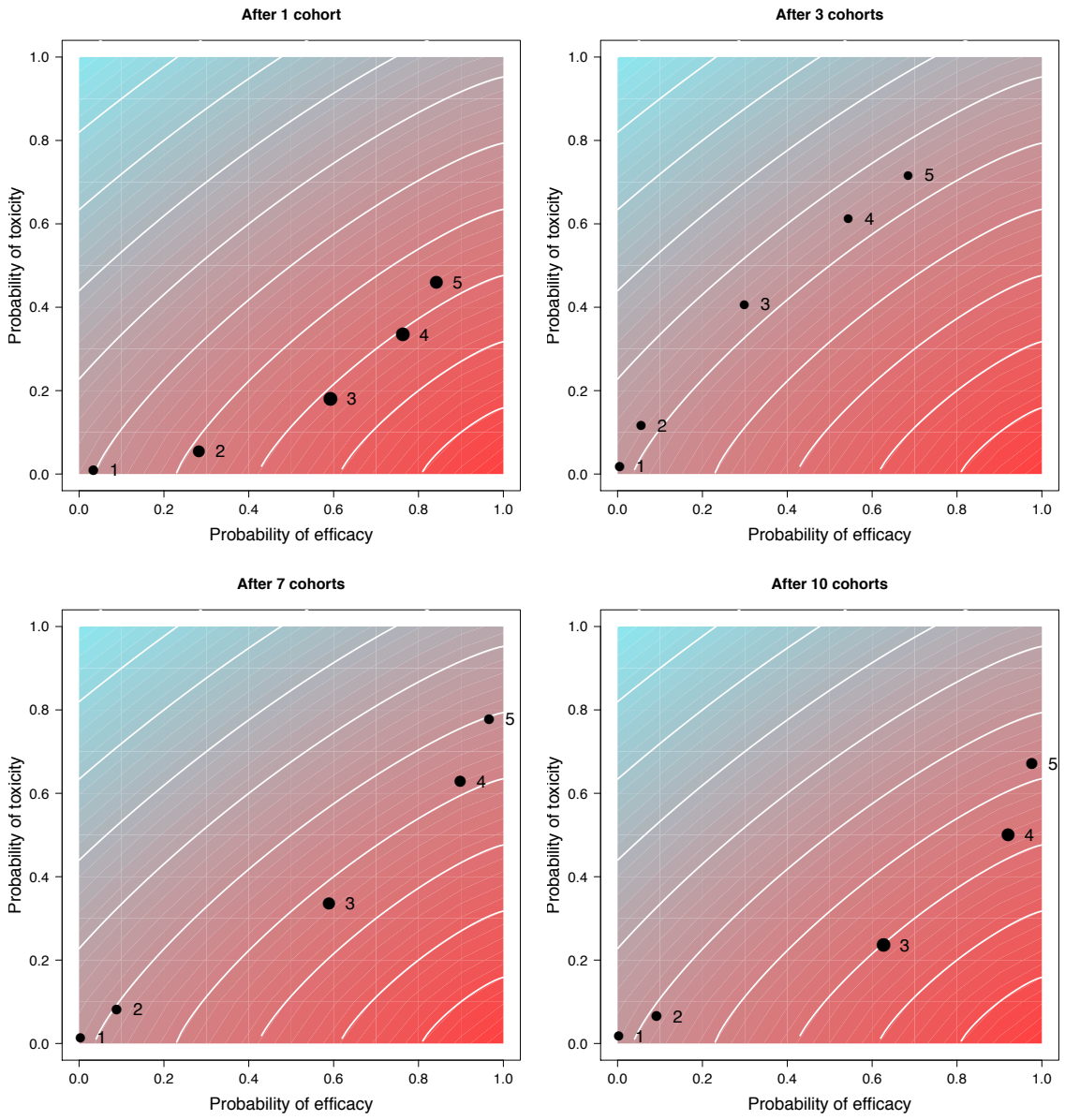

Figure S2. Estimated $[p_E(d), p_T(d)]$ pairs of doses d=1, 2, 3, 4, 5, after 1, 3, 7, and 10 cohorts of patients have been treated in an EffTox trial. The dot locations and associated numbers indicate the estimates and doses, respectively. The size of the each dot is proportional to its desirability.

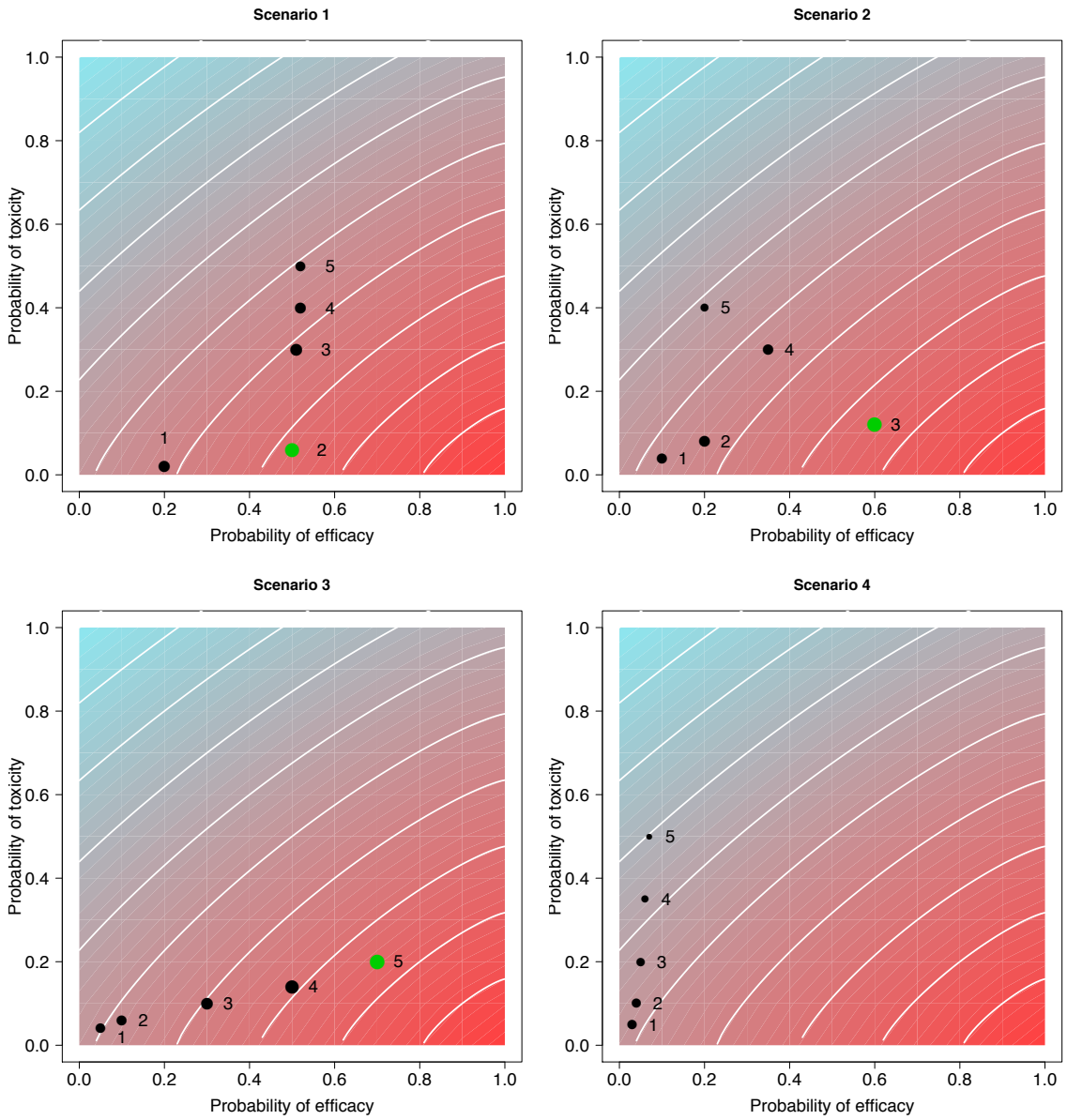

Figure S3. True values of $[p_E(d), p_T(d)]$ pairs for doses d=1, 2, 3, 4, 5, for each of the four simulation scenarios. The size of each dot is proportional to its desirability, and the optimal dose with highest desirability is highlighted in green. In Scenario 4, all doses have unacceptably low efficacy, thus no optimal dose exists.

Table S1. Simulation results for a trial with a maximum sample size of 30 patients.

Design		Dose level 1	2	3	4	5	No Dose Selected	Sample Size
				Scenario 1				
	Prob(toxicity)	0.02	0.06	0.30	0.40	0.50		
	Prob(efficacy)	0.2	0.5	0.51	0.52	0.52		
	Trade-off	0.58	0.72	0.61	0.56	0.51		
3+3	Selection %	3.6	**54.8**	28.2	10.6	0.0	2.8	
	No. patients	3.9	**13.4**	8.4	3.3	0.6		29.6
CRM-CE	Selection %	1.7	**45.1**	37.6	12.5	2.8	0.0	
	No. patients	3.8	**12.2**	10.5	2.9	0.6		30.0
CRM	Selection %	0.3	**34.4**	61.1	4.0	0.1	0.0	
	No. patients	3.7	**11.3**	12.0	2.5	0.4		29.9
EffTox	Selection %	2.0	**66.0**	29.0	2.0	0	1.0	
	No. patients	3.6	**15.0**	9.9	1.2	0.2		29.9
				Scenario 2				
	Prob(toxicity)	0.04	0.08	0.12	0.30	0.40		
	Prob(efficacy)	0.10	0.20	0.60	0.35	0.20		
	Trade-off	0.52	0.56	0.75	0.54	0.41		
3+3	Selection %	6.9	12.6	**43.5**	25.3	0.0	11.7	
	No. patients	4.8	5.8	**9.8**	6.4	1.7		28.5
CRM-CE	Selection %	5.1	23.1	**31.4**	27.4	12.4	0.6	
	No. patients	5.0	8.0	**8.5**	6.0	2.4		29.9
CRM	Selection %	0.9	13.5	**48.3**	32.3	3.9	1.1	
	No. patients	4.4	7.3	**10.2**	6.2	1.6		29.7
EffTox	Selection %	2.0	13.0	**65.0**	14.0	3.0	4.0	
	No. patients	3.7	6.0	**13.4**	4.9	1.3		29.3

				Scenario 3				
	Prob(toxicity)	0.04	0.06	0.10	0.14	0.20		
	Prob(efficacy)	0.05	0.10	0.30	0.50	0.70		
	Trade-off	0.50	0.52	0.60	0.69	0.75		
3+3	Selection %	3.5	8.5	17.8	18.4	**0.0**	51.8	
	No. patients	4.1	5.0	6.1	5.6	**2.9**		23.7
CRM-CE	Selection %	4.0	18.4	19.6	27.0	**30.4**	0.6	
	No. patients	4.6	7.0	6.6	6.1	**5.6**		29.9
CRM	Selection %	0.8	8.1	24.0	29.2	**36.8**	1.1	
	No. patients	4.2	6.1	7.4	6.1	**5.9**		29.7
EffTox	Selection %	1.0	3.0	16.0	29.0	**48.0**	4.0	
	No. patients	3.3	4.3	6.9	7.0	**7.8**		29.3
				Scenario 4				
	Prob(toxicity)	0.05	0.10	0.20	0.35	0.50		
	Prob(efficacy)	0.03	0.04	0.05	0.06	0.07		
	Trade-off	0.48	0.47	0.44	0.37	0.30		
3+3	Selection %	9.1	28.9	37.5	18.6	0.0	**5.9**	
	No. patients	5.3	8.7	9.1	4.9	1.0		29.0
CRM-CE	Selection %	9.2	35.8	31.8	17.1	5.6	**0.5**	
	No. patients	6.0	10.4	8.5	3.8	1.1		29.8
CRM	Selection %	2.7	28.8	53.3	13.5	0.6	**1.1**	
	No. patients	5.3	9.9	10.2	3.7	0.6		29.7
EffTox	Selection %	1.0	1.0	9.0	10.0	4.0	**77.0**	
	No. patients	3.6	4.2	6.1	4.5	1.5		19.9

Table S2. Simulation results for trial with a maximum sample size of 60 patients

Design		Dose level 1	2	3	4	5	No Dose Selected	Sample Size
				Scenario 1				
	Prob(toxicity)	0.02	0.06	0.30	0.40	0.50		
	Prob(efficacy)	0.2	0.5	0.51	0.52	0.52		
	Trade-off	0.58	0.72	0.61	0.56	0.51		
3+3	Selection %	3.6	**54.8**	28.2	10.6	0.0	2.8	
	No. patients	5.0	**29.8**	16.9	6.5	0.6		58.8
CRM-CE	Selection %	0.1	**35.2**	59.6	4.8	0.1	0.2	
	No. patients	3.8	**21.6**	30.1	4.0	0.3		59.8
CRM	Selection %	0.0	**33.8**	65.3	0.6	0.0	0.3	
	No. patients	3.7	**22.0**	30.4	3.3	0.4		59.8
EffTox	Selection %	2.0	**80.0**	18.0	0.0	0.0	1.0	
	No. patients	4.2	**37.3**	16.6	1.4	0.2		59.7
				Scenario 2				
	Prob(toxicity)	0.04	0.08	0.12	0.30	0.40		
	Prob(efficacy)	0.10	0.20	0.60	0.35	0.20		
	Trade-off	0.52	0.56	0.75	0.54	0.41		
3+3	Selection %	6.9	12.6	**43.5**	25.3	0	11.7	
	No. patients	6.9	9.6	**22.8**	14	1.7		55.0
CRM-CE	Selection %	0.8	13.6	**51.0**	30.9	3.3	0.4	
	No. patients	4.8	11	**25.6**	15.7	2.7		59.8
CRM	Selection %	0.2	6.3	**59.0**	33.8	0.4	0.3	
	No. patients	4.7	10	**26.4**	16.7	2.1		59.9
EffTox	Selection %	1.0	13.0	**76.0**	6.0	0	4.0	
	No. patients	4.2	10.2	**34.3**	7.9	1.7		58.3

				Scenario 3				
	Prob(toxicity)	0.04	0.06	0.10	0.14	0.20		
	Prob(efficacy)	0.05	0.10	0.30	0.50	0.70		
	Trade-off	0.50	0.52	0.60	0.69	0.75		
3+3	Selection %	3.5	8.5	17.8	18.4	**0.0**	51.8	
	No. patients	5.1	7.5	11.4	11.1	**2.9**		38.0
CRM-CE	Selection %	0.5	8.6	24.1	27.8	**38.6**	0.4	
	No. patients	4.5	8.4	14.5	14.7	**17.6**		59.7
CRM	Selection %	0.1	2.8	17.9	32.9	**46.0**	0.3	
	No. patients	4.4	7.6	13.1	15.9	**18.7**		59.7
EffTox	Selection %	0.0	1.0	16.0	30.0	**49.0**	3.0	
	No. patients	3.3	4.7	11.2	16.2	**22.6**		58.0
				Scenario 4				
	Prob(toxicity)	0.05	0.10	0.20	0.35	0.50		
	Prob(efficacy)	0.03	0.04	0.05	0.06	0.07		
	Trade-off	0.48	0.47	0.44	0.37	0.30		
3+3	Selection %	9.1	29.0	37.4	18.6	0.0	**5.9**	
	No. patients	8.0	17.4	20.3	10.4	1.0		57.1
CRM-CE	Selection %	2.3	27.3	55.4	13.1	0.8	**1.1**	
	No. patients	5.8	17.5	27.5	7.7	0.8		59.3
CRM	Selection %	0.1	22.4	69.8	6.8	0.0	**0.9**	
	No. patients	5.4	17.8	29.0	6.7	0.6		59.5
EffTox	Selection %	0.0	1.0	5.0	5.0	2.0	**87.0**	
	No. patients	3.4	4.4	7.0	4.7	1.7		21.2

Dose-outcome model assumed by the EffTox design

Given raw doses $d_1 < d_2 < \ldots < d_K$, standardized doses used in the model are defined as $x_k = \log(d_k) - (1/K)\{\log(d_1) + \ldots + \log(d_K)\}$. Denoting the linear terms $\eta(E,k) = \mu_E + \beta_{E,1} x_k + \beta_{E,2}(x_k)^2$ and $\eta(T,k) = \mu_T + \beta_T x_k$, the two marginal probabilities are $\pi(j,k) = \text{logit}^{-1}\{\eta(j,k)\}$ for j=E, T, and k=1,…,K, with $\beta_T > 0$ to ensure that $\pi(T,k)$ increases with dose. A bivariate distribution is obtained by assuming a Gumbel- Morgenstern copula. Priors on the model parameters are assumed to be normally distributed, with hyper-parameter means determined from elicited means of $\pi(j,k)$ for each (j,k) and hyper-parameter variances calibrated to obtain a given specified prior effective sample size. Additional details are given in Thall, et al.[23].

Construction of efficacy-toxicity trade-off contours

To construct efficacy-toxicity trade-off contours for a trial, one first specifies three equally desirable efficacy-toxicity probability pairs $\pi_1^* = (p_{E,1}^*,\ 0)$, $\pi_2^* = (1,\ p_{T,2}^*)$ and $\pi_3^* = (p_{E,3}^*,\ p_{T,3}^*)$, subject to the constraints $p_{E,1}^* < p_{E,3}^*$ and $p_{T,3}^* < p_{T,2}^*$. In our trial example, the three pairs are $\pi_1^* = (.15, 0)$, $\pi_2^* = (1, .50)$ and $\pi_3^* = (.30, .15)$. Thus, a dose with 15% efficacy and no toxicity, a dose with 100% efficacy and 50% toxicity, and a dose with 30% efficacy and 15% of toxicity are equally desirable. Based on these three equally desirable efficacy-toxicity probability pairs, the desirability function is defined as

$$\phi(p_E, p_T) = 1 - \left\{ \left(\frac{p_E - 1}{p_{E,1}^* - 1} \right)^r + \left(\frac{p_T - 0}{p_{T,2}^* - 0} \right)^r \right\}^{1/r}$$

where $r > 0$. We determine the value of r by solving equation

$$\phi\left(p_{E,3}^{*},\ p_{T,3}^{*}\right) = 0.$$

Once the value of r is determined, the family of efficacy-toxicity trade-off contours is determined by the above function. Specifically, for a given desirability δ, the efficacy-toxicity trade-off contour is defined as

$$C_{\delta} = \{(p_E, p_T):\ \phi(p_E, p_T) = \delta\}$$

That is, all (p_E, p_T) on C_{δ} have desirability δ.

Simulation configurations for the 3+3, CRM and EffTox Designs

There are many different 3+3 algorithms. In our simulation, we use the common version given in Table S3.

Table S3. A commonly used phase I trial 3+3 algorithm

General Rules	
1. Never re-escalate to a dose level after de-escalating from that level 2. If the decision is to de-escalate or choose one level lower but current level is lowest, stop and choose no level 3. If the decision is to escalate above highest level, stop and choose no level 4. If the decision is to stop and choose one level lower, but one level lower has 3 or fewer patients, treat 3 more at that lower level	
# toxicities/ # patients	**Decision**
0/3	Escalate one level, if allowed by General Rule 1, otherwise treat 3 more at current level.
0/3 + [0/3 or 1/3]	escalate one level, if allowed by General Rule 1, otherwise treat 3 more at the current level
0/3 + [2/3 or 3/3]	Stop, choose one level lower as MTD if allowed by General Rule 4
1/3	Treat 3 more at current level

1/3 + 0/3	Escalate one level if allowed by General Rule 1, otherwise choose current level as MTD if allowed by General Rule 4
1/3 + [1/3 or 2/3 or 3/3]	Stop, choose one level lower as MTD if allowed by General Rule 4
2/3 or 3/3	De-escalate one level

The CRM is based on the following power model for the probability of toxicity as a function of dose level,

$$p_{T(d=i)} = \pi_i^{\exp(\alpha)}$$

with the initial estimates of the toxicity rate (i.e., skeleton) $(\pi_1, \ldots, \pi_5) = (0.05,\ 0.15,\ 0.3,\ 0.45,\ 0.55)$ and prior $\alpha \sim N(0, 2)$. The CRM selects the MTD as the dose whose posterior estimate of the toxicity rate close to the target p^*=0.2. The skeleton typically is elicited from the clinician or clinicians planning the trial. In the case that the clinicians have limited prior information on the estimates of the toxicity rates, statistical methods are available to calibrate the initial estimate of the curve to obtain good operating characteristics (Cheung, 2011).

In the EffTox design, the upper limit on $p_T(d)$ was $A_T = 0.20$, lower limit on $p_E(d)$ was $A_E = 0.20$, and the equally desirable (p_E, p_T) pairs (.15, 0), (.25, .15) and (1, .70) were used to generate the target trade-off contour (see Figure 2). A dose is defined as acceptable if $\Pr(p_E(d) > A_E \mid \text{data}) > 0.10$ and $\Pr(p_T(d) < A_T \mid \text{data}) > 0.10$. For the 5 doses, the elicited prior means of $p_T(d)$ are (0.05, 0.1, 0.2, 0.3, 0.5), and the elicited prior means of $p_E(d)$ are (0.2, 0.4, 0.6, 0.8, 0.9), with prior effective sample size (ESS) of 0.9

patients. Simulations and trial conduct for the EffTox design were carried out using the desktop application EffTox version 4.0.12, which is freely available at the website https://biostatistics.mdanderson.org/softwaredownload/