

# Supplementary Materials for

## “Generalization of the time-to-event continual reassessment method to bivariate outcomes”

### 1 Additional simulation studies

In order to further examine the operating characteristics of the proposed method, we conducted additional simulation studies. There are 6 candidate doses from which the lowest safe dose that maximizes efficacy (i.e., optimal dose) is selected. We set the maximum sample size to be  $N = 60$  and the maximum toxicity tolerance to be  $\xi = 33\%$ . Under each scenario, each method is simulated for 1000 iterations and the average results are compared by the accuracy of optimal dose selection, number of patients treated at optimal doses and total trial duration.

Toxicity prior skeleton is chosen as  $(p_1, p_2, p_3, p_4, p_5, p_6) = (0.1, 0.20, 0.30, 0.40, 0.50, 0.60)$ .

$L = 11$  possible efficacy models are constructed as:

$$\mathbf{Q} = \begin{pmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \\ \mathbf{q}_3 \\ \mathbf{q}_4 \\ \mathbf{q}_5 \\ \mathbf{q}_6 \\ \mathbf{q}_7 \\ \mathbf{q}_8 \\ \mathbf{q}_9 \\ \mathbf{q}_{10} \\ \mathbf{q}_{11} \end{pmatrix} = \begin{pmatrix} \mathbf{0.60}, 0.50, 0.40, 0.30, 0.20, 0.10 \\ 0.50, \mathbf{0.60}, 0.50, 0.40, 0.30, 0.20 \\ 0.40, 0.50, \mathbf{0.60}, 0.50, 0.40, 0.30 \\ 0.30, 0.40, 0.50, \mathbf{0.60}, 0.50, 0.40 \\ 0.20, 0.30, 0.40, 0.50, \mathbf{0.60}, 0.50 \\ 0.10, 0.20, 0.30, 0.40, 0.50, \mathbf{0.60} \\ 0.20, 0.30, 0.40, 0.50, \mathbf{0.60}, \mathbf{0.60} \\ 0.30, 0.40, 0.50, \mathbf{0.60}, \mathbf{0.60}, \mathbf{0.60} \\ 0.40, 0.50, \mathbf{0.60}, \mathbf{0.60}, \mathbf{0.60}, \mathbf{0.60} \\ 0.50, \mathbf{0.60}, \mathbf{0.60}, \mathbf{0.60}, \mathbf{0.60}, \mathbf{0.60} \\ \mathbf{0.60}, \mathbf{0.60}, \mathbf{0.60}, \mathbf{0.60}, \mathbf{0.60}, \mathbf{0.60} \end{pmatrix}. \quad (1)$$

$\mathbf{q}_1$  through  $\mathbf{q}_6$  represent scenarios where the dose-efficacy peaked at dose  $d_1$  through  $d_6$ , respectively;  $\mathbf{q}_7$  through  $\mathbf{q}_{11}$  are scenarios when dose-efficacy plateaus after an intermediate dose. We assume no existing knowledge about the candidate models and set  $h(\ell) = \frac{1}{11}$  for  $\ell = 1, 2, \dots, 11$ .  $p(\ell) = \frac{1}{11}$  for  $\ell = 1, 2, \dots, 11$ . For all model parameters, for both toxicity and efficacy, we used  $\mathcal{N}(0, 1.34)$  prior distributions. The DLT observation window is  $\lambda_T = 4$  weeks and the efficacy observation window is  $\lambda_E = 12$  weeks.

## 1.1 TITE distributions

We assume the arrival of patients follows a Poisson distribution, with mean  $\mu = 0.25, 1$  or  $2$  denoting the expected number of patients per week.

The time-to-event of toxicity or efficacy response is assumed to follow either a conditional uniform distribution or a Weibull distribution. Since we do not know the underlying distri-

bution of the TITE variable, we simulate various truths for the TITE distribution. Note that ‘U’ denotes a conditionally Uniform distribution of the TITE variable; ‘W’ denotes a Weibull distribution. For a conditional uniform distribution, the time-to-event would be randomly generated on the interval  $(0, 4)$  for toxicity and  $(0, 12)$  for efficacy. For the Weibull model the shape parameter was fixed at a value of 4 and the scale parameters were chosen so that the cumulative distribution function at times  $\lambda_T = 4$  and  $\lambda_E = 12$  would be the probability of toxicity and efficacy, respectively, at each dose level. Both of these time-to-event models were used in Cheung and Chappell (2000). Therefore, for each scenario, we investigated four different specifications:

1. Tox-U, Eff-U: the patients time-to-event for both toxicity and efficacy were generated under a conditionally uniform model.
2. Tox-U, Eff-W: the patients time-to-toxicity were generated under a conditionally uniform model; the patients time-to-efficacy were generated under a Weibull model.
3. Tox-W, Eff-W: the patients time-to-event for both toxicity and efficacy were generated under a Weibull model.
4. Tox-W, Eff-U: the patients time-to-toxicity were generated under a Weibull model; the patients time-to-efficacy were generated under a conditionally uniform model.

## 1.2 Dose response scenarios

The dose-response scenarios are presented in Table 1. In scenario S1-S6, dose-toxicity is generally increasing but the overall toxicity rate is relatively low. Scenario S7 through S9 represents scenarios with high toxicity rate and the true MTD is dose 4. Dose-efficacy in scenarios S7 through S9 is constant, increasing, and decreasing respectively.

Table 1: *True probabilities of observing toxicity and efficacy responses at each dose level.*

Scenario		Dose level					
		1	2	3	4	5	6
S1	Toxicity	0.02	0.03	0.05	0.06	0.07	0.07
	Efficacy	0.40	0.25	0.20	0.15	0.10	0.05
S2	Toxicity	0.02	0.04	0.07	0.09	0.10	0.15
	Efficacy	0.15	0.45	0.30	0.25	0.16	0.10
S3	Toxicity	0.01	0.02	0.03	0.04	0.05	0.06
	Efficacy	0.10	0.15	0.35	0.18	0.12	0.07
S4	Toxicity	0.01	0.02	0.04	0.06	0.08	0.10
	Efficacy	0.05	0.15	0.30	0.45	0.35	0.30
S5	Toxicity	0.01	0.02	0.03	0.05	0.06	0.07
	Efficacy	0.15	0.25	0.33	0.47	0.60	0.40
S6	Toxicity	0.01	0.02	0.03	0.04	0.05	0.06
	Efficacy	0.05	0.15	0.30	0.35	0.40	0.50
S7	Toxicity	0.05	0.10	0.16	0.30	0.40	0.50
	Efficacy	0.30	0.30	0.30	0.30	0.30	0.30
S8	Toxicity	0.05	0.10	0.16	0.30	0.40	0.50
	Efficacy	0.10	0.20	0.30	0.40	0.50	0.60
S9	Toxicity	0.05	0.10	0.16	0.30	0.40	0.50
	Efficacy	0.50	0.40	0.30	0.20	0.10	0.05

### 1.3 Operating characteristics

Table 2 and Table 3 examine the operating characteristics of the proposed method regarding its ability to accurately select optimal doses as well as the number of patients treated at optimal doses. Overall, the proposed method showed encouraging selection accuracy under various dose-response scenarios and different underlying simulation assumptions. The chance of selecting optimal doses generally ranges from 50% to 70%, except for scenario S6 and S7 under which the chance is only about 30%. The proposed method showed robustness against varying TITE distribution as well as the rate at which patients are enrolled. Whether the TITE follows a Weibull or conditional Uniform distribution, the chance of selecting the optimal dose remains the same. Similarly, the rate of patient arrival do not diminish the ability to select the optimal dose.

The distribution of patient allocation centers at the optimal dose with an average of 25 patients treated at this dose. Note that even though the accuracy of optimal dose selection is not affected by patient arrival rate, the distribution of patient allocation significantly depends on how fast patients are enrolled. The slower patients are enrolled, the more likely they are treated at the optimal dose, as more information become available for each dose assignment.

We further analyze scenario S6 and scenario S7 as results from these two scenarios are less desired than the others. Table 4 is a subset of the simulation results detailing the dose selection and patient allocation distribution for scenario S6 and S7. In table 4, the patient enrollment follows a Poisson distribution with rate = 0.25. Toxicity outcomes are assessed within 4 weeks. In scenario S6, both toxicity and efficacy are monotonically increasing with the optimal dose being the highest dose. The results indicate that the proposed method is overly conservative and tends to select the highest 3 dose levels with similar probabilities. Patient allocation distribution centers at the next highest dose.

In scenario S7, all dose levels have the same efficacy probabilities while the MTD is at the 4<sup>th</sup> dose. The exact optimal dose is the first dose but dose 1-4 are also correct (i.e. they maximize efficacy and are also safe). Therefore, even though the probability of selecting the optimal dose is only about 30%, most of the time, the selected dose is still correct. Similarly, the vast majority of the participating patients are treated at an efficacious dose level with less than 33% DLT probability.

An advantage of the proposed method over that proposed by Wages and Tait (2015) is that the design significantly shortens the expected trial duration as dose assignment can be performed using partial information from patients who are still under observation. The trial duration shortens to approximately 5, 1.5 or 0.8 years when the rate of patient enrollment is respectively 0.25, 1, and 2 patients/week, as compared to 15 years if the next patient can only be enrolled when all current patients have completed follow up.

Table 2: *Simulation results: toxicity and efficacy outcomes are observed for 4 and 12 weeks, respectively. ‘U’ and ‘W’ respectively denotes a conditionally Uniform distribution and Weibull distribution of toxicity or efficacy TITE. The enrollment of patients follow a Poisson process, with Rate=0.25, 1, or 2. The estimated probability of selecting the optimal dose is summarized in column ‘Select’, and the average number of patients treat at the optimal dose is under column ‘Treat’.*

Tox	Eff	Rate	Select	Treat	Tox	Eff	Rate	Select	Treat	Tox	Eff	Rate	Select	Treat
Scenario 1					Scenario 2					Scenario 3				
U	U	0.25	67.6	28.8	U	U	0.25	69.3	27.4	U	U	0.25	69.3	26.9
U	U	1	72.4	28.8	U	U	1	67.9	24.3	U	U	1	73.3	26.7
U	U	2	75.2	26.9	U	U	2	62.3	20.9	U	U	2	70.6	23.7
U	W	0.25	66	28.3	U	W	0.25	66.9	26	U	W	0.25	71.7	27.6
U	W	1	72.6	26.6	U	W	1	69.5	23.6	U	W	1	71.3	24.6
U	W	2	76.3	23.7	U	W	2	66.1	18.5	U	W	2	71.3	21
W	U	0.25	67.9	28.9	W	U	0.25	66.6	26.4	W	U	0.25	71.6	28
W	U	1	71.7	28.4	W	U	1	66.4	23.9	W	U	1	72	26.3
W	U	2	73.6	26.6	W	U	2	63.5	20.6	W	U	2	69.6	23.4
W	W	0.25	72	29.5	W	W	0.25	67.3	25.8	W	W	0.25	68.5	26.4
W	W	1	73.4	26.8	W	W	1	68.9	22.6	W	W	1	75.1	25.7
W	W	2	76.9	23.1	W	W	2	66.4	18.9	W	W	2	71.2	20.7
Scenario 4					Scenario 5					Scenario 6				
U	U	0.25	59.2	24	U	U	0.25	59.7	24.6	U	U	0.25	29.4	12.5
U	U	1	58.8	22.2	U	U	1	63.9	24.1	U	U	1	38	14.1
U	U	2	59.6	20.8	U	U	2	63.5	21.5	U	U	2	40.3	13.4
U	W	0.25	61.3	24.5	U	W	0.25	60.6	25	U	W	0.25	30.9	12.6
U	W	1	60	22.1	U	W	1	62	22.4	U	W	1	35.1	12.3
U	W	2	57	18.5	U	W	2	61.7	19.1	U	W	2	41	11.4
W	U	0.25	63.4	24.7	W	U	0.25	60.3	25.1	W	U	0.25	29.4	12.6
W	U	1	60.1	22.2	W	U	1	62.4	23.7	W	U	1	35.8	13.8
W	U	2	61.7	20.2	W	U	2	62.5	21.4	W	U	2	40.3	14.1
W	W	0.25	62.9	24.6	W	W	0.25	61.8	24.9	W	W	0.25	31	12.5
W	W	1	59	21.9	W	W	1	62.4	22.6	W	W	1	36.9	12.6
W	W	2	59.4	18.8	W	W	2	63	19.1	W	W	2	44.8	12
Scenario 7					Scenario 8					Scenario 9				
U	U	0.25	30.2	15.7	U	U	0.25	47.4	21.8	U	U	0.25	71.8	31.4
U	U	1	32.5	16.1	U	U	1	48.4	21.5	U	U	1	69.4	29.3
U	U	2	33.9	16	U	U	2	48.8	20	U	U	2	73.8	28.5
U	W	0.25	27.9	14.6	U	W	0.25	46.8	21.5	U	W	0.25	65.9	29.1
U	W	1	33.2	15	U	W	1	48.9	20.9	U	W	1	72.5	28
U	W	2	35.4	15.1	U	W	2	50.2	19.7	U	W	2	74.8	24.4
W	U	0.25	28.6	15	W	U	0.25	45.7	21.4	W	U	0.25	65.3	29.6
W	U	1	30.7	14.6	W	U	1	49.2	22.1	W	U	1	69.7	28.8
W	U	2	33	15	W	U	2	48.1	21.7	W	U	2	69.5	25.9
W	W	0.25	27.9	14.8	W	W	0.25	43.6	21	W	W	0.25	68.7	30.8
W	W	1	31.9	14.5	W	W	1	47.7	21.6	W	W	1	72.8	27.7
W	W	2	36.6	13.8	W	W	2	50.2	20.5	W	W	2	73.1	23.3

Table 3: Simulation results: toxicity and efficacy outcomes are both observed 12 weeks. ‘U’ and ‘W’ respectively denotes a conditionally Uniform distribution and Weibull distribution of toxicity or efficacy TITE. The enrollment of patients follow a Poisson process, with Rate=0.25, 1, or 2. The estimated probability of selecting the optimal dose is summarized in column ‘Select’, and the average number of patients treat at the optimal dose is under column ‘Treat’.

Tox	Eff	Rate	Select	Treat	Tox	Eff	Rate	Select	Treat	Tox	Eff	Rate	Select	Treat
Scenario 1					Scenario 2					Scenario 3				
U	U	0.25	69.1	29.6	U	U	0.25	68.3	26.6	U	U	0.25	70.7	27.8
U	U	1	73.6	29.2	U	U	1	64.8	23.5	U	U	1	72.9	27.4
U	U	2	71.6	25.7	U	U	2	67.1	21.4	U	U	2	71.7	25.4
U	W	0.25	70.1	29.3	U	W	0.25	67.2	26.2	U	W	0.25	72.8	28.1
U	W	1	73.4	26.7	U	W	1	67.2	22.8	U	W	1	75.9	27.3
U	W	2	77.2	23.8	U	W	2	67.5	19	U	W	2	72.5	22.8
W	U	0.25	68	29.1	W	U	0.25	66.5	26	W	U	0.25	69.8	27.6
W	U	1	68.8	27.2	W	U	1	63.9	22.8	W	U	1	78.6	29
W	U	2	71.3	26.1	W	U	2	65.7	20.6	W	U	2	72.6	25.1
W	W	0.25	67.7	28.3	W	W	0.25	68.4	26.2	W	W	0.25	69.3	27.2
W	W	1	73.2	26.9	W	W	1	70.6	23.1	W	W	1	74.2	26.4
W	W	2	74.4	22.5	W	W	2	67.3	17.8	W	W	2	73.8	23
Scenario 4					Scenario 5					Scenario 6				
U	U	0.25	59.2	24.2	U	U	0.25	64.6	26.1	U	U	0.25	31.4	12.5
U	U	1	60.4	22.4	U	U	1	60	22.5	U	U	1	34.1	12.8
U	U	2	61.2	20.2	U	U	2	62.5	20.6	U	U	2	39.2	12.1
U	W	0.25	60.7	24.4	U	W	0.25	60.6	24.6	U	W	0.25	30.7	12.4
U	W	1	61.3	21.8	U	W	1	56.5	20.4	U	W	1	34.2	11
U	W	2	57.9	18.9	U	W	2	61.1	16.8	U	W	2	39.8	10.1
W	U	0.25	61.9	24.5	W	U	0.25	60.2	24.8	W	U	0.25	31.7	12.9
W	U	1	61.5	22.7	W	U	1	59.7	22.9	W	U	1	32.3	12.4
W	U	2	57.7	20.2	W	U	2	61.6	19.9	W	U	2	39	12.9
W	W	0.25	62.3	24.6	W	W	0.25	61.1	24.8	W	W	0.25	31.8	12.7
W	W	1	59.9	21.6	W	W	1	62	21.4	W	W	1	37.1	11.7
W	W	2	58.5	18.7	W	W	2	63.7	17.8	W	W	2	42.3	10.8
Scenario 7					Scenario 8					Scenario 9				
U	U	0.25	29.9	15.2	U	U	0.25	46.8	21.5	U	U	0.25	66.1	29.7
U	U	1	31.8	16.8	U	U	1	52.5	20.7	U	U	1	70.6	30
U	U	2	34.6	16.5	U	U	2	49	17.9	U	U	2	74.2	28.8
U	W	0.25	27.5	15.1	U	W	0.25	48.8	21.2	U	W	0.25	67.5	30
U	W	1	31.6	15.5	U	W	1	52.2	19.9	U	W	1	72.4	28.6
U	W	2	40.2	16.7	U	W	2	52.4	16.9	U	W	2	77.6	25.9
W	U	0.25	27.4	14.7	W	U	0.25	44.9	22	W	U	0.25	67.7	30.4
W	U	1	31.8	15.1	W	U	1	45.6	21.2	W	U	1	70.4	28.7
W	U	2	29.9	13.4	W	U	2	44.9	19	W	U	2	71.2	26.5
W	W	0.25	27.4	14.1	W	W	0.25	45.1	21.4	W	W	0.25	69.7	30.7
W	W	1	29.3	13.8	W	W	1	49.8	22	W	W	1	72.7	27.3
W	W	2	36	13.5	W	W	2	49.9	18.6	W	W	2	73	23.4



Table 4: Distribution of dose selection and patient allocation to each dose level after 1000 iterations of simulation. ‘U’ and ‘W’ respectively denotes a conditionally Uniform distribution and Weibull distribution of the toxicity and efficacy TITE variables. The enrollment of patients follows a Poisson process, with Rate=0.25. Maximum sample size is  $N = 60$ . The maximum tolerated DLT rate is set at  $\xi = 33\%$ . Dose-response scenarios are detailed in Table 1.

Scenario	Tox	Eff	Dose 1	Dose 2	Dose 3	Dose 4	Dose 5	Dose 6
Percentage of final selection								
6	U	W	0.5	2.4	20	22.1	24.3	30.7
6	U	U	0.9	2.4	17.5	24.1	23.7	31.4
6	W	U	0.3	2.8	17.6	22.8	24.8	31.7
6	W	W	0.6	2.7	17.9	22.6	24.4	31.8
Average number of patient treated								
6	U	W	2.5	4.6	11.5	13.6	15.4	12.4
6	U	U	2.6	4.3	10.8	14.0	15.9	12.5
6	W	U	2.4	4.4	11.06	13.3	15.9	12.9
6	W	W	2.5	4.3	10.9	13.9	15.7	12.7
Percentage of final selection								
7	U	W	27.5	27	26.9	17.2	1.4	0
7	U	U	29.9	24.4	28.7	16.2	0.8	0
7	W	U	27.4	25.6	30.6	15.5	0.9	0
7	W	W	27.4	24.4	30.4	16.3	1.5	0
Average number of patient treated								
7	U	W	15.1	14.9	16.4	11.0	2.3	0.3
7	U	U	15.5	15.0	17.0	10.2	2.4	0.3
7	W	U	14.7	14.6	17.0	10.8	2.5	0.3
7	W	W	14.1	14.5	17.2	10.9	3.0	0.4