# Supplementary Material to "A Utility-based Bayesian Optimal Interval (U-BOIN) Phase I/II Design to Identify the Optimal Biological Dose for Targeted and Immune Therapies"

This supplementary material is intended to be read in conjunction with the full article titled "A Utility-based Bayesian Optimal Interval (U-BOIN) Phase I/II Design to Identify the Optimal Biological Dose for Targeted and Immune Therapies."

## 1 | AN EXAMPLE TO SHOW HOW TO USE DECISION TABLES IN THE U-BOIN DESIGN

The U-BOIN design is a model-assisted phase I/II design. Its decision tables can be pre-tabulated to facilitate the conduct of a trial. Two decision tables are provided in the U-BOIN design: one is for dose escalation/de-escalation boundaries, mainly for use in stage I and step B1 in stage II, and the other table is for utility values, for use in stage II to determine which doses are admissible, as well as which dose the next cohort of patients will receive. We give an example to show how to use the U-BOIN design with a hypothetical phase I/II trial.

We consider a trial with binary efficacy and toxicity outcomes. Five doses are examined for the trial. Suppose the maximum acceptable dose limiting toxicity (DLT) rate is 0.3, the minimum acceptable efficacy rate is 0.2, and the utility value for the joint outcomes $(Y_E, Y_T) = \{(0,1),(0,0),(1,1),(1,0)\}$ is $\{0, 15, 25, 100\}$.

At an interim in stage I, we have the following data as shown in Table S1, where the current dose level is 3. Note that efficacy data also is collected, but not used in this stage. Given the observed data, we make dose assignment based on the decision rules in Table S2, where the escalation and de-escalation boundaries are calculated using $\pi_T^* = 0.25$. Since the current dose level is 3, on which 6 patients are treated, we check the column where the number of patients treated is 6 in Table S2. The decision rule shows that we should escalate the dose to dose level 4 when the number of DLT is 1 or less out of 6 observations.

**TABLE S1** Data available at an interim in stage I of the trial example

| | Dose level | | | | |
|---|---|---|---|---|---|
| Outcome | 1 | 2 | **3** | 4 | 5 |
| No. of patients treated | 3 | 3 | **6** | 0 | 0 |
| No. of DLT | 0 | 0 | **1** | 0 | 0 |

**TABLE S2** Escalation and de-escalation boundaries in the trial example

| Decision | Number of patients treated on a dose | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | **6** | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| Escalate if No. of DLT <= | 0 | 0 | 0 | 0 | 0 | **1** | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 4 | 4 | 4 | 4 |
| Deescalate if No. of DLT >= | 1 | 1 | 1 | 2 | 2 | **2** | 3 | 3 | 3 | 3 | 4 | 4 | 4 | 5 | 5 | 5 | 6 | 6 | 6 | 6 | 7 | 7 | 7 | 8 |
| Eliminate if No. of DLT >= | NA | NA | 3 | 3 | 4 | **4** | 5 | 5 | 5 | 6 | 6 | 7 | 7 | 8 | 8 | 8 | 9 | 9 | 9 | 10 | 10 | 11 | 11 | 11 |

Suppose at another interim time in stage II, the data for the five tried doses are shown in rows 1-4 in Table S3. We use Tables S4-S6 to determine which dose to use to treat the next cohort of patients, because theses tables show the estimated utility for a dose when the number of patients treated on the dose is 3, 6, and 9, respectively. The values in these tables are determined using the admissible criteria with the probability cutoff $C_T = 0.95$, $C_E = 0.9$, and the utility values $(0, 15, 25, 100)$. We can use Table

S4 to find utility values for dose level 1 and 2 because both doses have treated three patients. By the same token, use Table S5 for dose level 3 and 5, and Table S6 for level 4. The utility values found for the doses are bolded in Table S3. As shown, dose level 1 is not admissible. Among the admissible doses $2-5$, dose level 4 has the largest utility. Thus, the next cohort of patients will receive dose level 4. The online App will provide the decision tables, once the design parameters are input.

**TABLE S3** Data available at an interim monitoring in stage II of the trial example

| | Dose level | | | | |
|---|---|---|---|---|---|
| Outcome | 1 | 2 | 3 | 4 | 5 |
| No. of patients treated | 3 | 3 | 6 | 9 | 6 |
| No. of Efficacy | 0 | 1 | 2 | 5 | 3 |
| No. of DLT | 0 | 0 | 1 | 2 | 2 |
| No. of (Efficacy, No DLT) | 0 | 1 | 1 | 4 | 2 |
| **Utility** | **0** | **41.2** | **31.4** | **50.5** | **41.4** |

**TABLE S4** Utility given the sample size on a dose is 3

| No.Eff | No.Tox | No.(Eff=1, Tox=0) | Utility | No.Eff | No.Tox | No.(Eff=1 Tox=0) | Utility | No.Eff | No.Tox | No.(Eff=1 Tox=0) | Utility |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **<1** | **Any** | **Any** | **0** | 1 | > 2 | Any | 0 | 2 | > 2 | Any | 0 |
| **1** | **0** | **1** | **41.2** | 2 | 0 | 2 | 62.5 | 3 | 0 | 3 | 83.8 |
| 1 | 1 | 0 | 22.5 | 2 | 1 | 1 | 43.8 | 3 | 1 | 2 | 65.0 |
| 1 | 1 | 1 | 37.5 | 2 | 1 | 2 | 58.8 | 3 | 2 | 1 | 46.2 |
| 1 | 2 | 0 | 18.8 | 2 | 2 | 0 | 25.0 | 3 | > 2 | Any | 0 |
| 1 | 2 | 1 | 33.8 | 2 | 2 | 1 | 40.0 | | | | |

Note: "No.Eff" is the number of responses, "No. Tox" the number of toxicity, "No.(Eff=1,Tox=0)" the number of outcomes with response, but no toxicity. A utility value of 0 indicates that the dose is not admissible. Bolded text corresponds to the dose information and utility for dose level 1 and 2 in the example in this section.

**TABLE S5** Utility given the sample size on a dose is 6

| No.Eff | No.Tox | No.(Eff=1, Tox=0) | Utility | No.Eff | No.Tox | No.(Eff=1 Tox=0) | Utility | No.Eff | No.Tox | No.(Eff=1 Tox=0) | Utility |
|---|---|---|---|---|---|---|---|---|---|---|---|
| <1 | Any | Any | 0 | 2 | > 3 | Any | 0 | 4 | 3 | 1 | 34.3 |
| 1 | 0 | 1 | 30.0 | 3 | 0 | 3 | 54.3 | 4 | 3 | 2 | 42.9 |
| 1 | 1 | 0 | 19.3 | 3 | 1 | 2 | 43.6 | 4 | 3 | 3 | 51.4 |
| 1 | 1 | 1 | 27.9 | 3 | 1 | 3 | 52.1 | 4 | > 3 | Any | 0 |
| 1 | 2 | 0 | 17.1 | 3 | 2 | 1 | 32.9 | 5 | 0 | 5 | 78.6 |
| 1 | 2 | 1 | 25.7 | **3** | **2** | **2** | **41.4** | 5 | 1 | 4 | 67.9 |
| 1 | 3 | 0 | 15.0 | 3 | 2 | 3 | 50.0 | 5 | 1 | 5 | 76.4 |
| 1 | 3 | 1 | 23.6 | 3 | 3 | 0 | 22.1 | 5 | 2 | 3 | 57.1 |
| 1 | > 3 | Any | 0 | 3 | 3 | 1 | 30.7 | 5 | 2 | 4 | 65.7 |
| 2 | 0 | 2 | 42.1 | 3 | 3 | 2 | 39.3 | 5 | 3 | 2 | 46.4 |
| **2** | **1** | **1** | **31.4** | 3 | 3 | 3 | 47.9 | 5 | 3 | 3 | 55.0 |
| 2 | 1 | 2 | 40.0 | 3 | > 3 | Any | 0 | 5 | > 3 | Any | 0 |
| 2 | 2 | 0 | 20.7 | 4 | 0 | 4 | 66.4 | 6 | 0 | 6 | 90.7 |
| 2 | 2 | 1 | 29.3 | 4 | 1 | 3 | 55.7 | 6 | 1 | 5 | 80.0 |
| 2 | 2 | 2 | 37.9 | 4 | 1 | 4 | 64.3 | 6 | 2 | 4 | 69.3 |
| 2 | 3 | 0 | 18.6 | 4 | 2 | 2 | 45.0 | 6 | 3 | 3 | 58.6 |
| 2 | 3 | 1 | 27.1 | 4 | 2 | 3 | 53.6 | 6 | > 3 | Any | 0 |
| 2 | 3 | 2 | 35.7 | 4 | 2 | 4 | 62.1 | | | | |

Note: "No. Eff" is the number of responses, "No. Tox" the number of toxicity, "No.(Eff=1,Tox=0)" the number of outcomes with response, but no toxicity. A utility value of 0 indicates that the dose is not admissible. Bolded text corresponds to the dose information and utility for dose level 3 and 5 in the example in this section.

**TABLE S6** Utility given the sample size on a dose is 9

| No.Eff | No.Tox | No.(Eff=1, Tox=0) | Utility | No.Eff | No.Tox | No.(Eff=1 Tox=0) | Utility | No.Eff | No.Tox | No.(Eff=1 Tox=0) | Utility |
|---|---|---|---|---|---|---|---|---|---|---|---|
| <1 | Any | Any | 0 | 3 | > 4 | Any | 0 | 6 | 3 | 3 | 45.5 |
| 1 | 0 | 1 | 25.5 | 4 | 0 | 4 | 51.0 | 6 | 3 | 4 | 51.5 |
| 1 | 1 | 0 | 18.0 | 4 | 1 | 3 | 43.5 | 6 | 3 | 5 | 57.5 |
| 1 | 1 | 1 | 24.0 | 4 | 1 | 4 | 49.5 | 6 | 3 | 6 | 63.5 |
| 1 | 2 | 0 | 16.5 | 4 | 2 | 2 | 36.0 | 6 | 4 | 2 | 38.0 |
| 1 | 2 | 1 | 22.5 | 4 | 2 | 3 | 42.0 | 6 | 4 | 3 | 44.0 |
| 1 | 3 | 0 | 15.0 | 4 | 2 | 4 | 48.0 | 6 | 4 | 4 | 50.0 |
| 1 | 3 | 1 | 21.0 | 4 | 3 | 1 | 28.5 | 6 | 4 | 5 | 56.0 |
| 1 | 4 | 0 | 13.5 | 4 | 3 | 2 | 34.5 | 6 | > 4 | Any | 0 |
| 1 | 4 | 1 | 19.5 | 4 | 3 | 3 | 40.5 | 7 | 0 | 7 | 76.5 |
| 1 | > 4 | Any | 0 | 4 | 3 | 4 | 46.5 | 7 | 1 | 6 | 69.0 |
| 2 | 0 | 2 | 34.0 | 4 | 4 | 0 | 21.0 | 7 | 1 | 7 | 75.0 |
| 2 | 1 | 1 | 26.5 | 4 | 4 | 1 | 27.0 | 7 | 2 | 5 | 61.5 |
| 2 | 1 | 2 | 32.5 | 4 | 4 | 2 | 33.0 | 7 | 2 | 6 | 67.5 |
| 2 | 2 | 0 | 19.0 | 4 | 4 | 3 | 39.0 | 7 | 2 | 7 | 73.5 |
| 2 | 2 | 1 | 25.0 | 4 | 4 | 4 | 45.0 | 7 | 3 | 4 | 54.0 |
| 2 | 2 | 2 | 31.0 | 4 | > 4 | Any | 0 | 7 | 3 | 5 | 60.0 |
| 2 | 3 | 0 | 17.5 | 5 | 0 | 5 | 59.5 | 7 | 3 | 6 | 66.0 |
| 2 | 3 | 1 | 23.5 | 5 | 1 | 4 | 52.0 | 7 | 4 | 3 | 46.5 |
| 2 | 3 | 2 | 29.5 | 5 | 1 | 5 | 58.0 | 7 | 4 | 4 | 52.5 |
| 2 | 4 | 0 | 16.0 | 5 | 2 | 3 | 44.5 | 7 | 4 | 5 | 58.5 |
| 2 | 4 | 1 | 22.0 | **5** | **2** | **4** | **50.5** | 7 | > 4 | Any | 0 |
| 2 | 4 | 2 | 28.0 | 5 | 2 | 5 | 56.5 | 8 | 0 | 8 | 85.0 |
| 2 | > 4 | Any | 0 | 5 | 3 | 2 | 37.0 | 8 | 1 | 7 | 77.5 |
| 3 | 0 | 3 | 42.5 | 5 | 3 | 3 | 43.0 | 8 | 1 | 8 | 83.5 |
| 3 | 1 | 2 | 35.0 | 5 | 3 | 4 | 49.0 | 8 | 2 | 6 | 70.0 |
| 3 | 1 | 3 | 41.0 | 5 | 3 | 5 | 55.0 | 8 | 2 | 7 | 76.0 |
| 3 | 2 | 1 | 27.5 | 5 | 4 | 1 | 29.5 | 8 | 3 | 5 | 62.5 |
| 3 | 2 | 2 | 33.5 | 5 | 4 | 2 | 35.5 | 8 | 3 | 6 | 68.5 |
| 3 | 2 | 3 | 39.5 | 5 | 4 | 3 | 41.5 | 8 | 4 | 4 | 55.0 |
| 3 | 3 | 0 | 20.0 | 5 | 4 | 4 | 47.5 | 8 | 4 | 5 | 61.0 |
| 3 | 3 | 1 | 26.0 | 5 | 4 | 5 | 53.5 | 8 | > 4 | Any | 0 |
| 3 | 3 | 2 | 32.0 | 5 | > 4 | Any | 0 | 9 | 0 | 9 | 93.5 |
| 3 | 3 | 3 | 38.0 | 6 | 0 | 6 | 68.0 | 9 | 1 | 8 | 86.0 |
| 3 | 4 | 0 | 18.5 | 6 | 1 | 5 | 60.5 | 9 | 2 | 7 | 78.5 |
| 3 | 4 | 1 | 24.5 | 6 | 1 | 6 | 66.5 | 9 | 3 | 6 | 71.0 |
| 3 | 4 | 2 | 30.5 | 6 | 2 | 4 | 53.0 | 9 | 4 | 5 | 63.5 |
| 3 | 4 | 3 | 36.5 | 6 | 2 | 5 | 59.0 | 9 | > 4 | Any | 0 |

Note: "No. Eff" is the number of responses, "No. Tox" the number of toxicity, "No.(Eff=1,Tox=0)" the number of outcomes with response, but no toxicity. A utility value of 0 indicates that the dose is not admissible. Bolded text corresponds to the dose information and utility for dose level 4 in the example in this section.

## 2 | FROM THE TRADE-OFF BASED ON MARGINAL TOXICITY AND EFFICACY TO THE TRADE-OFF BASED ON UTILITY FUNCTION

We provide an example here to show how to get the values of $(\psi_2, \psi_3)$, given a pre-specified value for $w$.

Let $\pi = (\pi_1, \pi_2, \pi_3, \pi_4)$ denote the probability for the joint outcomes $Y = \{1, 2, 3, 4\} = \{(Y_E = 0, Y_T = 1), (Y_E = 0, Y_T = 0), (Y_E = 1, Y_T = 1), (Y_E = 1, Y_T = 0)\}$. Then the marginal DLT rate and efficacy rate are $\pi_T = \pi_1 + \pi_3$ and $\pi_E = \pi_3 + \pi_4$. According to Theorem 1, we have

$$\psi_2\pi_2 + \psi_3\pi_3 + 100\pi_4 = \xi(\pi_E - w\pi_T). \tag{1}$$

The following procedure can be used to find the values for $(\psi_2, \psi_3)$, if $w$ is pre-specified.

Step 1. Make sure the pre-specified $w$ is within its plausible range, given the probabilities of outcomes.
Note that the utility value in this article is non-negative by definition. Since $\xi > 0$, we have

$$w < \pi_E/\pi_T. \tag{2}$$

Step 2. Find the plausible range for $\xi$ given $w$.
As $0 \le \psi_2, \psi_3 \le 100$, it follows that

$$\frac{100\pi_4}{\pi_E - w\pi_T} \le \xi \le \frac{100(\pi_2 + \pi_3 + \pi_4)}{\pi_E - w\pi_T}. \tag{3}$$

Step 3. Determine the values for $(\psi_2, \psi_3)$ given $w$ and $\xi$.
Given $w$ satisfying (2), we can find the range for $\xi$ using (3). For each $w$ and $\xi$, we can search for the values of $(\psi_2, \psi_3)$. Given $w = 0.5$, $\pi = (0.1, 0.6, 0.2, 0.1)$, Table S7 gives the values for a grid of $\xi$ that satisfy the inequality in (3) and the corresponding value of $(\psi_2, \psi_3)$. In practice, the search can be simple, as we might have a range of plausible values for $\psi_2$ or $\psi_3$. For example, if we set $\psi_2 = 25$, then the possible values for $\psi_3$ are $(9, 30, 51, 72, 93)$.

**TABLE S7** Some values for $\xi$ and $(\psi_2, \psi_3)$ given $w$ and $\pi$

| $\xi$ | $\psi_2$ | $\psi_3$ | $\xi$ | $\psi_2$ | $\psi_3$ | $\xi$ | $\psi_2$ | $\psi_3$ |
|---|---|---|---|---|---|---|---|---|
| 95 | 5 | 6 | 263 | 35 | 42 | 432 | 65 | 79 |
| 123 | 5 | 27 | 291 | 35 | 63 | 460 | 65 | 100 |
| 151 | 5 | 48 | 319 | 35 | 84 | 375 | 75 | 7 |
| 179 | 5 | 69 | 263 | 45 | 12 | 404 | 75 | 28 |
| 207 | 5 | 90 | 291 | 45 | 33 | 432 | 75 | 49 |
| 151 | 15 | 18 | 319 | 45 | 54 | 460 | 75 | 70 |
| 179 | 15 | 39 | 347 | 45 | 76 | 488 | 75 | 91 |
| 207 | 15 | 60 | 375 | 45 | 97 | 432 | 85 | 19 |
| 235 | 15 | 81 | 291 | 55 | 3 | 460 | 85 | 40 |
| 179 | 25 | 9 | 319 | 55 | 24 | 488 | 85 | 61 |
| 207 | 25 | 30 | 347 | 55 | 46 | 516 | 85 | 82 |
| 235 | 25 | 51 | 375 | 55 | 67 | 460 | 95 | 10 |
| 263 | 25 | 72 | 404 | 55 | 88 | 488 | 95 | 31 |
| 291 | 25 | 93 | 347 | 65 | 16 | 516 | 95 | 52 |
| 207 | 35 | 0 | 375 | 65 | 37 | 544 | 95 | 73 |
| 235 | 35 | 21 | 404 | 65 | 58 | 572 | 95 | 94 |

## 3 | PERFORMANCE OF THE U-BOIN WHEN SAMPLE SIZE IS SMALL

To examine the performance of the U-BOIN when there are delayed outcomes at a relatively smaller sample size, we conduct another simulation using $N = 39$ and $s_1 = 9$. We present the percentage of correct selection and the number of patients treated for the following designs: EffTox, U-BOIN without delayed outcome (U-BOIN-CD), and U-BOIN with delayed response that is

predicted using multiple imputation (U-BOIN-MI) in Table S8. Results show that U-BOIN still maintains its superior operating characteristics, even when the sample size is relatively small. In contrast, EffTox performs much worse when the sample size is small.

**TABLE S8** Results of simulation when the sample size is 39. Operating characteristics include the selection percentage (selection %), the average number of patients treated at each dose (No. of patients), and the percentage of early stopping. The optimal biological dose (OBD) is bolded. In scenario 8, the OBD does not exist, and thus the percentage of early stopping is bolded.

| Design | | Dose Level 1 | 2 | 3 | 4 | 5 | % of early stopping | |
|---|---|---|---|---|---|---|---|---|
| | | | | **Scenario 1** | | | | |
| | DLT rate | 0.02 | **0.15** | 0.30 | 0.45 | 0.60 | | |
| | Efficacy rate | 0.20 | **0.65** | 0.65 | 0.65 | 0.65 | | |
| | Utility | 43.0 | **69.0** | 63.0 | 56.0 | 50.0 | | |
| EffTox | Selection % | 3.0 | **51.0** | 44.0 | 2.0 | 0.0 | 0.0 | 28.5 |
| | No. of patients | 4.1 | **15.4** | 17.0 | 2.2 | 0.3 | | |
| U-BOIN-CD | Selection % | 2.2 | **73.1** | 21.4 | 3.1 | 0.2 | 0.2 | 28.6 |
| | No. of patients | 6.1 | **19.9** | 9.9 | 2.9 | 0.4 | | |
| U-BOIN -MI | Selection % | 1.7 | **68.0** | 24.3 | 3.8 | 0.1 | 2.2 | 15.9 |
| | No. of patients | 6.1 | **19.1** | 10.1 | 3.0 | 0.4 | | |
| | | | | **Scenario 2** | | | | |
| | DLT rate | 0.03 | 0.08 | **0.15** | 0.28 | 0.40 | | |
| | Efficacy rate | 0.10 | 0.22 | **0.60** | 0.60 | 0.60 | | |
| | Utility | 36.0 | 43.0 | **66.0** | 60.0 | 55.0 | | |
| EffTox | Selection % | 0.0 | 4.0 | **58.0** | 30.0 | 9.0 | 0.0 | 26.2 |
| | No. of patients | 3.3 | 4.2 | **17.3** | 10.2 | 3.9 | | |
| U-BOIN-CD | Selection % | 2.3 | 4.2 | **63.7** | 24.3 | 5.1 | 0.7 | 26.6 |
| | No. of patients | 4.7 | 6.8 | **15.2** | 9.3 | 3.5 | | |
| U-BOIN -MI | Selection % | 0.2 | 0.2 | **62.2** | 23.1 | 5.8 | 8.6 | 15.6 |
| | No. of patients | 4.4 | 6.2 | **14.9** | 8.7 | 3.4 | | |
| | | | | **Scenario 3** | | | | |
| | DLT rate | 0.05 | **0.15** | 0.30 | 0.45 | 0.60 | | |
| | Efficacy rate | 0.08 | **0.46** | 0.25 | 0.20 | 0.10 | | |
| | Utility | 34.0 | **56.0** | 37.0 | 29.0 | 18.0 | | |
| EffTox | Selection % | 16.0 | **62.0** | 13.0 | 2.0 | 1.0 | 6.0 | 28.5 |
| | No. of patients | 8.9 | **15.8** | 9.8 | 2.5 | 0.9 | | |
| U-BOIN-CD | Selection % | 3.3 | **87.6** | 6.0 | 0.5 | 0.1 | 2.7 | 29.0 |
| | No. of patients | 6.6 | **20.9** | 8.2 | 2.6 | 0.4 | | |
| U-BOIN -MI | Selection % | 1.5 | **86.9** | 7.3 | 0.7 | 0.2 | 3.4 | 15.6 |
| | No. of patients | 6.4 | **20.6** | 8.3 | 2.6 | 0.4 | | |
| | | | | **Scenario 4** | | | | |
| | DLT rate | 0.15 | **0.25** | 0.40 | 0.45 | 0.50 | | |
| | Efficacy rate | 0.15 | **0.45** | 0.30 | 0.25 | 0.20 | | |
| | Utility | 36.0 | **52.0** | 36.0 | 32.0 | 27.0 | | |
| EffTox | Selection % | 38.0 | **43.0** | 8.0 | 2.0 | 3.0 | 7.0 | 31.3 |
| | No. of patients | 13.4 | **13.3** | 6.6 | 2.4 | 1.9 | | |
| U-BOIN-CD | Selection % | 15.7 | **71.0** | 4.4 | 0.9 | 0.0 | 8.2 | 31.1 |
| | No. of patients | 12.9 | **18.7** | 4.5 | 1.1 | 0.2 | | |
| U-BOIN -MI | Selection % | 8.9 | **66.8** | 5.1 | 0.8 | 0.1 | 18.3 | 14.7 |
| | No. of patients | 11.9 | **17.6** | 4.3 | 0.9 | 0.2 | | |
| | | | | **Scenario 5** | | | | |
| | DLT rate | **0.10** | 0.30 | 0.50 | 0.55 | 0.65 | | |
| | Efficacy rate | **0.45** | 0.45 | 0.45 | 0.45 | 0.45 | | |
| | Utility | **58.0** | 50.0 | 42.0 | 40.0 | 36.0 | | |
| EffTox | Selection % | **66.0** | 28.0 | 4.0 | 0.0 | 0.0 | 1.0 | 32.2 |
| | No. of patients | **20.1** | 12.9 | 4.4 | 0.9 | 0.3 | | |
| U-BOIN-CD | Selection % | **73.5** | 22.7 | 2.8 | 0.7 | 0.0 | 0.4 | 32.5 |

**Table S8 Continued:**

| Design | | Dose Level 1 | 2 | 3 | 4 | 5 | % of early stopping | |
|---|---|---|---|---|---|---|---|---|
| | No. of patients | **22.5** | 12.8 | 3.2 | 0.4 | 0.0 | | |
| U-BOIN -MI | Selection % | **67.2** | 26.7 | 3.2 | 0.2 | 0.0 | 2.7 | 15.8 |
| | No. of patients | **21.6** | 13.1 | 3.3 | 0.4 | 0.0 | | |
| | | | | **Scenario 6** | | | | |
| | DLT rate | 0.05 | 0.07 | 0.10 | 0.12 | **0.16** | | |
| | Efficacy rate | 0.35 | 0.45 | 0.50 | 0.55 | **0.75** | | |
| | Utility | 53.0 | 59.0 | 61.0 | 64.0 | **75.0** | | |
| EffTox | Selection % | 12.0 | 11.0 | 28.0 | 23.0 | **26.0** | 0.0 | 26.1 |
| | No. of patients | 7.3 | 6.4 | 10.5 | 7.5 | **7.3** | | |
| U-BOIN-CD | Selection % | 5.8 | 13.0 | 13.7 | 15.8 | **51.7** | 0.1 | 26.7 |
| | No. of patients | 5.5 | 7.0 | 7.2 | 7.2 | **12.9** | | |
| U-BOIN -MI | Selection % | 6.4 | 12.8 | 14.1 | 15.8 | **50.5** | 0.3 | 16.0 |
| | No. of patients | 5.8 | 7.0 | 7.2 | 7.4 | **11.8** | | |
| | | | | **Scenario 7** | | | | |
| | DLT rate | 0.03 | **0.08** | **0.25** | 0.40 | 0.55 | | |
| | Efficacy rate | 0.05 | **0.25** | **0.35** | 0.40 | 0.45 | | |
| | Utility | 33.0 | **45.0** | **45.0** | 43.0 | 40.0 | | |
| EffTox | Selection % | 2.0 | **30.0** | **56.0** | 8.0 | 2.0 | 2.0 | 28.3 |
| | No. of patients | 4.0 | **9.3** | **18.3** | 5.2 | 1.8 | | |
| U-BOIN-CD | Selection % | 3.3 | **43.7** | **39.3** | 11.5 | 1.3 | 1.0 | 28.7 |
| | No. of patients | 5.0 | **13.9** | **13.7** | 5.5 | 1.1 | | |
| U-BOIN -MI | Selection % | 1.1 | **40.2** | **40.8** | 13.2 | 1.8 | 3.1 | 15.7 |
| | No. of patients | 4.6 | **14.4** | **13.3** | 5.2 | 1.0 | | |
| | | | | **Scenario 8** | | | | |
| | DLT rate | 0.22 | 0.45 | 0.55 | 0.65 | 0.70 | | |
| | Efficacy rate | 0.03 | 0.10 | 0.20 | 0.35 | 0.40 | | |
| | Utility | 25.0 | 23.0 | 25.0 | 30.0 | 31.0 | | |
| EffTox | Selection % | 1.0 | 8.0 | 8.0 | 1.0 | 0.0 | **82.0** | 15.1 |
| | No. of patients | 4.6 | 5.6 | 6.0 | 2.3 | 1.3 | | |
| U-BOIN-CD | Selection % | 4.1 | 10.5 | 3.1 | 0.1 | 0.0 | **82.3** | 18.3 |
| | No. of patients | 13.8 | 8.7 | 1.3 | 0.1 | 0.0 | | |
| U-BOIN -MI | Selection % | 0.0 | 0.5 | 1.0 | 0.2 | 0.0 | **98.2** | 8.7 |
| | No. of patients | 11.2 | 4.8 | 1.2 | 0.1 | 0.0 | | |

# 4 | ANOTHER EIGHT SCENARIOS CONSIDERED FOR SIMULATION A AND B

## 4.1 | Simulation A

We also consider eight additional representative scenarios that differ in the shape of the dose-toxicity and dose-efficacy curves, as well as the location of the optimal biological dose (OBD). Scenario A1 and A2 are cases where dose-responses are monotonically increasing. In comparison to scenario A1, scenario A2 has efficacy rate increasing much more quickly with higher dose levels (from dose level 3 to 4). Scenarios A3 and A4 represent situations in which the efficacy increases with the dose, and then plateaus. Scenarios A5, A6, and A7 are circumstances where the dose-response curve increases to an optimal point, and then decreases. Scenario A8 is where no optimal dose exists, because all doses are overly toxic.

Table S9 summarizes the operating characteristics for the designs. In general, U-BOIN outperforms EffTox with a higher percentage of correct selection (PCS) of the OBD. For example, in scenario A1, the dose response curve monotonically increases, and a suboptimal dose (i.e., dose level 3) has a similar efficacy rate as the OBD (i.e., dose level 2), but a higher toxicity rate; the U-BOIN design yields a 9% higher PCS than EffTox. A similar gain is observed when sub-optimal doses have comparable DLT rates as the OBD, but lower efficacy rates (e.g., scenarios A2 and A4). The advantage of U-BOIN is more obvious when the OBD has higher efficacy and lower toxicity, as compared to other doses (e.g., scenario A6). In scenario 6, the PCS of U-BOIN

is 18% higher than that of EffTox. Scenario A7 is similar to scenario A6 in that they have a similar dose-response shape and the same OBD location (dose level 2). The dissimilarity is that the difference in both toxicity and efficacy in dose level 1 and 2 is larger in scenario A6 than in A7. In scenario A7, the U-BOIN design has a PCS of 71%, but EffTOX only has a PCS of 35% and mistakenly selects dose 1 as the OBD with a probability of 0.55. This further shows the robustness of the U-BOIN design and the sensitivity of a model-based design. Scenario A8 considers situations where an OBD does not exist, because all doses are overly toxic. In this case, the PCS is defined as the percentage of simulated trials stopped early, due to toxicity. We can see that U-BOIN is more likely to stop the trial correctly.

In terms of patient allocation, U-BOIN in general assigns more patients on the OBD, while the performance between the two designs is comparable in Scenarios A2 and A5.

**TABLE S9** Results of additional eight scenarios for Simulation A, including the selection percentage (selection %), the average number of patients treated at each dose (No. of patients), and the percentage of early stopping. The optimal biological dose (OBD) is bolded. In scenario 8, the OBD does not exist, and thus the percentage of early stopping is bolded.

| Design | | Dose Level 1 | 2 | 3 | 4 | 5 | % of early stopping |
|---|---|---|---|---|---|---|---|
| | | | | **Scenario A1** | | | |
| | DLT rate | 0.03 | **0.17** | 0.35 | 0.50 | 0.65 | |
| | Efficacy rate | 0.12 | **0.39** | 0.43 | 0.55 | 0.65 | |
| | Utility | 37.0 | **51.0** | 46.0 | 48.0 | 47.0 | |
| EffTox | Selection % | 6.0 | **57.0** | 31.0 | 2.0 | 0.0 | 3.0 |
| | No. of patients | 6.3 | **23.1** | 19.0 | 3.7 | 0.9 | |
| U-BOIN | Selection % | 5.3 | **70.6** | 19.1 | 3.4 | 0.0 | 1.6 |
| | No. of patients | 8.8 | **29.5** | 12.1 | 3.0 | 0.3 | |
| | | | | **Scenario A2** | | | |
| | DLT rate | 0.01 | 0.08 | 0.10 | **0.12** | 0.35 | |
| | Efficacy rate | 0.05 | 0.25 | 0.30 | **0.60** | 0.60 | |
| | Utility | 33.0 | 45.0 | 47.0 | **67.0** | 57.0 | |
| EffTox | Selection % | 0.0 | 4.0 | 15.0 | **64.0** | 16.0 | 0.0 |
| | No. of patients | 3.4 | 4.6 | 10.6 | **25.1** | 10.2 | |
| U-BOIN | Selection % | 0.1 | 6.3 | 6.2 | **70.8** | 16.3 | 0.2 |
| | No. of patients | 4.3 | 7.9 | 7.4 | **23.2** | 11.2 | |
| | | | | **Scenario A3** | | | |
| | DLT rate | **0.20** | 0.35 | 0.45 | 0.55 | 0.76 | |
| | Efficacy rate | **0.40** | 0.45 | 0.55 | 0.55 | 0.55 | |
| | Utility | **50.0** | 48.0 | 50.0 | 46.0 | 37.0 | |
| EffTox | Selection % | **57.0** | 29.0 | 8.0 | 1.0 | 0.0 | 5.0 |
| | No. of patients | **24.7** | 16.7 | 7.9 | 1.8 | 0.7 | |
| U-BOIN | Selection % | **71.5** | 20.4 | 4.7 | 0.2 | 0.0 | 3.1 |
| | No. of patients | **35.5** | 13.1 | 3.5 | 0.5 | 0.0 | |
| | | | | **Scenario A4** | | | |
| | DLT rate | 0.10 | **0.13** | 0.34 | 0.45 | 0.50 | |
| | Efficacy rate | 0.20 | **0.45** | 0.45 | 0.45 | 0.45 | |
| | Utility | 41.0 | **56.0** | 48.0 | 44.0 | 42.0 | |
| EffTox | Selection % | 15.0 | **57.0** | 23.0 | 3.0 | 1.0 | 2.0 |
| | No. of patients | 10.3 | **22.8** | 14.9 | 3.4 | 2.0 | |
| U-BOIN | Selection % | 7.9 | **72.1** | 16.8 | 1.8 | 0.4 | 1.1 |
| | No. of patients | 9.9 | **28.6** | 11.8 | 2.8 | 0.5 | |
| | | | | **Scenario A5** | | | |
| | DLT rate | 0.01 | 0.08 | **0.10** | 0.20 | 0.40 | |
| | Efficacy rate | 0.07 | 0.10 | **0.42** | 0.26 | 0.20 | |
| | Utility | 35.0 | 34.0 | **56.0** | 41.0 | 30.0 | |

**Table S9 Continued:**

| Design | | Dose Level 1 | 2 | 3 | 4 | 5 | % of early stopping |
|---|---|---|---|---|---|---|---|
| EffTox | Selection % | 1.0 | 4.0 | **69.0** | 19.0 | 3.0 | 3.0 |
| | No. of patients | 4.5 | 4.9 | **25.5** | 13.7 | 4.5 | |
| U-BOIN | Selection % | 2.1 | 3.5 | **78.1** | 13.6 | 1.1 | 1.7 |
| | No. of patients | 5.3 | 6.9 | **24.6** | 11.7 | 5.1 | |
| | | | | **Scenario A6** | | | |
| | DLT rate | 0.05 | **0.15** | 0.30 | 0.45 | 0.60 | |
| | Efficacy rate | 0.08 | **0.46** | 0.25 | 0.20 | 0.10 | |
| | Utility | 34.0 | **56.0** | 37.0 | 29.0 | 18.0 | |
| EffTox | Selection % | 9.0 | **71.0** | 10.0 | 2.0 | 1.0 | 7.0 |
| | No. of patients | 10.8 | **25.1** | 11.7 | 3.0 | 1.0 | |
| U-BOIN | Selection % | 1.4 | **91.5** | 4.7 | 0.4 | 0.0 | 2.1 |
| | No. of patients | 6.7 | **34.4** | 8.9 | 2.7 | 0.4 | |
| | | | | **Scenario A7** | | | |
| | DLT rate | 0.15 | **0.20** | 0.40 | 0.50 | 0.55 | |
| | Efficacy rate | 0.25 | **0.45** | 0.30 | 0.25 | 0.25 | |
| | Utility | 42.0 | **54.0** | 36.0 | 30.0 | 28.0 | |
| EffTox | Selection % | 55.0 | **35.0** | 4.0 | 1.0 | 0.0 | 5.0 |
| | No. of patients | 25.0 | **17.3** | 6.7 | 2.0 | 1.3 | |
| U-BOIN | Selection % | 21.4 | **72.4** | 3.0 | 0.2 | 0.0 | 2.8 |
| | No. of patients | 17.0 | **28.1** | 6.2 | 1.3 | 0.2 | |
| | | | | **Scenario A8** | | | |
| | DLT rate | 0.45 | 0.50 | 0.55 | 0.60 | 0.65 | |
| | Efficacy rate | 0.15 | 0.18 | 0.25 | 0.25 | 0.25 | |
| | Utility | 26.0 | 26.0 | 28.0 | 26.0 | 25.0 | |
| EffTox | Selection % | 10.0 | 4.0 | 1.0 | 1.0 | 0.0 | **84.0** |
| | No. of patients | 10.3 | 4.7 | 3.2 | 1.5 | 1.2 | |
| U-BOIN | Selection % | 8.2 | 2.1 | 0.4 | 0.0 | 0.0 | **89.2** |
| | No. of patients | 17.2 | 2.6 | 0.4 | 0.0 | 0.0 | |

## 4.2 | Simulation B

Table S10 shows results for Simulation B. The PCS of the OBD and the number of the patients allocated to the OBD are generally comparable to these in Simulation A (i.e., the optimal benchmark with fully observed data), indicating that U-BOIN efficiently handles the delayed efficacy response. Because U-BOIN does not need to suspend accrual to wait for $Y_E$ to be fully observed and allows real-time decision making, it has great potential to shorten the trial duration.

**TABLE S10** Results of the additional eight scenarios for Simulation B, including the selection percentage (selection %), the average number of patients treated at each dose (No. of patients), the percentage of early stopping, and the trial duration. The optimal biological dose (OBD) is bolded. In scenario 8, the OBD does not exist, and thus the percentage of early stopping is boldeded.

| Design | | Dose Level 1 | 2 | 3 | 4 | 5 | % of early stopping | Duration (month) |
|---|---|---|---|---|---|---|---|---|
| | | | | **Scenario A1** | | | | |
| EffTox | Selection % | 6.0 | **57.0** | 31.0 | 2.0 | 0.0 | 3.0 | 47.5 |
| | No. of patients | 6.3 | **23.1** | 19.0 | 3.7 | 0.9 | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| U-BOIN | Selection % | 2.1 | **65.1** | 21.9 | 3.8 | 0.1 | 7.0 | 20.4 |
| | No. of patients | 7.5 | **28.8** | 12.8 | 2.9 | 0.3 | | |
| | | | | **Scenario A2** | | | | |
| EffTox | Selection % | 0.0 | 4.0 | 15.0 | **64.0** | 16.0 | 0.0 | 43.0 |
| | No. of patients | 3.4 | 4.6 | 10.6 | **25.1** | 10.2 | | |
| U-BOIN | Selection % | 0.0 | 6.0 | 5.3 | **70.9** | 15.9 | 1.8 | 20.9 |
| | No. of patients | 4.1 | 8.1 | 8.2 | **22.8** | 10.3 | | |
| | | | | **Scenario A3** | | | | |
| EffTox | Selection % | **57.0** | 29.0 | 8.0 | 1.0 | 0.0 | 5.0 | 51.1 |
| | No. of patients | **24.7** | 16.7 | 7.9 | 1.8 | 0.7 | | |
| U-BOIN | Selection % | **68.1** | 20.8 | 5.6 | 0.4 | 0.0 | 5.1 | 20.3 |
| | No. of patients | **35.0** | 12.9 | 3.4 | 0.5 | 0.0 | | |
| | | | | **Scenario A4** | | | | |
| EffTox | Selection % | 15.0 | **57.0** | 23.0 | 3.0 | 1.0 | 2.0 | 48.1 |
| | No. of patients | 10.3 | **22.8** | 14.9 | 3.4 | 2.0 | | |
| U-BOIN | Selection % | 2.1 | **67.1** | 17.4 | 3.2 | 0.3 | 9.9 | 19.9 |
| | No. of patients | 7.9 | **27.2** | 11.9 | 3.1 | 0.6 | | |
| | | | | **Scenario A5** | | | | |
| EffTox | Selection % | **1.0** | 4.0 | 69.0 | 19.0 | 3.0 | 3.0 | 43.8 |
| | No. of patients | **4.5** | 4.9 | 25.5 | 13.7 | 4.5 | | |
| U-BOIN | Selection % | **0.1** | 1.6 | 80.8 | 12.1 | 1.4 | 4.0 | 20.5 |
| | No. of patients | **4.3** | 6.6 | 24.9 | 11.7 | 5.2 | | |
| | | | | **Scenario A6** | | | | |
| EffTox | Selection % | 9.0 | 71.0 | 10.0 | 2.0 | **1.0** | 7.0 | 45.6 |
| | No. of patients | 10.8 | 25.1 | 11.7 | 3.0 | **1.0** | | |
| U-BOIN | Selection % | 0.2 | 90.2 | 5.1 | 0.4 | **0.0** | 4.0 | 20.4 |
| | No. of patients | 5.8 | 34.1 | 9.4 | 2.7 | **0.4** | | |
| | | | | **Scenario A7** | | | | |
| EffTox | Selection % | 55.0 | **35.0** | **4.0** | 1.0 | 0.0 | 5.0 | 49.3 |
| | No. of patients | 25.0 | **17.3** | **6.7** | 2.0 | 1.3 | | |
| U-BOIN | Selection % | 20.2 | **66.2** | **3.3** | 0.3 | 0.0 | 10.0 | 19.9 |
| | No. of patients | 16.9 | **26.3** | **6.1** | 1.1 | 0.1 | | |
| | | | | **Scenario A8** | | | | |
| EffTox | Selection % | 10.0 | 4.0 | 1.0 | 1.0 | 0.0 | **84.00** | 19.3 |
| | No. of patients | 10.3 | 4.7 | 3.2 | 1.5 | 1.2 | | |
| U-BOIN | Selection % | 0.0 | 0.8 | 0.0 | 0.0 | 0.0 | **99.20** | 6.7 |
| | No. of patients | 9.6 | 1.3 | 0.2 | 0.0 | 0.0 | | |

## 5 | SENSITIVITY ANALYSIS WITH DIFFERENT $S_2$ IN SIMULATION A

Figure S1 Panel (1) shows the percentage of correct selection when the values of $s_2$ vary. When $s_2$ are 18 and 21 (recommended in step B3 in the dose finding algorithm), the change in the percentage of correct selection is negligible. Panel (2) shows that the corresponding sample sizes significantly decrease when compared to EffTox or the U-BOIN with $s_2 = 54$, the maximum sample size in the trial.

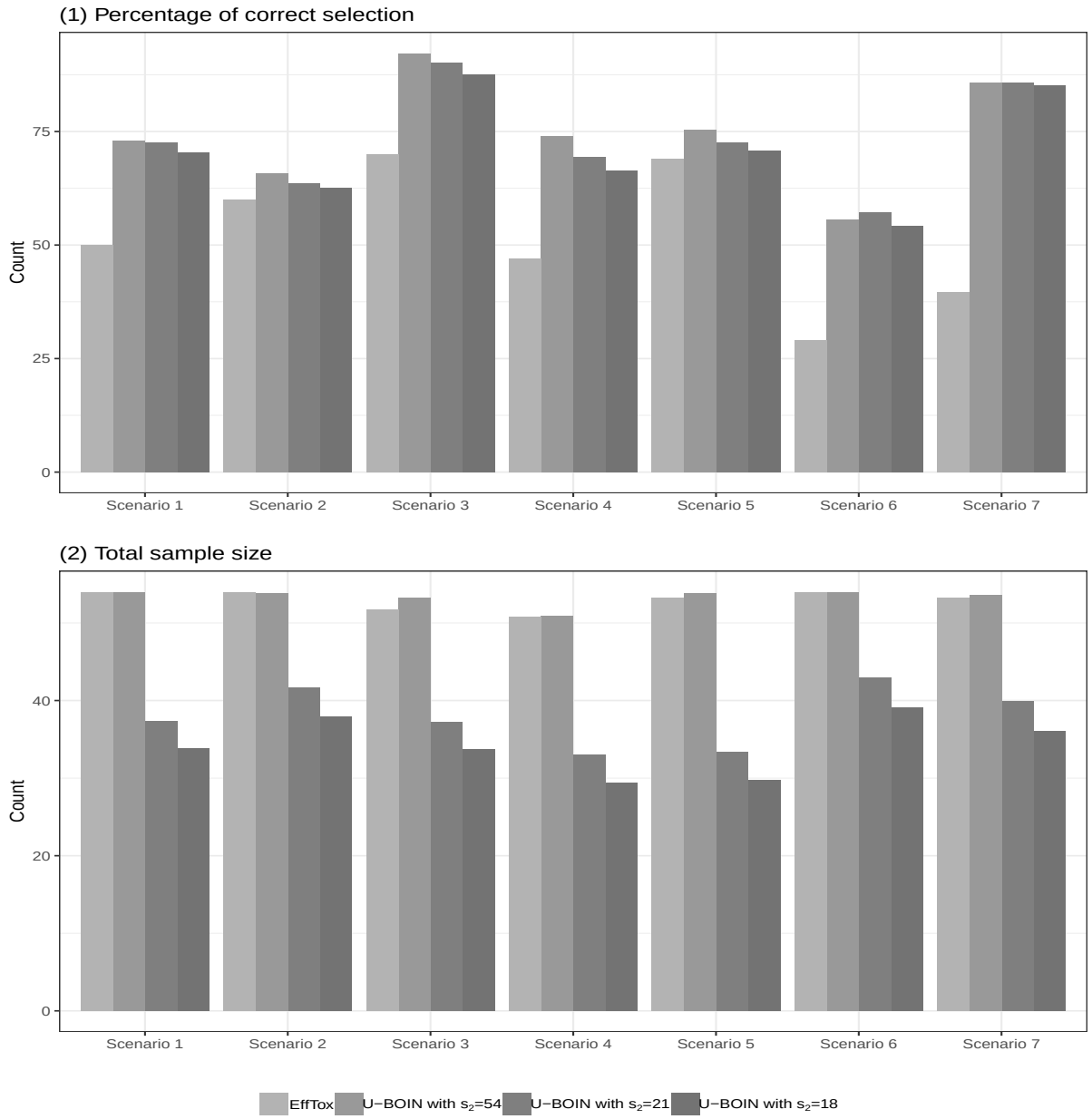


**FIGURE S1** Sensitivity analysis with different values of $s_2$.

# 6 | ADDITIONAL SENSITIVITY ANALYSIS FOR SIMULATION B

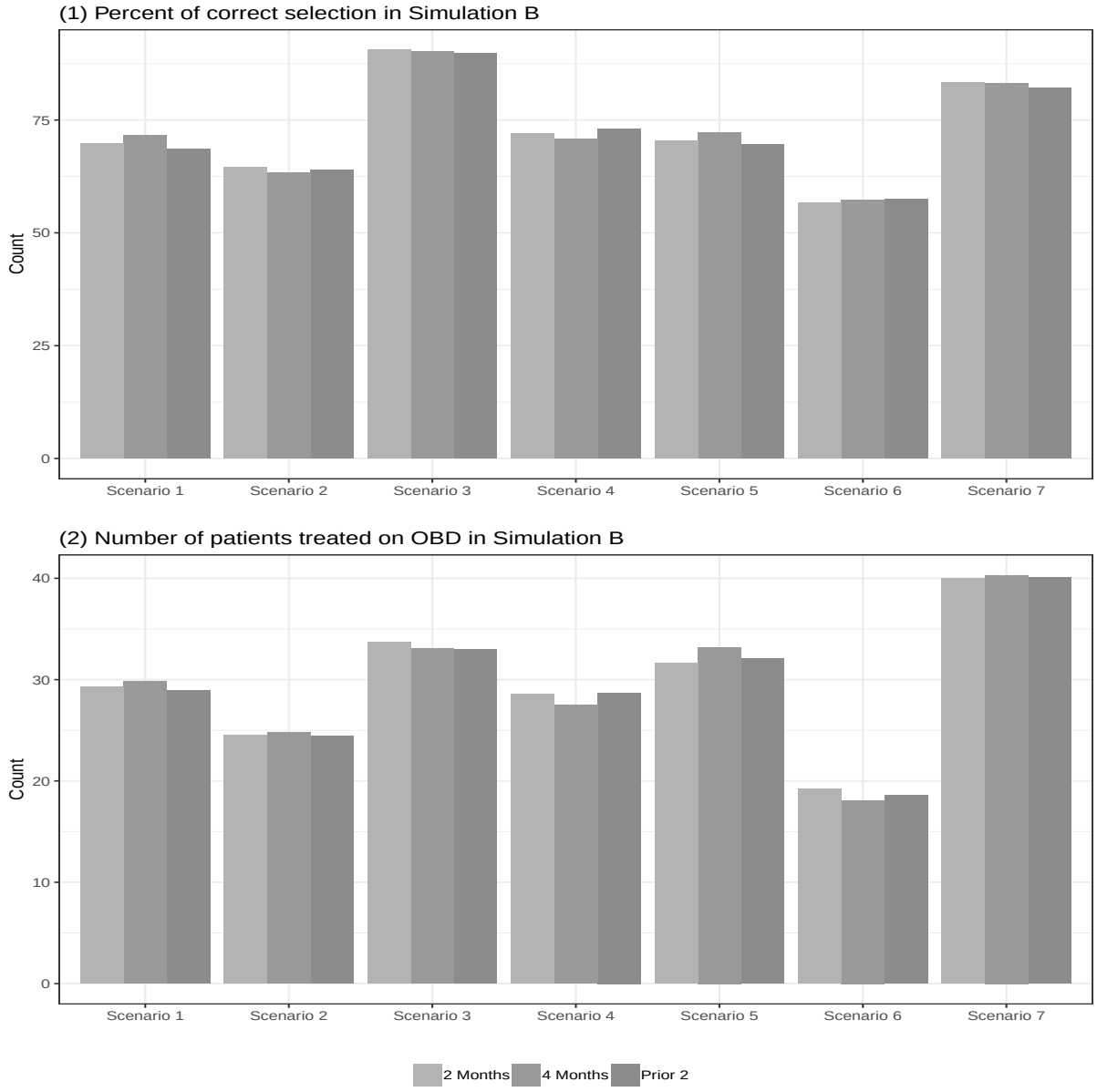


**FIGURE S2** Sensitivity analysis with 2-month and 4-month efficacy assessment window, and a different specification of the prior (denoted as prior 2) for the prediction model parameters: $\beta_0 \sim N(0, 3.75^2)$ and $\beta_1 \sim Gamma(shape = 1, rate = 0.4)$.

**TABLE S11** Results of sensitivity analysis in two scenarios where efficacy ($Y_E$) and immune response ($Y_I$) are weakly associated.

| Design | | Dose Level 1 | 2 | 3 | 4 | 5 | % of early stopping | Duration (month) |
|---|---|---|---|---|---|---|---|---|
| | | **Scenario B1** (Correlation coefficient = -0.15) | | | | | | |
| | DLT rate | 0.08 | 0.10 | **0.15** | 0.32 | 0.40 | | |
| | Efficacy rate | 0.10 | 0.20 | **0.45** | 0.50 | 0.55 | | |
| | Utility | 34 | 41 | **56** | 52 | 52 | | |
| EffTox | Selection % | 1.0 | 7.0 | **53.0** | 25.0 | 13.0 | 0 | 45.5 |
| | No. of patients | 4.1 | 5.9 | **22.4** | 13.0 | 8.4 | | |
| U-BOIN | Selection % | 2.5 | 12.2 | **60.2** | 13.9 | 5.0 | 6.2 | 20.3 |
| | No. of patients | 7.0 | 10.9 | 21.7 | 9.1 | 3.2 | | |
| | | **Scenario B2** (Correlation coefficient = 0.2) | | | | | | |
| | DLT rate | 0.15 | **0.20** | 0.40 | 0.50 | 0.55 | | |
| | Efficacy rate | 0.20 | **0.45** | 0.30 | 0.25 | 0.25 | | |
| | Utility | 39 | **54** | 36 | 30 | 28 | | |
| EffTox | Selection % | 45.0 | **43.0** | 5.0 | 1.0 | 1.0 | 6.0 | 49.3 |
| | No. of patients | 21.2 | **19.3** | 7.5 | 2.4 | 1.6 | | |
| U-BOIN | Selection % | 10.2 | **74.6** | 3.7 | 0.3 | 0 | 11.2 | 19.6 |
| | No. of patients | 13.9 | **28.6** | 6.1 | 1.1 | 0.2 | | |