A Bayesian Bivariate Probit Model For Phase I-II Dose Finding Trials

Clay Olsen

June 2021

Abstract

Traditionally, in dose-finding trials, researchers adopt a rule-based methodology for separate phase 1 and 2 trials to find the maximum tolerated dose level and evaluate the efficacy of an experimental drug. Alternatively, a single model-based, early-phase trial design can be utilized to simultaneously account for efficacy and toxicity. We propose a bivariate probit design to model the binary outcomes of efficacy and toxicity under a Bayesian framework while accounting for potential dependencies between dose-efficacy and dose-toxicity. We evaluate dose levels with an intuitive utility function, proposed in Liu and Johnson [2016], that accounts for desired trade-offs between efficacy and toxicity. We conduct extensive simulation studies to examine the operating characteristics of our proposed design versus the design in [Liu and Johnson, 2016] under various practical dose-efficacy and dose-toxicity relationships. The results show that the proposed design possesses favorable operating characteristics and can capture the shape of the dose-efficacy and dose-toxicity curves.

KEY WORDS: Dose finding, Adaptive design, Bayesian models, Bi-variate probit model, Utility function

1 Introduction

A primary objective for early phase clinical trials is to identify the maximum tolerated dose level (MTD) of a new drug. The MTD is the dose level where side effects are serious enough to prevent an increase in dose. Often, for the first phase of a trial, several dose levels are tested to build a toxicity profile of the drug. In the second phase, the drug is administered to a small cohort of patients to build an efficacy profile. With successful results, a third phase begins with a larger cohort of patients [F. Yan and Yuan, 2018]. There are some disadvantages to having separate phase I and II trials. The MTD results from a phase I trial are often unreliable due to the likely, small sample size obtained. Thus, the results are susceptible to inconsistencies in data collection. Consequently, MTD can have high variability from phase I results. Dose adjustments may need to be made for phase II if excessive toxicity is reached. Additionally, physicians commonly use rule-based trial designs, the most common being the "3+3" trial design. The "3+3" design proceeds with cohorts of three patients with the first cohort being treated at starting dose level considered safe, from animal toxicological data. Binary outcomes of toxicity are observed after the necessary response time of the patients are met. If none of the three patients have toxic outcomes, another three patients will be treated at the next dose level. If one patient of the three has a toxic outcome, another

three patients are treated at the same dose level. If two or more of the three patients have toxic outcomes, dose escalation is stopped and the MTD is determined [Christophe Le Tourneau and Siu, 2009]. Rule-based trial designs tend to have limited flexibility, are unable to account for inconsistencies in data, and do not properly utilize previously collected data in further testing, such as information from other dose levels. Additionally, few patients with this method would be treated at the dose level determined to be most optimal.

An alternative is using a model-based approach with a singular, early-phase trial aiming to find the optimal dose level of an experimental drug by maximizing a utility function. For example, Liu and Johnson [2016] build an adaptive Bayesian design that estimates the dose-toxicity and dose-efficacy curves of an experimental drug. Patients are assigned to designated dose levels and binary outcomes of efficacy and toxicity are observed. The design adapts to assign each cohort with probability proportional to the utility of each dose level determined by the utility function described in Section 4. Independent, default prior densities are used to provide information of the expected efficacy and toxicity, which monotonically increase with dose level. Despite the constraint of efficacy and toxicity increasing with dose level, the dose-efficacy and dose-toxicity curves remain fairly malleable, allowing patient results to influence these relationships. Offering model flexibility and accounting for the efficacy and toxicity trade-offs provide important ethical implications, which may benefit patient recruiting.

In our paper, we experiment with two model based, dose finding trial designs. We compare the design in Liu and Johnson [2016], that assumes independent Bernoulli distributions, with our proposed bi-variate probit regression model that incorporates dependencies between efficacy and toxicity. The bivariate model introduces latent variables for efficacy and toxicity effects to estimate correlated binary outcomes jointly. On the otherhand, the model in Liu and Johnson [2016] captures the marginal effects of efficacy and toxicity, ignoring the effects correlation.

The rest of the paper is organized as follows. In Sections 2 and 3, we review the Liu and Johnson model and propose a bivariate probit model used for estimating efficacy and toxicity. In Section 4, we define a utility function as a trade-off between toxicity and efficacy. In Section 5, we describe the dose-finding algorithm and decision rules for the proposed trial design based on the utility function. In Section 6, we present simulation studies to examine the operating characters of this design for different dose-efficacy and dose-toxicity scenarios followed by a brief discussions of the results.

2 Liu and Johnson's Trial Design

2.1 Sampling Distribution and Prior Specification

Let y_{Ei} and y_{Ti} represent the binary outcomes for efficacy and toxicity for patient i, where $i = 1, ..., n(t)$, for trial time t. $y_{Ei} = 1$ means the dose given to patient i had an effective response and $y_{Ti} = 1$ means the dose given to patient i showed symptoms resulting from toxicity. Likewise, $y_{Ei} = 0$ and $y_{Ti} = 0$ represent no efficacy or toxicity responses in patient i. For the dose level j, where $j = 1, ..., J$, p_{Ej} and p_{Tj} represent the probability of efficacy and toxicity responses. $d_{[i]}$ refers to the dose level given to patient i. y_{Ei} and y_{Ti} are modeled with the following Bernoulli distribution,

$$
\begin{aligned}
&y_{ki}|d_{[i]} = j \overset{ind}{\sim} Bernoulli(p_{kj}), \;\; k \in \{E, T\}, \\
&p_{k1} = \beta_{k1}, \\
&p_{kj} = p_{k,j-1} + (1 - p_{k,j-1})\beta_{kj}, \;\; j = 2, ..., J, \\
&\beta_{kj} \overset{ind}{\sim} Beta(a_{kj}, b_{kj}).
\end{aligned} \tag{1}
$$

In the model, p_{Ej} and p_{Tj} follow a random variable Markov structure, where p_{kj}'s scale through Equation (1), where the $\beta_{kj} \in (0, 1)$ is generated from Beta distributions. Thus, p_{kj} increases with dose level, $p_{k1} < p_{k2} < ... < p_{kJ}$. We can rewrite p_{kj} as $p_{kj} = 1 - (1 - p_{k,j-1})(1 - \beta_{k,j})$, which implies;

$$
p_{kj} = 1 - \prod_{r=1}^{j}(1 - \beta_{kr}). \tag{2}
$$

Given n patients, J dose levels and dose $d_{[i]}$ for patient i at trial time t, the model's likelihood function is,

$$
L(\boldsymbol{\beta_E}, \boldsymbol{\beta_T}|\mathcal{D}_{(t)}) = \prod_{k=\{E,T\}} \prod_{i=1}^{n(t)} \left\{ 1 - \prod_{r=1}^{d_{[i]}}(1 - \beta_{kr}) \right\}^{y_{ki}} \prod_{r=1}^{d_{[i]}} (1 - \beta_{kr})^{1-y_{ki}}, \tag{3}
$$

where $\boldsymbol{\beta_E} = (\beta_{E1}, ..., \beta_{EJ})$, $\boldsymbol{\beta_T} = (\beta_{T1}, ..., \beta_{TJ})$ and $\mathcal{D}_{(t)} = \{(d_{[i]}, y_{Ei}, y_{Ti}), i = 1, ..., n(t)\}$. The priors for $\boldsymbol{\beta_E}$ and $\boldsymbol{\beta_T}$, $\pi(\boldsymbol{\beta_E})$ and $\pi(\boldsymbol{\beta_T})$, are Beta distributions as follows,

$$
\begin{aligned}
\pi(\boldsymbol{\beta_E}) &= \prod_{j=1}^{J} \frac{\Gamma(\alpha_{Ej} + b_{Ej})}{\Gamma(\alpha_{Ej})\Gamma(b_{Ej})} (\beta_{Ej})^{\alpha_{Ej}-1}(1 - \beta_{Ej})^{b_{Ej}-1}, \\
\pi(\boldsymbol{\beta_T}) &= \prod_{j=1}^{J} \frac{\Gamma(\alpha_{Tj} + b_{Tj})}{\Gamma(\alpha_{Tj})\Gamma(b_{Tj})} (\beta_{Tj})^{\alpha_{Tj}-1}(1 - \beta_{Tj})^{b_{Tj}-1}.
\end{aligned} \tag{4}
$$

2.2 Prior Calibration

The values of hyper-parameters a_{kj} and b_{kj}, in the Beta distributions, can be derived from prior probabilities for efficacy and toxicity given from physicians. Let $\tilde{p}_{kj}$ represent the prior probabilities given from physicians for dose j and outcome $k \in \{E, T\}$. By re-arranging Equation (2), we can solve for β_{kj},

$$
\beta_{kj} = \frac{\tilde{p}_{kj} - \tilde{p}_{k,j-1}}{1 - \tilde{p}_{k,j-1}}, \tag{5}
$$

where $\tilde{p}_{k0} = 0$. Since β_{kj} follows a Beta distribution, we can equate the prior mean, $E(\beta_{kj}) = \frac{a_{kj}}{a_{kj}+b_{kj}}$ to $\frac{\tilde{p}_{kj}-\tilde{p}_{k,j-1}}{1-\tilde{p}_{k,j-1}}$. Additionally, we set a constraint by fixing $a_{kj} + b_{kj} = m$, where m is a constant representing the "prior sample size." A weakly informative prior is

achieved by setting m to a low positive value like $m = 1$. With this constraint, we can obtain values for the hyper-parameters.

$$a_{kj} = m\frac{\tilde{p}_{kj} - \tilde{p}_{k,j-1}}{1 - \tilde{p}_{k,j-1}} \quad \text{and} \quad b_{kj} = m\frac{1 - \tilde{p}_{kj}}{1 - \tilde{p}_{k,j-1}}. \tag{6}$$

Assuming that the prior probabilities provided by physicians are $\tilde{p}_{ej} = (0.2, 0.3, 0.4, 0.5, 0.6)$, and $\tilde{p}_{tj} = (0.05, 0.1, 0.2, 0.3, 0.35)$, we derive α_{kj} and b_{kj} using equation 6 resulting in, $a_{ej} = (0.2, 0.3, 0.4, 0.5, 0.6)$, $b_{ej} = (0.8, 0.7, 0.6, 0.5, 0.4)$, $a_{tj} = (0.05, 0.1, 0.2, 0.3, 0.35)$, and $b_{tj} = (0.95, 0.9, 0.8, 0.7, 0.65)$.

2.3 Posterior Simulation

The posterior distribution, $f(\boldsymbol{\beta_E}, \boldsymbol{\beta_T}|\mathcal{D}_{(t)})$, can be written as

$$\begin{aligned}
f(\boldsymbol{\beta_E},&\boldsymbol{\beta_T}|\mathcal{D}_{(t)}) \propto L(\boldsymbol{\beta_E}, \boldsymbol{\beta_T}|\mathcal{D}_{(t)})\pi(\boldsymbol{\beta_E})\pi(\boldsymbol{\beta_T}) \\
&= \prod_{i=1}^{n(t)} \left\{1 - \prod_{r=1}^{d_{[i]}}(1-\beta_{Er})\right\}^{y_{Ei}} \prod_{r=1}^{d_{[i]}}(1-\beta_{Er})^{1-y_{Ei}} \\
&\quad \left\{1 - \prod_{r=1}^{d_{[i]}}(1-\beta_{Tr})\right\}^{y_{Ti}} \prod_{r=1}^{d_{[i]}}(1-\beta_{Tr})^{1-y_{Ti}} \prod_{j=1}^{J} \frac{\Gamma(\alpha_{Ej}+b_{Ej})}{\Gamma(\alpha_{Ej})\Gamma(b_{Ej})} \\
&\quad (\beta_{Ej})^{\alpha_{Ej}-1}(1-\beta_{Ej})^{b_{Ej}-1} \prod_{j=1}^{J} \frac{\Gamma(\alpha_{Tj}+b_{Tj})}{\Gamma(\alpha_{Tr})\Gamma(b_{Tj})} (\beta_{Tj})^{\alpha_{Tj}-1}(1-\beta_{Tj})^{b_{Tj}-1} \\
&\propto \prod_{i=1}^{n(t)} \left\{1 - \prod_{r=1}^{d_{[i]}}(1-\beta_{Er})\right\}^{y_{Ei}} \prod_{r=1}^{d_{[i]}}(1-\beta_{Er})^{1-y_{Ei}} \left\{1 - \prod_{r=1}^{d_{[i]}}(1-\beta_{Tr})\right\}^{y_{Ti}} \\
&\quad \prod_{r=1}^{d_{[i]}}(1-\beta_{Tr})^{1-y_{Ti}} \prod_{j=1}^{J} (\beta_{Ej})^{\alpha_{Ej}-1}(1-\beta_{Ej})^{b_{Ej}-1}(\beta_{Tj})^{\alpha_{Tj}-1}(1-\beta_{Tj})^{b_{Tj}-1}.
\end{aligned} \tag{7}$$

.

Since y_{Ei} and y_{Ti} are assumed independent, $\boldsymbol{\beta_E}$ and $\boldsymbol{\beta_T}$ are independent a postriori. Let $\boldsymbol{y_k} = \{y_{ki}; i = 1, ..., n(t)\}$ and $\boldsymbol{d} = \{d_{[i]}; i = 1, ..., n(t)\}$.

$$\begin{aligned}
&f(\boldsymbol{\beta_E}, \boldsymbol{\beta_T}|\mathcal{D}_{(t)}) = f(\boldsymbol{\beta_E}|\boldsymbol{y_E}, \boldsymbol{d}) f(\boldsymbol{\beta_T}|\boldsymbol{y_T}, \boldsymbol{d}), \\
&f(\boldsymbol{\beta_E}|\boldsymbol{y_E}, \boldsymbol{d}) \propto \prod_{i=1}^{n(t)} \left\{1 - \prod_{r=1}^{d_{[i]}}(1-\beta_{Er})\right\}^{y_{Ei}} \prod_{r=1}^{d[i]}(1-\beta_{Er})^{1-y_{Ei}} \prod_{j=1}^{J}(\beta_{Ej})^{\alpha_{Ej}-1}(1-\beta_{Ej})^{b_{Ej}-1}, \\
&f(\boldsymbol{\beta_T}|\boldsymbol{y_T}, \boldsymbol{d}) \propto \prod_{i=1}^{n(t)} \left\{1 - \prod_{r=1}^{d[i]}(1-\beta_{Tr})\right\}^{y_{Ti}} \prod_{r=1}^{d[i]}(1-\beta_{Tr})^{1-y_{Ti}} \prod_{j=1}^{J}(\beta_{Tj})^{\alpha_{Tj}-1}(1-\beta_{Tj})^{b_{Tj}-1}.
\end{aligned} \tag{8}$$

With the full conditional posterior distributions, $f(\boldsymbol{\beta_E}|\boldsymbol{y_E}, \boldsymbol{d})$ and $f(\boldsymbol{\beta_T}|\boldsymbol{y_T}, \boldsymbol{d})$, we obtain posterior samples of β_{Ej} and β_{Tj} through Monte Carlo Markov Chain (MCMC). We use those samples to get posterior samples of p_{Ej} and p_{Tj}. The prior values of $\tilde{p}_{Ej}$ and $\tilde{p}_{Tj}$, which are discussed in Section 2.2, are used for the initial values of β_{Ej} and β_{Tj}.

Since β_{Ej}'s and β_{Tj}'s have bounds $(0, 1)$, it is advisable to perform a Metropolis-Hastings algorithm on the logit transformed scale for their update. Additionally, we implemented an adaptive Metropolis Hastings for better mixing of the chain [Roberts and Rosenthal, 2006]. With posterior estimates of β_{kj} we can get our posterior distributions of p_{kj}.

3 Bivariate Probit Model

3.1 Sampling Distribution and Prior Specification

For the bi-variate probit model, with our two binary response variables y_{Ei} and y_{Ti}, we introduce continuous latent variables $\tilde{y}_{ki} \in R$ and let,

$$y_{ki} = \begin{cases} 1 \text{ if } \tilde{y}_{ki} \geq 0, \\ 0 \text{ if } \tilde{y}_{ki} < 0. \end{cases} \tag{9}$$

We let $\tilde{\boldsymbol{y}}_{\boldsymbol{i}} = [\tilde{y}_{Ei}, \tilde{y}_{Ti}]'$ follow the bivariate normal distribution with parameters $\mu_{d_{[i]}}$ and Σ. We assume that μ_{kj}'s monotonically increase with dose level and introduce hyper-parameter α_{kj}. We model the distributions of $\tilde{\boldsymbol{y}}$ conditional on $\boldsymbol{\mu}$, and Σ using the following model,

$$\begin{aligned} &\tilde{\boldsymbol{y}}_{\boldsymbol{i}} \mid \boldsymbol{\mu}, \Sigma, d_{[i]} = j \sim N_2(\boldsymbol{\mu}, \Sigma), \\ &\text{where, } \boldsymbol{\mu_j} = \begin{bmatrix} \mu_{Ej} \\ \mu_{Tj} \end{bmatrix}, \ \tilde{\boldsymbol{y}}_{\boldsymbol{i}} = \begin{bmatrix} \tilde{y}_{Ei} \\ \tilde{y}_{Ti} \end{bmatrix}, \text{ and } \Sigma = \begin{bmatrix} \sigma_E^2 & \rho\sigma_E\sigma_T \\ \rho\sigma_E\sigma_T & \sigma_T^2 \end{bmatrix}. \end{aligned} \tag{10}$$

Then we assume,

$$\begin{aligned} &\mu_{k1} \sim N(\bar{\alpha}_{k1}, \tau_{k1}^2), \\ &\mu_{kj} = \mu_{k,j-1} + \alpha_{kj}, \ j = 2, ..., J, \\ &\alpha_{kj} \sim N_+(\bar{\alpha}_{kj}, \tau_{k2}^2), \text{ a truncated normal distribution where, } \alpha_{kj} > 0 \end{aligned} \tag{11}$$

The assumption of efficacy and toxicity monotonically increasing with dose level is maintained with the added variable α_{kj}, so that, $\mu_{k1} \leq \mu_{k2} \leq ... \leq \mu_{kJ}$. Setting, $\mu_{k1} = \alpha_{k1}$, it follows that for dose j, $\mu_{kj} = \sum_{r=1}^{j} \alpha_{kr}$. Note that α_{k1} is on the real line, while α_{kj} is strictly positive for $j > 1$. We model α_{k1} with variance τ_{k1} and α_{kj}, with variance τ_{k2}, $j > 2$. We set the values of $\sigma_E^2 = \sigma_T^2 = 1$. Let $\rho = 2(\tilde{\rho} - \frac{1}{2})$ and assume,

$$\tilde{\rho} \sim \text{Beta}(a_p, b_p). \tag{12}$$

We use this form of ρ to maintain $\rho \in (-1, 1)$. With this model, for dose level $j = 1, ...J$, $p(y_{kj} = 1) = p(\tilde{y}_{kj} > 0), k \in \{E, T\}$.

3.2 Prior Calibration

We propose to determine the values of the hyper-parameters $\bar{\alpha}_{Ej}$, $\bar{\alpha}_{Tj}$, τ_{E1}, τ_{T1}, τ_{E2}, τ_{T2}, a_ρ, and b_ρ by matching the prior probabilities elicited by clinicians in Section 2.2. $\bar{\alpha}_{kj}$'s are derived from $\tilde{p}_{kj}$'s, the prior probabilities for efficacy and toxicity proposed by physicians. We set $\tau_{k1} = 1.15$ and $\tau_{k2} = 0.45$. Then we can derive $\bar{\alpha}_{Ej}$ and $\bar{\alpha}_{Tj}$ from,

$$\tilde{p}_{kj} = p(\tilde{y}_k > 0|j) = 1 - \Phi(0|\sum_{r=1}^{j} \bar{\alpha}_{kr}, 1). \tag{13}$$

Using the same elicited probabilities of toxicity and efficacy, $\tilde{p}_{kj}$, described in Section 2.2, we obtain hyper-parameter values of $\bar{\alpha}_E = [-0.84, 0.317, 0.255, 0.255, 0.255]$ and $\bar{\alpha}_T = [-1.645, 0.34, 0.44, 0.31, 0.15]$. The prior for $\tilde{\rho}$ follows Beta(a_ρ, b_ρ). We set the prior parameters $a_\rho = b_\rho = 0.5$ to assume no dependence between the probability of efficacy and toxicity with an "prior sample size" of 1.

3.3 Posterior Simulation

The full joint posterior distribution is,

$$\mathcal{D}_{(t)} = \{y_{Ei}, y_{Ti}, d_{[i]}, i = 1, ..., n(t)\},$$
$$p(\tilde{\boldsymbol{y}}, \alpha_{kj}, \rho|\mathcal{D}_{(t)}) \propto \prod_{k\in\{E,T\}} \prod_{i=1}^{n(t)} p(\tilde{\boldsymbol{y}}|\mathcal{D}_{(t)}, \alpha_{kj}, \rho)p(\alpha_{kj})p(\rho)\{1(\tilde{y}_{ki} \geq 0)y_{ki} + 1(\tilde{y}_{ki} < 0)(1 - y_{ki})\}. \tag{14}$$

The full conditional distributions of our random variables are displayed in Equations 19-24. With the full conditional distributions for $\tilde{y}_{Ei}, \tilde{y}_{Ti}, \alpha_{Ej}$, and α_{Tj}, we obtain posterior distributions with Gibbs sampling. Since we can not easily decipher the posterior distribution for $\tilde{\rho}$, we can generate posterior samples with the Metropolis Hastings Algorithm. The posterior probability of efficacy and toxicity, p_{kj} is then calculated with equation (13) using the posterior samples of α_{kj}.

4 Utility

We characterize the most optimal dose level as the dose level with the highest utility. The utility of a dose level, U_j, accounts for trade-offs between efficacy and toxicity with an extra, negative weight assigned to dose levels with toxicity above an upper threshold. We implement the utility function from Liu and Johnson [2016] defined as,

$$U_j(p_{Ej}, p_{Tj}) = p_{Ej} - w_1 p_{Tj} - w_2 p_{Tj} I(p_{Tj} > \phi_T). \tag{15}$$

Here, p_{Ej} and p_{Tj} are the probabilities of observing efficacy and toxicity outcomes for dose j. $I(p_{Tj} > \phi_t)$ indicates excessive toxicity where ϕ_T is the upper toxicity threshold determined by physicians. When this threshold is met, an additional penalty is added to the

utility function with weight w_2. The non-negative weights, w_1 and w_2, demonstrate the desired trade offs of higher toxicity levels. A more conservative approach would see higher values for the weights, which would put a larger negative affect of toxicity. A high value for w_2 likely favors dose levels that have toxicity below the toxicity threshold regardless of the efficacy. For example, with a $w_1 = 0.33$ and a $w_2 = 1.09$, we are putting a 33% penalty on dose levels with toxicity below ϕ_t and a $w1 + w2 = 142\%$ penalty on on dose levels with toxicity higher than ϕ_T. With these weights, the utility of a dose is bound between -1.42 and 1.

The values of w_1 and w_2 can be given directly or shaped to fit the desires of patients or physicians. They can be determined when given at least three pairs of efficacy and toxicity probabilities that are determined to be equally desirable. With all three pairs (p^*_{Ei}, p^*_{Ti}) having utilities $U^*_1 = U^*_2 = U^*_3$, where $U^*_i = U(p^*_{Ei}, p^*_{Ti})$, we can calculate the weights by minimizing the expression:

$$(w_1, w_2) = \underset{(w_1, w_2)}{\operatorname{argmin}} \sum_{i=1}^{I} (U^*_i - \sum_{i=1}^{I} U^*_i / I). \tag{16}$$

This establishes a contour for the utility values when plotting p_E against p_T. With this contour, we can visually see how the utility responds to the relationship between efficacy and toxicity. Figure 1 shows the the relationship between utility, p_E and p_T with weights $w_1 = 0.33, w_2 = 1.09$. Note that the scale changes as $p_T > 0.3$ from the second weight component.

5 Trial Conduct

5.1 Stopping Rule

The stopping rule is implemented to prevent investigating doses with excessive toxicity or low efficacy. We let $\phi_E = 0.2$ be the lower threshold for efficacy and let $\phi_T = 0.3$ be the upper threshold for toxicity determined by physicians. We stop the trial early if all dose levels meet at least one of the following inequalities,

$$P(p_{Ej} < \phi_E | \mathcal{D}_t) > 1 - C_E \text{ and } P(p_{Tj} > \phi_T | \mathcal{D}_t) > 1 - C_T. \tag{17}$$

Here, C_E and C_T are predetermined thresholds, determined by statisticians from preliminary study, that require enough p_{Ej} estimates above ϕ_E and enough p_{Tj} estimates below ϕ_T to consider selecting a dose level as most optimal. In our simulation, we set $C_E = C_T = 0.2$. If these criteria are met, there is evidence that the dose levels may not be beneficial or may even be too dangerous to continue testing. We add the condition that the dose chosen as the optimal dose level at the end of the trial must meet the criteria in the inequalities from equation (17). This insures that we do not ultimately pick a dose level that we consider in-efficacious or too toxic. The original work of Liu and Johnson [2016] does not have this condition, but the condition is added to their design as well as our design for fair comparison.

5.2 Dose finding Algorithm

Given the utility function, the algorithm is as follows:

1. Treat the first cohort at the pre-determined starting dose level. We let the starting dose be the lowest dose level, $d_{[i=1]} = 1$.

2. Obtain $\mathcal{D}_{(t)}$ and update posterior distributions of p_{kj}. Then update the posterior distributions U_j's for $j = 1, ..., J$.

3. Let $j_{\max}$ be the highest dose that has been administered in the trial so far. Let j^* be the dose of highest expected utility, $j^* = \underset{j \leq (j_{\max}+1)}{\operatorname{argmax}} E(U_j|\mathcal{D}_t)$. Note that this implies a dose level will not be tested unless the previous dose level has been tested.

4. Let A_t be the group of potential dose levels j to be administered to the next cohort, where $A_t = \{j; j \in \{j^* - 1, j^*, j^* + 1\}, j \leq j_{\max} + 1\}$. The dose level administered to the next cohort is sampled from A_t with probabilities proportional to $E(U_j|\mathcal{D}_{(t)})$. Since $E(U_j|\mathcal{D}_{(t)})$ has lower bound $-(w_1 + w_2)$, we add $(w_1 + w_2)$ to $E(U_j|\mathcal{D}_{(t)})$ so that the proportions are strictly positive.

5. The stopping rule described in Equation (17) is enacted to end a trial early if acceptability precautions are breached. In such a case, no dose is chosen as optimal.

6. Repeat steps 2-5.

7. When the maximum sample size of $N_{\max} = 48$ is reached, we determine that the optimal dose level is the dose level with the highest expected utility amongst the dose levels that have been tested and do not exceed the limitations set by the stopping rule in equation (17).

6 Simulation Studies

6.1 Simulation Setup

For consistency, we imitated the operating characteristics from Liu and Johnson [2016] studying doses at levels (0.25, 0.5, 0.75, 1.00, 1.25), $j = 1, ..., 5$, over $n = 48$ patients, with cohorts of size 3. We simulated with 5 different dose–response relationships from Liu and Johnson [2016] to compare algorithms in a series of practical scenarios. We simulated 300 trials for each scenario.

For testing the dose-finding methods, we simulate y_{ki}'s to represent results from patients. We can model y_{Ei} and y_{Ti} jointly, given dose $d_{[i]}$ for patient i, as $\pi_{a,b,j} = \Pr(Y_E = a, Y_T = b|j = d_{[i]})$ for $a, b \in \{0, 1\}$. The joint distribution of y_{Ei} and y_{Ti} is the Gumbel distribution,

$$\begin{aligned}\pi_{a,b,j} =&(p_{Ej}^{TR})^a(1 - p_{Ej}^{TR})^{1-a}(p_{Tj}^{TR})^b(1 - p_{Tj}^{TR})^{1-b} \\ &+ (-1)^{a+b}p_{Ej}^{TR}(1 - p_{Ej}^{TR})p_{Tj}^{TR}(1 - p_{Tj}^{TR})(\frac{e^\gamma - 1}{e^\gamma + 1}).\end{aligned} \tag{18}$$

In Equation (18), a and b represent efficacy and toxicity outcomes. We use the Gumbel distribution to simulate patient outcomes for possible efficacy and toxicity outcome pairings,

$(a, b) \in \{(1,1),(1,0),(0,1),(0,0)\}$. p_{Ej}^{TR} and p_{Tj}^{TR} represent the true probabilities of efficacy and toxicity for a given dose level shown in Table 1. The true utilities are computed with P_{Ej}^{TR} and P_{Tj}^{TR} and are shown in Table 1. We considered five simulation scenarios from Liu and Johnson [2016] which cover a realistic range of dose–response relationships. For generating data with the Gumbel model, we use a $\gamma = 3$.

6.2 Simulation Results

The simulation results are shown in Table 1. The table displays the selection percentage of each dose level as the optimal dose level and the percentage of patients treated at each dose level throughout the trial in parentheses, with the true optimal dose level in bold. The "None" section identifies the percentage of simulated trials that were stopped early due to excessive toxicity or inadequate efficacy. The same utility function and trial methodology was used for both designs. Five dose-toxicity and dose-efficacy scenarios from Liu and Johnson [2016] were used to compare the designs.

Scenarios where the most optimal dose level has a true utility significantly higher than the neighboring dose levels or scenarios where the neighboring dose levels are not accepted because of Equation (17), the bivariate probit model had more success. This is most noticeable in scenario 2, where the most optimal dose level, dose level 2, has neighboring dose levels have true efficacy and toxicity levels not accepted by Equation (17). Looking at the proportion of doses assigned to patients, we see that the selection of dose levels given to patients for the probit model were more centered around the true optimal dose level in scenarios 1 and 3, but not in scenarios 2 and 4. In scenario 2, the probit model gave much less doses to dose level 1, where both efficacy and toxicity were very low. Alternatively, in scenario 4, where dose level 1 had higher toxicity than efficacy, the first dose level treated more patients, possibly assuming a stronger efficacy because of the dependence with toxicity. In scenario 5, where the true efficacy and toxicity probabilities all trigger our stopping rule, we see a significant difference between the two models.

7 Conclusion

The Bayesian model-based approaches have the advantage of borrowing information across doses, while not imposing rigid assumptions on the dose-toxicity and dose-efficacy curves. The proposed bivariate probit model yields effective results for a variety of dose-efficacy and dose-toxicity relationships, comparable with the JL model. For further comparison between models, a different utility function could be applied that jointly accounts for efficacy and toxicity. The dependency between efficacy and toxicity captured in the bivariate probit model may be more significant under with a this criterion like the bivariate utility function purposed in Yuxi Tao and Yan [2013].

Additional study into optimizing the dose choice strategy for future patients may provide a better distribution of dose samples. Treating cohorts with dose levels far from optimal may not provide us useful information, especially if we have learned, from neighboring dose levels, that a particular dose level is not suitable. For example, in scenario 3, we see that the doses were administered pretty evenly amongst dose levels 2-5. We may want the current utilities to have more influence on the dose choice for the next cohort.

Scenario	Summary	None	Dose1	Dose2	Dose3	Dose4	Dose5
1	$({p_E}^{TR}, {p_T}^{TR})$		**(0.28, 0.15)**	(0.3, 0.32)	(0.44, 0.45)	(0.6, 0.55)	(0.74, 0.62)
	U^{TR}		**0.23**	-0.15	-0.2	-0.18	-0.14
	LJ	0.107	0.62(0.386)	0.23(0.437)	0.04(0.126)	0(0.042)	0(0.009)
	PR	0.144	0.676(0.492)	0.18(0.443)	0(0.064)	0(0.002)	0(0)
2	$({p_E}^{TR}, {p_T}^{TR})$		(0.1, 0.04)	**(0.27, 0.18)**	(0.44, 0.37)	(0.58, 0.54)	(0.69, 0.67)
	U^{TR}		0.07	**0.2**	-0.03	-0.02	-0.01
	LJ	0	0.103(0.298)	0.627(0.404)	0.223(0.228)	0.027(0.065)	0(0.006)
	PR	0.14	0.01(0.18)	0.76(0.32)	0.08(0.32)	0.01(0.16)	0(0.02)
3	$({p_E}^{TR}, {p_T}^{TR})$		(0.05, 0.02)	(0.08, 0.05)	(0.15, 0.07)	(0.28, 0.10)	**(0.43, 0.12)**
	U^{TR}		0.04	0.06	0.13	0.25	**0.39**
	LJ	0.01	0.00(0.099)	0.01(0.227)	0.147(0.23)	0.247(0.263)	0.587(0.181)
	PR	0.125	0(0.144)	0(0.210)	0.04(0.176)	0.17(0.266)	0.665(0.204)
4	$({p_E}^{TR}, {p_T}^{TR})$		(0.02, 0.10)	(0.10, 0.12)	**(0.42, 0.15)**	(0.45, 0.3)	(0.5, 0.6)
	U^{TR}		-0.01	0.06	**0.37**	0.35	-0.35
	LJ	0.02	0(0.096)	0.033(0.296)	0.713(0.319)	0.23(0.233)	0.003(0.055)
	PR	0.14	0(0.189)	0.02(0.350)	0.71(0.278)	0.12(0.155)	0.01(0.028)
5	$({p_E}^{TR}, {p_T}^{TR})$		(0.2, 0.1)	(0.05, 0.25)	(0.35, 0.55)	(0.40, 0.60)	(0.52, 0.70)
	U^{TR}		-0.01	-0.03	-0.43	-0.45	-0.47
	LJ	0.607	0(0.368)	0.06(0.431)	0.32(0.139)	0.0067(0.046)	0(0.015)
	PR	0.93	0(0.489)	0.037(0.454)	0.033(0.064)	0(0.002)	0(0.000)

Table 1: Simulation results with the proportion of times dose j was selected as the most optimal dose at the end of a trial, and in parentheses, the proportion of administered doses to patients. Here PR are the results for the probit model and LJ are the results for Liu and Johnson [2016].

The dose finding algorithm assumes that the patients full results are reported before the testing another cohort of patients. In practice, this may not be very realistic due to varying drug response times, patient recruitment, and time sensitive needs of testing. A model that incorporates time delay may be beneficial. An example is demonstrated in Liu and Johnson [2016], where they model p_{kj} with a weight representing the proportion of time a patient has been tested of the total time necessary to have conclusive results. This way, patients that have had the necessary time to show responses will be weighted more heavily than patients that have not.

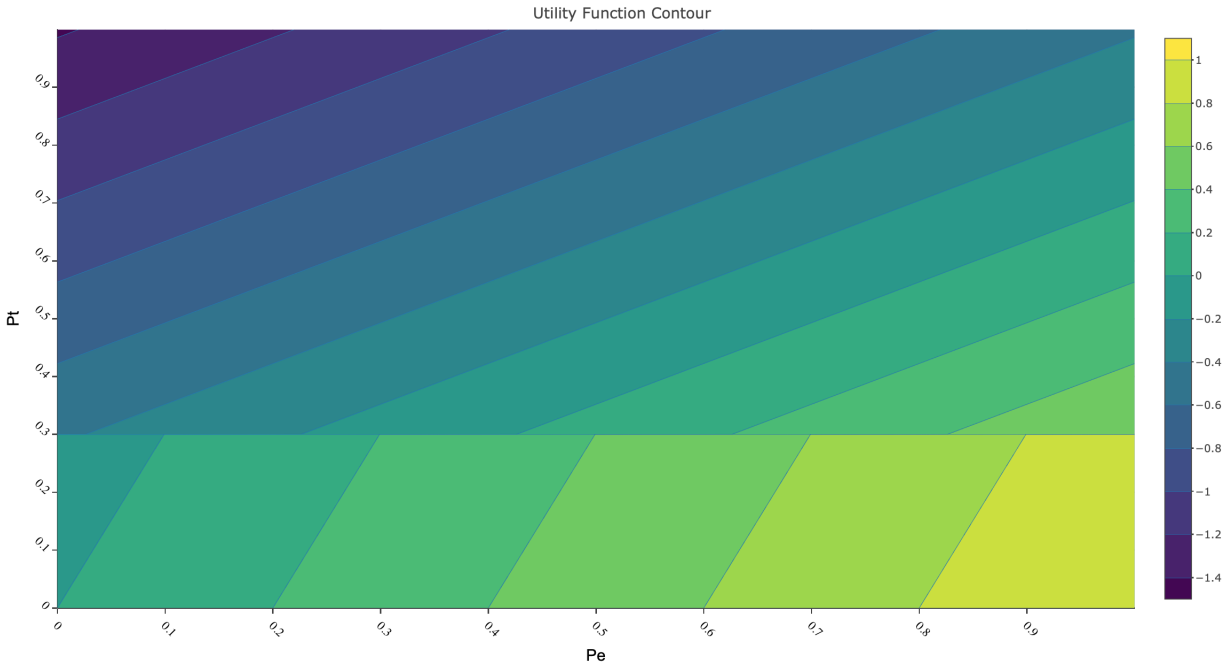

Figure 1: Contour of the utility function with weights $w_1 = 0.33$ and $w_2 = 1.09$

8 Appendix

8.1 Derivation of Full Conditional Posteriors for Bivariate Probit Model

8.1.1 Update $\tilde{y}_{Ei}$

$$
\begin{aligned}
p(\tilde{y}_{Ei}|\tilde{y}_{Ti}, \mu_{Ej}, \mu_{Tj}, \tilde{\rho}, \mathcal{D}) \propto & \{1(\tilde{y}_{Ei} \geq 0) y_{Ei} + 1(\tilde{y}_{Ei} < 0)(1 - y_{Ei})\} \\
& N(\mu_{Ej} + [2(\tilde{\rho} - \frac{1}{2})]\frac{\sigma_E}{\sigma_T}(\tilde{y}_{Ti} - \mu_{Tj}), \sigma_E^2(1 - [2(\tilde{\rho} - \frac{1}{2})]^2) \\
= & \{1(\tilde{y}_{Ei} \geq 0) y_{Ei} + 1(\tilde{y}_{Ei} < 0)(1 - y_{Ei})\} \\
& N(\sum_{r=1}^{d_{[i]}} \alpha_{Er} + (2\tilde{\rho} - 1)(\tilde{y}_{Ti} - \sum_{r=1}^{d_{[i]}} \alpha_{Tr}), 1 - (2\tilde{\rho} - 1)^2)
\end{aligned}
$$

$$
\text{If } y_{Ei} = 1, \;\; \tilde{y}_{Ei}|\tilde{y}_{Ti}, \mu_{Ej}, \mu_{Tj}, \tilde{\rho}, \mathcal{D} \sim N_+(\sum_{r=1}^{d_{[i]}} \alpha_{Er} + (2\tilde{\rho} - 1)(\tilde{y}_{Ti} - \sum_{r=1}^{d_{[i]}} \alpha_{Tr}), 1 - (2\tilde{\rho} - 1)^2)
$$

$$
\text{If } y_{Ei} = 0, \;\; \tilde{y}_{Ei}|\tilde{y}_{Ti}, \mu_{Ej}, \mu_{Tj}, \tilde{\rho}, \mathcal{D} \sim N_-(\sum_{r=1}^{d_{[i]}} \alpha_{Er} + (2\tilde{\rho} - 1)(\tilde{y}_{Ti} - \sum_{r=1}^{d_{[i]}} \alpha_{Tr}), 1 - (2\tilde{\rho} - 1)^2) \tag{19}
$$

8.1.2 Update $\tilde{y}_{Ti}$

$$
\begin{aligned}
p(\tilde{y}_{Ti}|\tilde{y}_{Ei},\mu_{Ej},\mu_{Tj},\tilde{p},\mathcal{D}) \propto & \{1(\tilde{y}_{Ti}\geq 0)y_{Ti}+1(\tilde{y}_{Ti}<0)(1-y_{Ti})\} \\
& N(\mu_{Tj}+[2(\tilde{p}-\frac{1}{2})]\frac{\sigma_E}{\sigma_T}(\tilde{y}_{Ei}-\mu_{Ej}),\sigma_E^2(1-[2(\tilde{p}-\frac{1}{2})]^2) \\
= & \{1(\tilde{y}_{Ti}\geq 0)y_{Ti}+1(\tilde{y}_{Ti}<0)(1-y_{Ti})\} \\
& N(\sum_{r=1}^{d_{[i]}}\alpha_{Tr}+(2\tilde{p}-1)(\tilde{y}_{Ei}-\sum_{r=1}^{d_{[i]}}\alpha_{Er}),1-(2\tilde{p}-1)^2)
\end{aligned}
$$

$$
\text{If } y_{Ti}=1,\ \ \tilde{y}_{Ti}|\tilde{y}_{Ei},\mu_{Ej},\mu_{Tj},\tilde{\rho},\mathcal{D}\sim N_+(\sum_{r=1}^{d_{[i]}}\alpha_{Tr}+(2\tilde{p}-1)(\tilde{y}_{Ei}-\sum_{r=1}^{d_{[i]}}\alpha_{Er}),1-(2\tilde{p}-1)^2)
$$

$$
\text{If } y_{Ti}=0,\ \ \tilde{y}_{Ti}|\tilde{y}_{Ei},\mu_{Ej},\mu_{Tj},\tilde{\rho},\mathcal{D}\sim N_-(\sum_{r=1}^{d_{[i]}}\alpha_{Tr}+(2\tilde{p}-1)(\tilde{y}_{Ei}-\sum_{r=1}^{d_{[i]}}\alpha_{Er}),1-(2\tilde{p}-1)^2) \tag{20}
$$

8.1.3 Update α_{Ej}

$$
\begin{aligned}
p(\alpha_{Ej}|\tilde{y}_{Ti},\tilde{y}_{Ei},\alpha_{Tj},\tilde{\rho},\mathcal{D},\alpha_{E,-j}) \propto & \prod_{i=1}^{n(t)}p(\tilde{y}_{Ei}|\tilde{y}_{Ti},\mu_{Ej},\mu_{Tj},\tilde{\rho})p(\mu_{Ej})p(\mu_{Tj})p(\tilde{\rho}) \\
\propto & p(\tilde{y}_{Ei}|\tilde{y}_{Ti},\alpha_{Ej},\alpha_{Tj},\tilde{p})p(\alpha_{Ej}) \\
& \text{Let: } I(\tau_E)\begin{cases}\tau_{E1}^2 \text{ if } j=1\\ \tau_{E2}^2 \text{ if } j>1\end{cases} \\
= & \prod_{i=1}^{n(t)}N(\tilde{y}_{Ei};\sum_{r=1}^{d_{[i]}}\alpha_{Er}+(2\tilde{\rho}-1)(\tilde{y}_{Ti}-\sum_{r=1}^{d_{[i]}}\alpha_{Tr}),1-[2\tilde{\rho}-1]^2)N(\alpha_{Ej};\bar{\alpha}_{Ej},I(\tau_E)^2) \\
\propto & \prod_{i=1}^{n(t)}\exp\left\{-\frac{[\tilde{y}_{Ei}-\sum_{r=1}^{d_{[i]}}\alpha_{Er}-\rho(\tilde{y}_{Ti}-\sum_{r=1}^{d_{[i]}}\alpha_{Tr})]^2}{2(1-[2\tilde{\rho}-1]^2)}\right\}\exp\left\{-\frac{(\alpha_{Ej}-\bar{\alpha}_{Ej})^2}{2I(\tau_E)^2}\right\}
\end{aligned}
$$

$$
\begin{aligned}
\text{Let, } z_{Ei}=\tilde{y}_{Ei}-\rho(\tilde{y}_{Ti}-\sum_{r=1}^{d_{[i]}}\alpha_{Tr}) & \\
= & \exp\left\{-\sum_{i=1}^{n(t)}\frac{(z_{Ei}-\sum_{r=1}^{d_{[i]}}\alpha_{Er})^2}{2(1-[2\tilde{\rho}-1]^2)}\right\}\exp\left\{-\frac{(\alpha_{Ej}-\bar{\alpha_{Ej}})^2}{2I(\tau_E)^2}\right\} \\
= & \exp\left\{-\sum_{d_{[i]}\geq j}^{n(t)}\frac{(z_{Ei}-\sum_{r=1}^{d_{[i]}}1(r\neq d_{[i]})\alpha_{Er}-\alpha_{Ej})^2}{2(1-[2\tilde{\rho}-1]^2)}\right\}\exp\left\{-\frac{(\alpha_{Ej}-\bar{\alpha_{Ej}})^2}{2I(\tau_E)^2}\right\}
\end{aligned}
$$

$$
\alpha_{Ej}|\tilde{y}_{Ti},\tilde{y}_{Ei},\alpha_{Tj},\tilde{\rho},\mathcal{D},\alpha_{E,-j}\sim\begin{cases}N(\mu_{\alpha_{E1}},\sigma^2_{\alpha_{E1}})\text{if } j=1\\ N_+(\mu_{\alpha_{Ej}},\sigma^2_{\alpha_{Ej}}) \text{ if } j>1\end{cases} \tag{21}
$$

8.1.4 Update α_{Tj}

$$
\begin{aligned}
p(\alpha_{Tj}|\tilde{y}_{Ti},\tilde{y}_{Ei},\alpha_{Ej},\tilde{\rho},\mathcal{D},\alpha_{E,-j}) &\propto \prod_{i=1}^{n(t)} p(\tilde{y}_{Ti}|\tilde{y}_{Ei},\mu_{Ej},\mu_{Tj},\tilde{\rho})p(\mu_{Ej})p(\mu_{Tj})p(\tilde{\rho}) \\
&\propto p(\tilde{y}_{Ti}|\tilde{y}_{Ei},\alpha_{Ej},\alpha_{Tj},\tilde{p})p(\alpha_{Tj}) \\
&\quad \text{Let: } I(\tau_T)\begin{cases}\tau_{T1}^2 \text{ if } j=1\\ \tau_{T2}^2 \text{ if } j>1\end{cases} \\
&= \prod_{i=1}^{n(t)} N(\tilde{y}_{Ei};\sum_{r=1}^{d_{[i]}}\alpha_{Tr}+(2\tilde{\rho}-1)(\tilde{y}_{Ei}-\sum_{r=1}^{d_{[i]}}\alpha_{Er}),1-[2\tilde{\rho}-1]^2)N(\alpha_{T1};\bar{\alpha}_{T1},I(\tau_T)^2) \\
&\propto \prod_{i=1}^{n(t)} \exp\left\{-\frac{[\tilde{y}_{Ti}-\sum_{r=1}^{d_{[i]}}\alpha_{Tr}-\rho(\tilde{y}_{Ei}-\sum_{r=1}^{d_{[i]}}\alpha_{Er})]^2}{2(1-[2\tilde{\rho}-1]^2)}\right\}\exp\left\{-\frac{(\alpha_{Tj}-\bar{\alpha}_{Tj})^2}{2I(\tau_T)^2}\right\}
\end{aligned}
$$

$$
\begin{aligned}
\text{Let, } z_{Ti} &= \tilde{y}_{Ti}-\rho(\tilde{y}_{Ei}-\sum_{r=1}^{d_{[i]}}\alpha_{Er}) \\
&= \exp\left\{-\sum_{i=1}^{n(t)}\frac{(z_{Ti}-\sum_{r=1}^{d_{[i]}}\alpha_{Tr})^2}{2(1-[2\tilde{\rho}-1]^2)}\right\}\exp\left\{-\frac{(\alpha_{Tj}-\bar{\alpha_{Tj}})^2}{2I(\tau_T)^2}\right\} \\
&= \exp\left\{-\sum_{d_{[i]}\geq j}^{n(t)}\frac{(z_{Ti}-\sum_{r=1}^{d_{[i]}}1(r\neq d_{[i]})\alpha_{Tr}-\alpha_{Tj})^2}{2(1-[2\tilde{\rho}-1]^2)}\right\}\exp\left\{-\frac{(\alpha_{Tj}-\bar{\alpha_{Tj}})^2}{2I(\tau_T)^2}\right\}
\end{aligned}
$$

$$
\alpha_{Tj}|\tilde{y}_{Ei},\tilde{y}_{Ti},\alpha_{Ej},\tilde{\rho},\mathcal{D},\alpha_{T,-j} \sim \begin{cases}N(\mu_{\alpha_{T1}},\sigma^2_{\alpha_{T1}}) \text{if } j=1\\ N_+(\mu_{\alpha_{Tj}},\sigma^2_{\alpha_{Tj}}) \text{ if } j>1\end{cases} \tag{22}
$$

$$
\begin{aligned}
\mu_{\alpha_{E1}} &= \sigma^2_{\alpha_{E1}} \left\{ \sum_{i|d_{[i]}\geq j}^{n(t)} \frac{(z_{Ei} - \sum_{r=1}^{d_{[i]}} 1(r \neq 1)\alpha_{Er})}{1-(2\tilde{\rho}-1)^2} + \frac{\bar{\alpha}_{E1}}{{\tau_{E1}}^2} \right\} \\
\mu_{\alpha_{Ej}} &= \sigma^2_{\alpha_{Ej}} \left\{ \sum_{i|d_{[i]}\geq j}^{n(t)} \frac{(z_{Ei} - \sum_{r=1}^{d_{[i]}} 1(r \neq j)\alpha_{Er})}{1-(2\tilde{\rho}-1)^2} + \frac{\bar{\alpha}_{Ej}}{{\tau_{E2}}^2} \right\} \\
\sigma^2_{\alpha_{E1}} &= \left\{ \frac{\sum_{i=1}^{n(t)} 1(d[i] > j)}{[1-(2\tilde{p}-1)^2]} + \frac{1}{{\tau_{E1}}^2} \right\}^{-1} \\
\sigma^2_{\alpha_{Ej}} &= \left\{ \frac{\sum_{i=1}^{n(t)} 1(d[i] > j)}{[1-(2\tilde{p}-1)^2]} + \frac{1}{{\tau_{E2}}^2} \right\}^{-1} \\
\mu_{\alpha_{T1}} &= \sigma^2_{\alpha_{T1}} \left\{ \sum_{i|d_{[i]}\geq j}^{n(t)} \frac{(z_{Ti} - \sum_{r=1}^{d_{[i]}} 1(r \neq 1)\alpha_{Tr})}{1-(2\tilde{\rho}-1)^2} + \frac{\bar{\alpha}_{T1}}{{\tau_{T1}}^2} \right\} \\
\mu_{\alpha_{Tj}} &= \sigma^2_{\alpha_{Tj}} \left\{ \sum_{i|d_{[i]}\geq j}^{n(t)} \frac{(z_{Ti} - \sum_{r=1}^{d_{[i]}} 1(r \neq j)\alpha_{Tr})}{1-(2\tilde{\rho}-1)^2} + \frac{\bar{\alpha}_{Tj}}{{\tau_{T2}}^2} \right\} \\
\sigma^2_{\alpha_{T1}} &= \left\{ \frac{\sum_{i=1}^{n(t)} 1(d[i] > j)}{[1-(2\tilde{p}-1)^2]} + \frac{1}{{\tau_{T1}}^2} \right\}^{-1} \\
\sigma^2_{\alpha_{Tj}} &= \left\{ \frac{\sum_{i=1}^{n(t)} 1(d[i] > j)}{[1-(2\tilde{p}-1)^2]} + \frac{1}{{\tau_{T2}}^2} \right\}^{-1}
\end{aligned}
\tag{23}
$$

8.1.5 Update

$$
\begin{aligned}
p(\tilde{\rho}|\tilde{y}_{Ti}, \tilde{y}_{Ei}, \mu_{Ej}, \mu_{Tj}, \mathcal{D}_{(t)}) &\propto p(\tilde{\boldsymbol{y}}_{\boldsymbol{i}}|\mu_{Ej}, \mu_{Tj}, \tilde{\rho}) p(\mu_{Ej}) p(\mu_{Tj}) p(\tilde{\rho}) \\
&\propto p(\tilde{\boldsymbol{y}}_{\boldsymbol{i}}|\alpha_{Ej}, \alpha_{Tj}, \tilde{\rho}) p(\tilde{\rho}) \\
&= \prod_{i=1}^{n(t)} N(\tilde{\boldsymbol{y}}_{\boldsymbol{i}}; \begin{bmatrix} \sum_{r=1}^{d_{[i]}} \alpha_{Er} \\ \sum_{r=1}^{d_{[i]}} \alpha_{Tr} \end{bmatrix}, \begin{bmatrix} 1 & 2\tilde{\rho}-1 \\ 2\tilde{\rho}-1 & 1 \end{bmatrix}) Be(\tilde{\rho}; a_\rho, b_\rho)
\end{aligned}
\tag{24}
$$

Bibliography

J. Jack Lee Christophe Le Tourneau and Lillian L. Siu. Dose escalation methods in phase i cancer clinical trials. (101):708–720, 2009.

K. H. Lu M. R. Gilbert F. Yan, P. F. Thall and Y. Yuan. Phase i–ii clinical trial design: a state-of-the-art paradigm for dose finding. (29):694–699, 2018.

Suyu Liu and Valen E. Johnson. A robust bayesian dose-finding design for phase i/ii clinical trials. (2):249–263, 2016.

Gareth O. Roberts and Jeffrey S. Rosenthal. Examples of adaptive mcmc. 2006.

Zhihui Li Jinguan Lin Tao Lu Yuxi Tao, Junlin Liu and Fangrong Yan. Dose-finding based on bivariate efficacy-toxicity outcome using archimedean copula. 2013.