Biometrics 62, 777–787
September 2006

DOI: 10.1111/j.1541-0420.2006.00534.x

Bayesian Dose-Finding in Phase I/II Clinical Trials Using Toxicity and Efficacy Odds Ratios

Guosheng Yin,* Yisheng Li, and Yuan Ji

Department of Biostatistics and Applied Mathematics,
University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030, U.S.A.
**email:* gsyin@mdanderson.org

Summary. A Bayesian adaptive design is proposed for dose-finding in phase I/II clinical trials to incorporate the bivariate outcomes, toxicity and efficacy, of a new treatment. Without specifying any parametric functional form for the drug dose–response curve, we jointly model the bivariate binary data to account for the correlation between toxicity and efficacy. After observing all the responses of each cohort of patients, the dosage for the next cohort is escalated, deescalated, or unchanged according to the proposed odds ratio criteria constructed from the posterior toxicity and efficacy probabilities. A novel class of prior distributions is proposed through logit transformations which implicitly imposes a monotonic constraint on dose toxicity probabilities and correlates the probabilities of the bivariate outcomes. We conduct simulation studies to evaluate the operating characteristics of the proposed method. Under various scenarios, the new Bayesian design based on the toxicity–efficacy odds ratio trade-offs exhibits good properties and treats most patients at the desirable dose levels. The method is illustrated with a real trial design for a breast medical oncology study.

Key words: Bayesian adaptive design; Bivariate binary model; Equivalence contour; Gibbs sampling; Trade-offs.

1. Introduction

Phase I clinical trials usually aim to find the maximum tolerated dose (MTD) for an investigational drug. It is common to assume that higher doses induce more severe toxicities, while efficacy is not considered in the design (Storer, 1989; O'Quigley, Pepe, and Fisher, 1990; Korn et al., 1994; Goodman, Zahurak, and Piantadosi, 1995; Møller, 1995; O'Quigley and Shen, 1996; Leung and Wang, 2002; among others). Traditional phase II trials are designed to examine the potential efficacy or any responsive activity of a new drug based on the MTD obtained from the phase I trials. However, with a very limited sample size in the phase I trial, the MTD might not be obtained in a reliable way, which thus affects the subsequent phase II and III trials. Therefore, it is critically important to find the optimal dosage of a drug which has the highest effectiveness as well as tolerable toxicity.

A typical dose-finding design is driven by one-level dose escalation or deescalation based on the cumulated data in a trial. Dose-finding solely based on toxicity while ignoring efficacy may not be the best strategy. Recently, there has been increasing research in the development of dose-finding methodologies based on both toxicity and efficacy outcomes (Gooley et al., 1994; Thall and Russell, 1998; Thall and Cheng, 1999; O'Quigley, Hughes, and Fenton, 2001; Braun, 2002; Ivanova, 2003; Thall and Cook, 2004; Bekele and Shen, 2005; among others). Many of these methods assume a parametric function to model the relationship between the toxicity–efficacy responses and the dose levels. However, at the early stage of drug evaluation, little is known about the behavior of the drug dose–response curve. It thus could be difficult to capture the response pattern based on the commonly used parametric forms since the underlying model structure is usually unknown (Mukhopadhyay, 2000). When searching for the optimal dose using a small number of subjects, this type of parametric modeling might lead to undesirable and unstable results when the specified parametric model is incorrect (Korn et al., 1994). Gasparini and Eisele (2000) proposed a curve-free dose-finding method by relaxing the rigid functional form between toxicities and dose levels. They focused on phase I trials by only considering toxicity and studied a product of the beta prior (PBP) to simplify the computation. However, the PBP could lead to undesirable operating characteristics under certain circumstances (Cheung, 2002). The work of Braun (2002) generalized the continual reassessment method (O'Quigley et al., 1990) to two competing outcomes, which however did not construct the proposed bivariate distribution from the marginal probabilities of the two outcomes of interest. Thall and Cook (2004) proposed to partition the two-dimensional toxicity–efficacy probability domain by introducing a trade-off contour. The set of contours was constructed with a polynomial model based on three physician-specified equivalent points, which might be subjective. In an adaptive fashion, the design by Thall and Cook (2004) updates the dose information coherently as more data are observed in the trial, and the next cohort of patients will then be treated at the current best dose in an admissible set. Bekele and Shen

(2005) investigated a joint distribution of a binary and a continuous outcome by introducing latent variables in a probit model.

Our aim is to make a design more efficient and ethical, as well as to save resources, by incorporating the different perspectives of a dose-finding trial. Toward this goal, we propose a single-arm Bayesian adaptive design based on toxicity and efficacy odds ratio trade-offs. The design evaluates doses by considering the toxicity and efficacy outcomes simultaneously, which is thus suitable for a combination of traditional phase I/II clinical trials. In order to find the optimal dose at which as many patients as possible can be treated, we derive a Bayesian adaptive decision-making procedure based on the odds ratio equivalence contour between toxicity and efficacy. This method eliminates the subjectivity of the trade-off contours according to the three physician-specified equivalent points, and inherits a natural interpretation of the odds ratio. In this procedure, patients are recruited sequentially in cohorts, and decisions are made on dose escalation, deescalation, or staying at the current dose level after observing the responses of each cohort. An order constraint in the dose toxicity probabilities is imposed so that higher doses are associated with more excessive toxicities. Although constrained parameter problems often make Bayesian computation and inference difficult, we propose a class of transformations on the parameters that naturally accommodates the order constraint on the probabilities of toxicity. No constraints are assumed for the probabilities of efficacy. Hence, the proposed method is quite general and suitable for clinical trials with cytotoxic and biological agents.

The research work in this article was motivated from a phase I/II clinical trial design conducted at the M. D. Anderson Cancer Center. The primary objective was to design an open-label, multi-center, dose-finding study to evaluate the efficacy and tolerability of RAD001 in combination with a standard 3-week cycle of docetaxel therapy in patients with metastatic breast cancer. As a novel macrolide, RAD001 was developed as an antiproliferative drug with applications as an immunosuppressant and antitumor agent. It acts on interleukin and the growth-factor-dependent proliferation of cells through their high affinity for an intracellular receptor protein. Docetaxel exerts its cytotoxic effect through the promotion of tubulin assembly and the delay of its depolymerization, which prevents normal mitosis, and alters the normal functions and skeletal structure of the cell. The investigators are interested in finding the optimal dose of the combination of RAD001 (three dose levels) and docetaxel (a constant dose level) in treating breast cancer patients.

The rest of the article is organized as follows. In Section 2.1, we propose the probability model and derive the likelihood function. In Section 2.2, we specify the joint prior distributions based on logit transformations and derive the posterior distributions for the implementation of Gibbs sampling in the Bayesian paradigm. In Section 2.3, we present the dose selection rule and the toxicity–efficacy odds ratio criteria. In Section 3, we conduct simulation studies to evaluate the properties of our Bayesian adaptive design. In Section 4, we illustrate the proposed methodologies with the phase I/II dose-finding breast oncology trial. Concluding remarks follow in Section 5.

2. Probability Model

2.1 *Likelihood Function*

In phase I/II clinical trials, we are concerned with finding a therapeutic dose of a new drug that maximizes the efficacy as well as controls the toxicity. Let Dose_j be the jth dose level for $j = 1, \ldots, d$, and $\text{Dose}_1 < \text{Dose}_2 < \cdots < \text{Dose}_d$. Let p_j and q_j be the probabilities of toxicity and efficacy associated with Dose_j. For toxicity, one usually assumes a monotonically increasing relationship between p_j and Dose_j, that is,

$$p_{j-1} < p_j, \quad j = 2, \ldots, d. \tag{1}$$

We do not impose the monotonic constraint for q_j since the efficacy for a certain therapy (e.g., cytotoxic agents) may decrease as the dose level increases.

We consider jointly modeling the toxicity and efficacy outcomes using a bivariate binary model. Let X_{ij} denote the toxicity outcome for subject i under dose j, and let Y_{ij} denote the efficacy outcome of the same subject, i.e., $X_{ij} = 1$ with probability p_j, or 0 with probability $1 - p_j$; $Y_{ij} = 1$ with probability q_j, or 0 with probability $1 - q_j$. Dale (1986) proposed the global cross-ratio as a measure of association which is suitable for bivariate, discrete, ordered responses. The global cross-ratio model can be formulated for the bivariate outcomes as follows (Dale, 1986). Define $\pi_{xy}^{(j)} = \Pr(X_{ij} = x, Y_{ij} = y)$ where $x, y = 0, 1$, and at dose level j, let

$$\theta_j = \frac{\pi_{00}^{(j)} \pi_{11}^{(j)}}{\pi_{01}^{(j)} \pi_{10}^{(j)}}, \quad j = 1, \ldots, d,$$

which quantifies the association between the two responses. Then, the probabilities $\pi_{xy}^{(j)}$ can be obtained from θ_j and the marginal probabilities p_j and q_j, i.e.,

$$\pi_{11}^{(j)} = \begin{cases} \left(a_j - \sqrt{a_j^2 + b_j}\right)/\{2(\theta_j - 1)\} & \theta_j \neq 1 \\ p_j q_j & \theta_j = 1, \end{cases}$$

$$\pi_{10}^{(j)} = p_j - \pi_{11}^{(j)}, \quad \pi_{01}^{(j)} = q_j - \pi_{11}^{(j)}, \quad \pi_{00}^{(j)} = 1 - p_j - q_j + \pi_{11}^{(j)}, \tag{2}$$

where $a_j = 1 + (p_j + q_j)(\theta_j - 1)$ and $b_j = -4\theta_j(\theta_j - 1)p_j q_j$. Let $\mathbf{p} = (p_1, \ldots, p_d)'$, $\mathbf{q} = (q_1, \ldots, q_d)'$, and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)'$. Suppose that n_j subjects are treated at Dose_j for $j = 1, \ldots, d$, then the likelihood function under (2) is

$$L(\mathbf{p}, \mathbf{q}, \boldsymbol{\theta} \,|\, \text{Data}) = \prod_{j=1}^{d} \prod_{i=1}^{n_j} \prod_{x=0}^{1} \prod_{y=0}^{1} \left\{\pi_{xy}^{(j)}\right\}^{I(X_{ij}=x, Y_{ij}=y)},$$

where $I(\cdot)$ is the indicator function.

2.2 *Prior Specification*

We propose two different transformations to the p_j's and q_j's for the specification of the priors with or without incorporating the inherent ordering constraint, respectively. To model toxicity, let

$$\phi_1 = \log\frac{p_1}{1-p_1}, \quad \phi_j = \log\left(\frac{p_j}{1-p_j} - \frac{p_{j-1}}{1-p_{j-1}}\right), \quad j = 2, \ldots, d,$$

and then

$$p_1 = \frac{e^{\phi_1}}{1+e^{\phi_1}}, \quad p_j = \frac{e^{\phi_1}+\cdots+e^{\phi_j}}{1+e^{\phi_1}+\cdots+e^{\phi_j}}, \quad j = 2, \ldots, d.$$

Clearly, the p_j's satisfy the monotonic condition (1). For efficacy, we do not enforce the ordering constraint with respect to the dose level, thus define

$$\psi_1 = \log\frac{q_1}{1-q_1}, \quad \psi_j = \log\left(\frac{q_j}{1-q_j}\right) - \log\left(\frac{q_{j-1}}{1-q_{j-1}}\right), \quad j = 2, \ldots, d.$$

Therefore,

$$q_1 = \frac{e^{\psi_1}}{1+e^{\psi_1}}, \quad q_j = \frac{e^{\psi_1+\cdots+\psi_j}}{1+e^{\psi_1+\cdots+\psi_j}}, \quad j = 2, \ldots, d.$$

Let $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_d)'$ and $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_d)'$. To induce the prior correlations between and within the probabilities of efficacy and toxicity among different dose levels, we assume a $2d$-dimensional multivariate normal prior distribution for $(\boldsymbol{\phi}, \boldsymbol{\psi})$,

$$\pi(\boldsymbol{\phi}, \boldsymbol{\psi}) = (2\pi)^{-d}|\boldsymbol{\Sigma}|^{-1/2}\exp\left\{-\frac{1}{2}(\boldsymbol{\phi}', \boldsymbol{\psi}')\boldsymbol{\Sigma}^{-1}(\boldsymbol{\phi}', \boldsymbol{\psi}')'\right\},$$

where $\boldsymbol{\Sigma}$ is a $2d \times 2d$ symmetric covariance matrix.

We take the prior for $\boldsymbol{\theta}$, $\pi(\boldsymbol{\theta})$, to be a log-normal distribution, and for ease of computation and implementation, all the θ_j's are assumed to be independent ($j = 1, \ldots, d$). An alternative prior for $\boldsymbol{\theta}$ is a gamma distribution since the support for $\boldsymbol{\theta}$ is the positive real line. If we assume *a priori* independence between $\boldsymbol{\theta}$ and $(\boldsymbol{\phi}, \boldsymbol{\psi})$, the joint posterior distribution is given by

$$\pi(\boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\theta} \,|\, \text{Data}) \propto L(\mathbf{p}(\boldsymbol{\phi}), \mathbf{q}(\boldsymbol{\psi}), \boldsymbol{\theta} \,|\, \text{Data})\pi(\boldsymbol{\phi}, \boldsymbol{\psi})\pi(\boldsymbol{\theta}),$$

from which the full conditional distributions of the parameters (ϕ_j, ψ_j, θ_j; $j = 1, \ldots, d$) can be obtained. Under noninformative priors, the likelihood dominates the posterior estimation and inference.

2.3 *Dose-Finding Criteria*

Following Thall and Russell (1998), we define a set of acceptable doses, $\mathcal{A}$(Dose), containing those Dose$_j$'s ($j = 1, \ldots, d$) for which the following two conditions are satisfied:

$$\Pr(p_j < \bar{\pi}_T) > p^*, \quad \Pr(q_j > \underline{\pi}_E) > q^*. \tag{3}$$

Here, $\bar{\pi}_T$ and $\underline{\pi}_E$ are physician-specified upper toxicity and lower efficacy limits, and p^* and q^* are fixed probability cutoffs.

To facilitate the dose selection procedure, we formulate the toxicity–efficacy odds ratio contour in the following way. For the optimal dose j, we expect its efficacy and toxicity probabilities (q_j, p_j) to be the closest to the lower-right corner (1, 0) in the two-dimensional efficacy and toxicity domain, as shown in Figure 1. The horizontal and vertical lines which cross point A (q_j, p_j) partition the probability square into four rectangles. It follows that the odds ratio $\omega_j^{(2)}$ between the toxicity and efficacy of dose j, defined by

$$\omega_j^{(2)} = \frac{p_j/(1-p_j)}{q_j/(1-q_j)} = \frac{p_j(1-q_j)}{(1-p_j)q_j}, \tag{4}$$

is exactly the ratio of the area of the lower-right versus that of the upper-left rectangle. A dose with a smaller value of $\omega_j^{(2)}$ is more desirable since it indicates higher efficacy and lower toxicity. Figure 1 shows an equivalent odds ratio contour, i.e., all the points along the curve have the same toxicity–efficacy odds ratio, $\omega_j^{(2)}$.

The odds ratio $\omega_j^{(2)}$ is based on the marginal probabilities (q_j, p_j). To incorporate the correlation between toxicity and efficacy in dose responses, we construct an alternative three-dimensional volume ratio by adding a third dimension of the conditional probability of efficacy given no toxicity, $\pi_{\mathrm{E}|\mathrm{T}^c}^{(j)}$. Thus, for dose j, let

$$\omega_j^{(3)} = \frac{p_j(1-q_j)\left(1-\pi_{\mathrm{E}|\mathrm{T}^c}^{(j)}\right)}{(1-p_j)q_j\pi_{\mathrm{E}|\mathrm{T}^c}^{(j)}} = \omega_j^{(2)}\frac{1-\pi_{\mathrm{E}|\mathrm{T}^c}^{(j)}}{\pi_{\mathrm{E}|\mathrm{T}^c}^{(j)}} = \omega_j^{(2)}\frac{\pi_{00}^{(j)}}{\pi_{01}^{(j)}}. \tag{5}$$

Figure 2 presents a three-dimensional probability space by adding an axis of the probability of efficacy given no toxicity ($\pi_{\mathrm{E}|\mathrm{T}^c}^{(j)}$) as the third scale. The point (1, 1, 0) represents the best combination for a given dose, i.e., $q_j = 1, \pi_{\mathrm{E}|\mathrm{T}^c}^{(j)} = 1$, and $p_j = 0$. In this probability cube, the horizontal and vertical planes across point A $(q_j, \pi_{\mathrm{E}|\mathrm{T}^c}^{(j)}, p_j)$ partition the unit cube into eight pieces. Focusing on the two cubes along the diagonal line, $\omega_j^{(3)}$ is the ratio between the volumes of the lower-left and the upper-right cubes. Therefore, the dose that yields the smallest value of $\omega_j^{(3)}$ is considered the best one that may be used to treat the next cohort of patients. Figure 2 presents an equivalent odds ratio surface crossing point A, i.e., all the points on this smooth surface have the same $\omega_j^{(3)}$.

Let $\tilde{q}_j = E(q_j \,|\, \text{Data})$ and $\tilde{p}_j = E(p_j \,|\, \text{Data})$ be the posterior means of the probabilities of efficacy and toxicity, and let $\tilde{\pi}_{00}^{(j)}$ and $\tilde{\pi}_{01}^{(j)}$ be the posterior means of $\pi_{00}^{(j)}$ and $\pi_{01}^{(j)}$, respectively. To implement a dose-finding clinical trial, we replace the unknown quantities in (4) and (5) by the corresponding posterior means. Based on the odds ratio criteria $\omega_j^{(2)}$ and $\omega_j^{(3)}$, the dose-finding algorithm is described as follows:

1. Patients in the first cohort are treated with the lowest dose level.
2. The dose will be escalated to the lowest untried dose level if the toxicity probability of the highest tried dose, denoted by p^{last}, satisfies

$$\Pr\left(p^{\text{last}} < \bar{\pi}_T\right) > p^{\text{escl}}, \tag{6}$$

   for some chosen cutoff probability of escalation, $p^{\text{escl}} \geq p^*$, where p^* is given in (3).
3. If a given dose j satisfies the two conditions in (3), Dose$_j \in \mathcal{A}$(Dose). If (6) is not satisfied and $\mathcal{A}$(Dose) is an empty set, then the trial is terminated and no dose is selected as long as the minimum sample size is reached.
4. Otherwise, patients in the next cohort are treated at the most desirable dose from $\mathcal{A}$(Dose) as determined by the

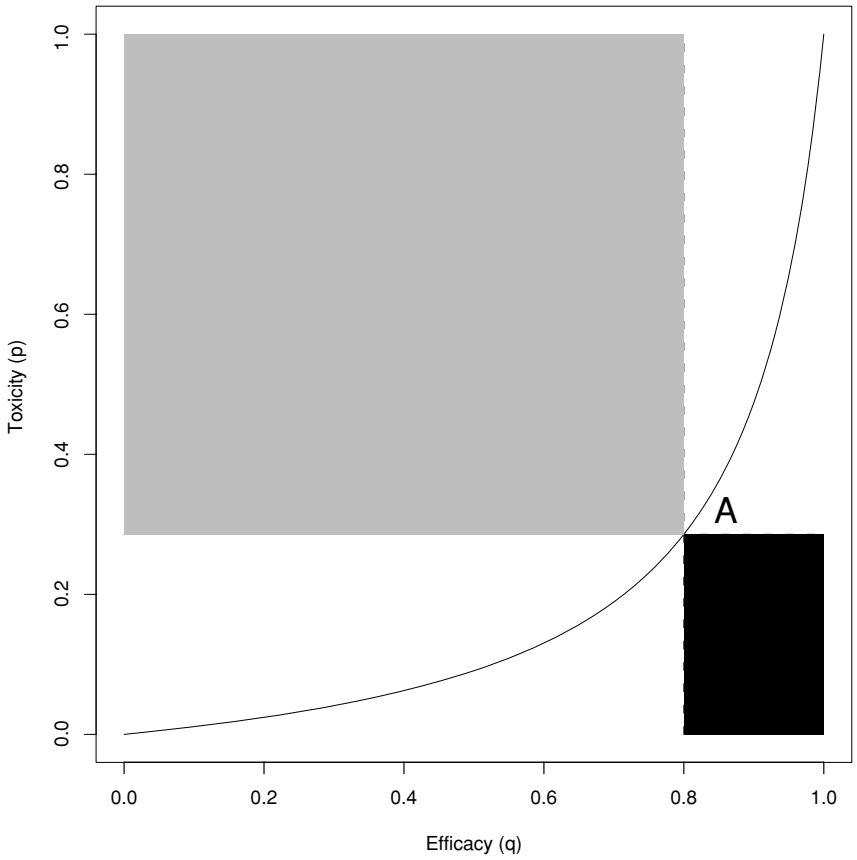

Figure 1. Two-dimensional toxicity–efficacy odds ratio trade-off contours with point A (q_j, p_j) corresponding to dose j.

odds ratio criterion, under the restriction that untried doses cannot be skipped when escalating or deescalating.

5. Once the maximum sample size is reached, the final dose Dose_f is selected such that $\text{Dose}_f \in \mathcal{A}(\text{Dose})$, and minimizes the toxicity–efficacy odds ratio.

Decisions on dose escalation, deescalation, or staying at the current dose should be determined by the observed data on the tried doses. Hence, the dose needs to be escalated as long as the highest tried dose level has not exceeded the toxicity threshold in (6). If (6) is not satisfied and $\mathcal{A}(\text{Dose})$ is an empty set, while the minimum sample size is not reached, the next cohort of patients will be treated at the most desirable dose as determined by the odds ratio criterion despite that $\mathcal{A}(\text{Dose})$ is empty. In our computer simulations, we set the minimum sample size at three, so the minimum sample size requirement is automatically met after the first cohort of patients is accrued to the study.

3. Simulation Studies

We conducted simulations to examine the operating characteristics of the proposed Bayesian dose-finding design. The maximum sample size was 60, the minimum was 3, and patients were accrued in cohorts of size 3. We considered 5 dose levels and 13 scenarios with different true probabilities of toxicity and efficacy, i.e., $\{(p_j, q_j), j = 1, \ldots, 5\}$. The actual dose levels were (0.25, 0.5, 0.75, 1, 2), and the first cohort of patients was treated at the lowest dose level. The upper toxicity and lower efficacy limits were taken to be $\bar{\pi}_T = \underline{\pi}_E = 0.3$, and the cutoff probabilities were $p^* = 0.25$ and $q^* = 0.1$ in (3), and $p^{\text{escl}} = 0.5$ in (6). For each configuration, we conducted 1000 simulated trials.

The priors for the Dale model parameters in our proposed designs were taken to be noninformative, e.g., $\text{var}(\phi_j) = \text{var}(\psi_j) = 100$ for $j = 1, \ldots, 5$. The off-diagonal elements in $\boldsymbol{\Sigma}$ characterizing the correlations were assigned as 0, since the prior correlation coefficients did not have much influence on the results based on our simulations. This robustness feature is very attractive since there is usually little prior information on these correlations and it is also difficult to elicit from physicians. Figure 3 shows the prior distributions of the five probabilities of toxicity and efficacy. The large prior variances of $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$ indeed induce noninformative priors on $\mathbf{p}$ and $\mathbf{q}$, and there appears to be an obvious trend of shifting to the right for toxicity due to the monotonic constraint on those probabilities. For the θ_j's, we took noninformative priors, i.e., $\log \theta_j \sim N(0, 10)$ for $j = 1, \ldots, 5$.

To implement Gibbs sampling, we derived the full conditional distribution for each model parameter and obtained the posterior samples using the adaptive rejection Metropolis sampling (ARMS) algorithm proposed by Gilks, Best, and Tan (1995). We took 1000 burn-ins and then recorded every fifth sample out of 5000 iterations to reduce the autocorrelation in the Markov chain, and based all the computations and

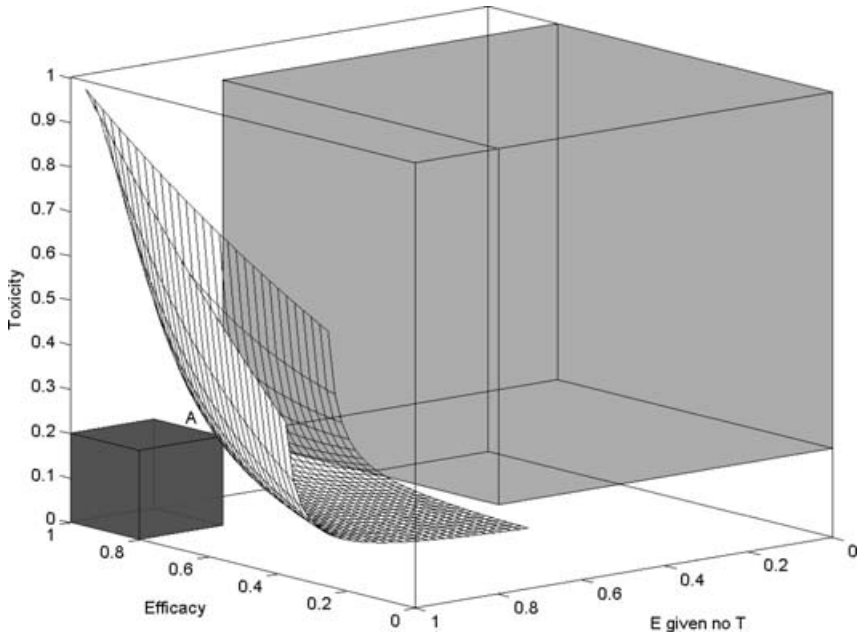

Figure 2. Three-dimensional toxicity–efficacy odds ratio trade-offs incorporating an axis of efficacy given no toxicity, with point A $(q_j, \pi^{(j)}_{\mathrm{E|T^c}}, p_j)$ corresponding to dose j.

inference strategies on the resulting 1000 posterior samples. The Markov chains converged very fast and mixed well. In a coherent and adaptive fashion, after the bivariate binary outcomes for each cohort of patients were observed, we sequentially updated the posterior distribution and reestimated the posterior means of parameters $(\tilde{p}_j, \tilde{q}_j, \tilde{\pi}^{(j)}_{00}, \tilde{\pi}^{(j)}_{01})$ for each dose. We then computed the odds ratios to determine the best dose at the current stage of the trial to treat the next cohort of patients.

For comparison, we examined the dose-finding method by Thall and Cook (2004) under the same 13 scenarios. They generated the bivariate binary toxicity–efficacy outcomes based on a Gumbel model (Murtaugh and Fisher, 1990). For $j = 1, \ldots, d$, the joint probability for the bivariate outcomes (x, y) is

$$\pi^{(j)}_{xy} = p_j^x (1-p_j)^{1-x} q_j^y (1-q_j)^{1-y} + (-1)^{x+y} p_j (1-p_j) q_j (1-q_j) \frac{e^\gamma - 1}{e^\gamma + 1}, \qquad (7)$$

where γ is an association parameter related to the correlation between toxicity and efficacy. A concave toxicity–efficacy trade-off contour was constructed based on three physician-specified points on the two-dimensional probability domain. For each given dose, they could obtain an intersection point between the trade-off contour and a straight line across the points of toxicity–efficacy probabilities and (1, 0). The desirability parameter δ_j is defined as a ratio of the Euclidean distance from the intersection point to (1, 0) versus that from the point of the toxicity–efficacy probabilities. A larger value of δ_j indicates a more desirable dose. To implement the design in Thall and Cook (2004), we set the prior mean probabilities of toxicity as (0.05, 0.15, 0.25, 0.35, 0.45) and those of efficacy as (0.15, 0.30, 0.45, 0.60, 0.75). The hyperparameters in the prior distributions were obtained by minimizing an objective function that quantifies the distance of the prior means and standard deviations from those elicited from physicians. The prior was specified to reasonably represent a physician's knowledge and uncertainty with respect to the drug. For details, see Thall and Cook (2004).

For ease of exposition, we refer to the methods based on $\omega^{(2)}_j$ and $\omega^{(3)}_j$ as the 2d-OR (two-dimensional odds ratio) and 3d-OR designs, and the proposal by Thall and Cook (2004) as the TC design. As suggested by a referee, another reasonable criterion is to select the dose with the largest joint posterior probability of $\pi^{(j)}_{01} = \Pr(\mathrm{E} = 1, \mathrm{T} = 0)$ to treat the next cohort of patients. Hence, we compare the operating characteristics of the above four different clinical trial designs in the simulation studies. The bivariate binary outcomes were simulated from the Dale model based on prespecified marginal probabilities of toxicity and efficacy with $\theta_j = 1$ (corresponding to $\gamma = 0$ in the Gumbel model, i.e. independent cases). Table A1 (available at `http://www.tibs.org/biometrics`) summarizes the simulation results with respect to the 13 different toxicity–efficacy probability configurations using the TC (δ_j), 2d-OR $(\omega^{(2)}_j)$, 3d-OR $(\omega^{(3)}_j)$, and $\pi^{(j)}_{01}$ criteria $(j = 1, \ldots, 5)$. For each scenario, the first row represents the true probability combinations (p_j, q_j) multiplied by 100; the second is the desirability value δ_j from the TC method; the third shows the selection percentages of the dose levels using the δ_j desirability criterion with the average number of patients treated at each dose given in the parentheses based on 1000 simulations; the fourth gives the true values of $\omega^{(2)}_j$; the fifth exhibits dose selection percentages and average numbers of treated patients using the 2d-OR criterion; the sixth corresponds to the true values of $\omega^{(3)}_j$; the seventh row is the selection percentages of doses using $\omega^{(3)}_j$; the eighth row gives the true value of the cell

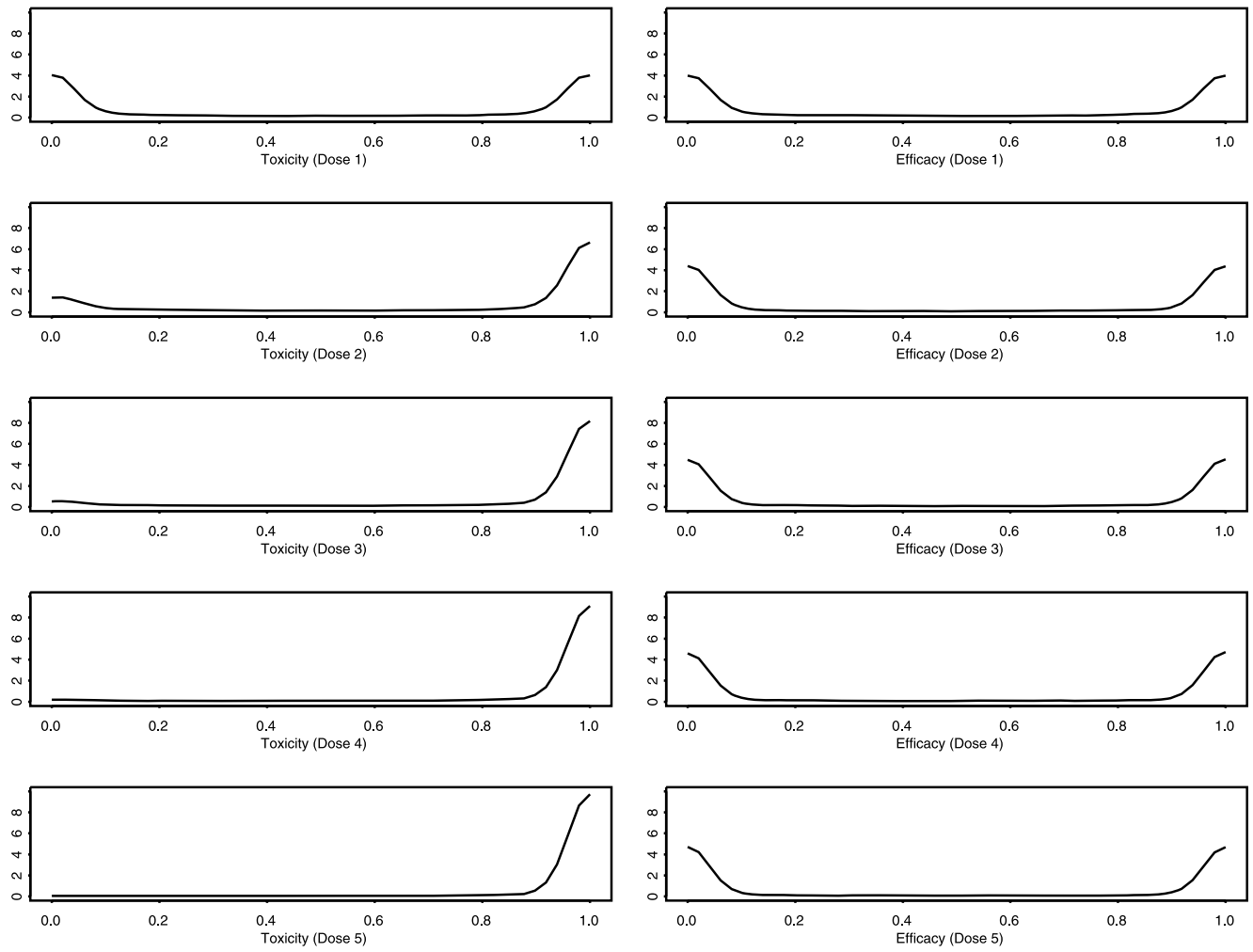

Figure 3. Prior densities for toxicity (p_j) and efficacy (q_j) for $j = 1, \ldots, 5$.

probability $\pi_{01}^{(j)}$; and the last row is the selection percentages corresponding to the proposal using $\pi_{01}^{(j)} = \Pr(\mathrm{E} = 1, \mathrm{T} = 0)$. The column of "None" represents the percentages of the 1000 simulated trials that did not select any of the doses, i.e., inconclusive trials.

Scenarios 1–9 represent the cases in which both toxicity and efficacy increase as the dose level increases, at various increasing rates. When toxicity increases substantially with respect to the dose level while efficacy does not change much (Scenario 1), all four methods yield very similar results by treating most patients at the first dose. Scenario 2 is the opposite case where toxicity is negligible but efficacy increases significantly over dose levels. The 3d-OR and $\pi_{01}^{(j)}$ criteria perform similarly while the TC and 2d-OR methods select the first dose with substantial percentages of around 20%. This scenario clearly demonstrates the advantage and importance of incorporating a third dimension $\pi_{\mathrm{E}|\mathrm{T}^c}^{(j)}$ as opposed to only using the marginal probabilities (p_j, q_j). Scenario 3 is a special case where all the $\omega_j^{(2)}$'s are equal to 1. In Scenario 4, all the four methods select the fifth dose with the highest percentages though the TC design performs relatively better. In Scenario 5, all five dose levels are toxic and thus none of them should be selected using any of these methods. The methods using $\omega_j^{(2)}$, $\omega_j^{(3)}$, and $\pi_{01}^{(j)}$ are more conservative and safer. In Scenario 6, the TC design behaves slightly better than other methods by mostly selecting the fourth dose. Scenarios 7–9 are constructed to accommodate other possible situations in the real trial, and we see that the proposed 2d-OR and 3d-OR methods behave well and are comparable to the TC and $\pi_{01}^{(j)}$ methods. In particular, when all the five doses are neither efficacious nor toxic (Scenario 9), all the designs are able to terminate the trial after trying all the doses while the TC method has 3.2% selection of dose 1 or dose 5.

In Scenarios 10 and 11, efficacy is not monotonically increasing over the dose level, i.e., the dose–response curve may have a parabolic shape: efficacy first increases and then decreases with increasing dose levels. It shows that our designs are quite conservative and robust, and tend not to treat patients at excessively toxic doses, although the percentage of selecting the optimal dose is slightly lower than that of the TC method.

We conducted a sensitivity analysis to examine the efficiency loss of the proposed nonparametric method compared to the TC method (the parametric model) in Scenarios 12

and 13. We generated the bivariate binary outcomes from the parametric logistic models and thus the probabilities of toxicity and efficacy were obtained from the following marginal logistic models with the dose level as a covariate,

$$\text{logit}(p_j) = \alpha_0 + \alpha_1 \text{Dose}_j,$$
$$\text{logit}(q_j) = \beta_0 + \beta_1 \text{Dose}_j + \beta_2 \text{Dose}_j^2, \quad j = 1, \ldots, 5.$$

In Scenario 12, the regression parameters were taken to be $\alpha_0 = -4.482$ and $\alpha_1 = 0.578$ for toxicity, and $\beta_0 = -0.059$, $\beta_1 = 1.207$, and $\beta_2 = 0$ for efficacy. Scenario 13 is the case with a nonmonotonic relationship between efficacy and the dose level, which therefore requires a quadratic covariate term in the model, with $\alpha_0 = -4.082$ and $\alpha_1 = 1.876$ for toxicity, and $\beta_0 = -3.937$, $\beta_1 = 7.605$, and $\beta_2 = -2.918$ for efficacy. Our methods perform better under Scenario 12 with higher percentages of selecting the last dose, while the TC method appears to be superior with a higher percentage of selecting the fourth dose under Scenario 13.

To assess the robustness of the proposed design with respect to the association parameter γ between toxicity and efficacy, we generated data with $\gamma = -2$ ($\theta_j \approx 0.47$) and $\gamma = 2$ ($\theta_j \approx 2.2$) based on the true marginal probabilities under Scenario 1. The likelihoods of selecting each dose are very close. For doses 1–5, the selection percentages out of 1000 simulations are (60.1, 21.7, 4.4, 0.9, 0.4) based on 2d-OR and (53.6, 26.8, 7.1, 0.9, 0.1) based on 3d-OR when $\gamma = 2$; and are (60.0, 20.1, 6.4, 0.7, 0.3) based on 2d-OR and (52.1, 24.4, 7.5, 1.3, 0.1) based on 3d-OR when $\gamma = -2$. These results demonstrate the robustness of our proposed designs regardless of the association parameters in the data generation.

4. Trial Conduct

As an illustration, we applied the proposed method to the phase I/II clinical trial design for the dose-finding of RAD001 in combination with docetaxel. RAD001 is taken as a single, once-weekly oral dose of 30 mg, 50 mg, or 70 mg according to the treatment schedule, while docetaxel has a constant dose level of 75 mg/m^2. After docetaxel treatment is stopped, RAD001 alone may continue to be given as a once-weekly dose until progression of the disease or appearance of unacceptable toxicity. Dose escalation or deescalation within a cohort is not allowed. The target is to find the highest tolerated dose among the three dose levels of RAD001, which will be combined with docetaxel. The efficacy of the combination of the two treatments in patients with metastatic breast cancer is assessed by a tumor response characterized as a complete response (disappearance of all target lesions) or a partial response (at least a 30% decrease in the sum of the longest diameter of target lesions with respect to the baseline). Of scientific interest for further investigation will be a finding of a response rate of at least 30%. Toxicity is continuously monitored and is defined as grade 4 hematological toxicity, grade 3–4 nonhematological toxicity, or other well-defined serious adverse events. The target toxicity of the maximum tolerated dose is assumed to be 25%. Therefore, we took $\bar{\pi}_T = 0.25$, $\underline{\pi}_E = 0.3$ and $p^* = 0.2$, $q^* = 0.1$ as threshold values for acceptable doses, and $p^{\text{escl}} = 0.55$ in (6).

Patients will be recruited by a cohort size of three, starting at (RAD001 30 mg + docetaxel 75 mg/m^2). A minimum of 3 and a maximum of 36 patients will be accrued to the study. Cumulated bivariate outcomes will be evaluated during the trial and decisions will be sequentially made on which dose level the next cohort of patients will be treated. Table A2 (available at `http://www.tibs.org/biometrics`) summarizes the simulation results for four scenarios with respect to the prespecified toxicity–efficacy probabilities of the three doses. To implement the TC design, we specified the prior mean probabilities of toxicity and efficacy for the three doses to be (0.1, 0.2, 0.3) and (0.05, 0.2, 0.4), respectively. In Scenario 1 where the second dose is the best, the $\pi_{01}^{(j)}$ criterion has a slightly lower selection percentage of dose 2 while the TC design seems to have a relatively higher likelihood of treating patients at dose 3. In Scenario 2, all four methods behave similarly except that the TC design treats a few more patients at dose 2. In Scenario 3, the TC and $\pi_{01}^{(j)}$ designs yield better results with higher percentages of selecting the last dose. When all the doses are excessively toxic as in Scenario 4, the four methods are able to stop the trial early.

The designs based on the $\omega_j^{(2)}$ and $\omega_j^{(3)}$ criteria are generally conservative and lead to similar results. The 3d-OR criterion accounts for the correlation between toxicity and efficacy and intuitively puts more weight on efficacy relative to the 2d-OR criterion. The operating characteristics under these four scenarios are quite satisfactory and reasonable. Based on various configurations of the underlying toxicity–efficacy probabilities in the simulations, the designs using the 2d-OR and 3d-OR criteria seem to be robust and reliable, and may be recommended for general practical use.

5. Discussion

We have proposed a new dose-finding method by combining phase I and II clinical trials. The odds ratio trade-off contours between toxicity and efficacy are intuitively attractive and practically feasible. It has an objective and meaningful interpretation of quantifying the relative degree of toxicity versus efficacy. In the proposed design, we constrained the parameters associated with toxicity probabilities in a monotonic order, while leaving efficacy probabilities free of constraints. The prior specification can be easily adapted to allow either the probabilities of toxicity or efficacy to be ordered with respect to the dose levels, depending on practical situations.

The new design inherits the curve-free advantage, which is nonparametric and thus does not depend on the dose settings. The conventional methods usually assume a proportional odds or a continual ratio model which might be unstable under certain dosage configurations. For example, if the dose levels double or triple in a consecutive order, the parametric model could result in undesirable operating characteristics, or even become computationally intractable in some cases. As an alternative to the Dale model, one may consider modeling the correlation between toxicity and efficacy at different dose levels using a single association parameter as in the Gumbel model (Murtaugh and Fisher, 1990). The feasibility of the proposed design requires a full evaluation of toxicity and efficacy responses for each cohort before assigning patients in the next cohort to a certain dose level. Therefore, the method may not be directly applicable to a trial in which there is a long period of lagged time to observe the bivariate outcomes.

Acknowledgements

We would like to thank the associate editor and three referees for very constructive suggestions which have led to substantial improvements of the article. We also thank Peter Thall and John Cook for very helpful and stimulating discussions, and for providing the software for dose-finding based on toxicity–efficacy trade-offs.

References

Bekele, B. N. and Shen, Y. (2005). A Bayesian approach to jointly modeling toxicity and biomarker expression in a phase I/II dose-finding trial. *Biometrics* **61,** 343–354.

Braun, T. M. (2002). The bivariate continual reassessment method: Extending the CRM to phase I trials of two competing outcomes. *Controlled Clinical Trials* **23,** 240–256.

Cheung, Y. K. (2002). On the use of nonparametric curves in phase I trials with low toxicity tolerance. *Biometrics* **58,** 237–240.

Dale, J. R. (1986). Global cross-ratio models for bivariate, discrete, ordered responses. *Biometrics* **42,** 909–917.

Gasparini, M. and Eisele, J. (2000). A curve-free method for phase I clinical trials. *Biometrics* **56,** 609–615.

Gilks, W. R., Best, N. G., and Tan, K. K. C. (1995). Adaptive rejection Metropolis sampling within Gibbs sampling. *Applied Statistics* **44,** 455–472.

Goodman, S. N., Zahurak, M. L., and Piantadosi, S. (1995). Some practical improvements in the continual reassessment method for phase I studies. *Statistics in Medicine* **14,** 1149–1161.

Gooley, T. A., Martin, P. J., Fisher, L. D., and Pettinger, M. (1994). Simulation as a design tool for phase I/II clinical trials: An example from bone marrow transplantation. *Controlled Clinical Trials* **15,** 450–462.

Ivanova, A. (2003). A new dose-finding design for bivariate outcomes. *Biometrics* **59,** 1001–1007.

Korn, E. L., Midthune, D., Chen, T. T., Rubinstein, L. V., Christian, M. C., and Simon, R. M. (1994). A comparison of two phase I trial designs. *Statistics in Medicine* **13,** 1799–1806.

Leung, D. H.-Y. and Wang, Y.-G. (2002). An extension of the continual reassessment method using decision theory. *Statistics in Medicine* **21,** 51–63.

Møller, S. (1995). An extension of the continual reassessment methods using a preliminary up-and-down design in a dose finding study in cancer patients, in order to investigate a greater range of doses. *Statistics in Medicine* **14,** 911–922.

Mukhopadhyay, S. (2000). Bayesian nonparametric inference on the dose level with specified response rate. *Biometrics* **56,** 220–226.

Murtaugh, P. A. and Fisher, L. D. (1990). Bivariate binary models of efficacy and toxicity in dose-ranging trials. *Communications in Statistics, Part A—Theory and Methods* **19,** 2003–2020.

O'Quigley, J. and Shen, L. Z. (1996). Continual reassessment method: A likelihood approach. *Biometrics* **52,** 673–684.

O'Quigley, J., Pepe, M., and Fisher, L. (1990). Continual reassessment method: A practical design for phase I clinical trials in cancer. *Biometrics* **46,** 33–48.

O'Quigley, J., Hughes, M. D., and Fenton, T. (2001). Dose-finding designs for HIV studies. *Biometrics* **57,** 1018–1029.

Storer, B. E. (1989). Design and analysis of phase I clinical trials. *Biometrics* **45,** 925–937.

Thall, P. F. and Cheng, S.-C. (1999). Treatment comparisons based on two–dimensional safety and efficacy alternatives in oncology trials. *Biometrics* **55,** 746–753.

Thall, P. F. and Cook, J. (2004). Dose-finding based on toxicity–efficacy trade-offs. *Biometrics* **60,** 684–693.

Thall, P. F. and Russell, K. E. (1998). A strategy for dose-finding and safety monitoring based on efficacy and adverse outcomes in phase I/II clinical trials. *Biometrics* **54,** 251–264.

Received November 2004. *Revised November* 2005. *Accepted November* 2005.

Table A1
Simulation results with (p_j, q_j) under the Dale model for five dose levels

	Selection percentage of dose (# of patients)					
	1	2	3	4	5	None
Scenario 1	(15, 55)	(25, 58)	(35, 60)	(45, 62)	(55, 65)	
δ_j	0.2926	0.2320	0.1313	0.0101	−0.1164	
	51.7 (28.5)	34.9 (20.3)	6.7 (5.5)	1.0 (1.2)	0 (0.2)	5.7
$\omega_j^{(2)}$	0.1444	0.2414	0.3590	0.5015	0.6581	
	57.4 (30.1)	22.1 (13.5)	6.4 (6.0)	1.4 (2.6)	0.1 (1.1)	12.6
$\omega_j^{(3)}$	0.1181	0.1748	0.2393	0.3074	0.3544	
	54.1 (28.5)	25.4 (15.0)	6.4 (6.2)	1.2 (2.7)	0.2 (1.0)	12.7
$\pi_{01}^{(j)}$	0.4675	0.4350	0.3900	0.3410	0.2925	
	53.7 (28.7)	23.4 (14.4)	8.2 (7.0)	2.4 (2.8)	0.3 (1.0)	12.0

Table A1
(*Continued*)

	Selection percentage of dose (# of patients)					
	1	2	3	4	5	None
Scenario 2	(1, 52)	(1.5, 61)	(2, 71)	(2.5, 82)	(3, 90)	
δ_j	0.3464	0.4653	0.5975	0.7427	0.8455	
	22.1 (16.4)	4.2 (8.9)	5.3 (7.0)	29.6 (17.3)	38.8 (10.5)	0
$\omega_j^{(2)}$	0.0093	0.0097	0.0083	0.0056	0.0034	
	21.8 (11.5)	11.5 (8.4)	11.5 (8.8)	17.4 (11.9)	37.5 (19.2)	0.3
$\omega_j^{(3)}$	0.0086	0.0062	0.0034	0.0012	0.0004	
	5.8 (6.1)	7.3 (6.8)	12.1 (9.1)	25.1 (14.5)	49.3 (23.2)	0.4
$\pi_{01}^{(j)}$	0.5148	0.6008	0.6958	0.7995	0.8730	
	1.2 (4.4)	1.9 (5.5)	9.5 (8.6)	29.6 (16.8)	57.8 (24.8)	0
Scenario 3	(15, 15)	(25, 25)	(35, 35)	(45, 45)	(55, 55)	
δ_j	−0.2358	−0.1789	−0.1484	−0.1536	−0.1973	
	2.7 (7.7)	18.1 (10.6)	17.6 (10.5)	4.9 (5.7)	0.1 (1.4)	56.6
$\omega_j^{(2)}$	1	1	1	1	1	
	3.1 (8.6)	12.8 (9.7)	11.1 (7.5)	2.1 (2.7)	0.2 (1.0)	70.7
$\omega_j^{(3)}$	5.6667	3.0000	1.8571	1.2222	0.8182	
	3.3 (8.3)	14.9 (10.5)	8.8 (6.9)	2.0 (2.9)	0.1 (1.0)	70.9
$\pi_{01}^{(j)}$	0.1275	0.1875	0.2275	0.2475	0.2475	
	4.2 (8.4)	12.9 (9.6)	10.9 (7.8)	3.9 (3.9)	0.1 (1.3)	68.0
Scenario 4	(1, 5)	(2, 20)	(3, 35)	(4, 60)	(5, 80)	
δ_j	−0.2886	−0.0911	0.1061	0.4377	0.6988	
	0 (3.4)	0 (3.0)	0.2 (3.4)	4.6 (9.4)	95.2 (40.8)	0
$\omega_j^{(2)}$	0.1919	0.0816	0.0574	0.0278	0.0132	
	0 (3.1)	1.3 (4.2)	6.3 (6.6)	28.0 (16.0)	61.5 (28.5)	2.9
$\omega_j^{(3)}$	3.6465	0.3265	0.1067	0.0185	0.0033	
	0 (3.1)	0.5 (3.7)	5.3 (5.6)	26.4 (14.8)	65.1 (31.3)	2.7
$\pi_{01}^{(j)}$	0.0495	0.1960	0.3395	0.5760	0.7600	
	0 (3.1)	0.3 (3.3)	1.6 (4.4)	17.1 (12.5)	79.1 (35.7)	1.9
Scenario 5	(30, 5)	(40, 20)	(50, 35)	(60, 50)	(70, 60)	
δ_j	−0.4795	−0.3815	−0.3229	−0.3140	−0.3873	
	0.5 (3.6)	6.0 (5.1)	0.4 (3.2)	0.1 (1.2)	0.1 (0.3)	92.9
$\omega_j^{(2)}$	8.1429	2.6667	1.8571	1.5000	1.5556	
	0 (3.8)	0.4 (2.6)	0 (1.4)	0 (0.3)	0 (0.1)	99.6
$\omega_j^{(3)}$	154.7143	10.6667	3.4490	1.5000	1.0370	
	0.3 (3.8)	1.0 (3.2)	0.1 (1.7)	0 (0.4)	0 (0.0)	98.6
$\pi_{01}^{(j)}$	0.0350	0.1200	0.1750	0.2000	0.1800	
	0 (4.0)	0.8 (3.3)	0.1 (1.6)	0 (0.4)	0 (0.1)	99.1
Scenario 6	(1, 20)	(3, 40)	(4, 60)	(5, 70)	(35, 75)	
δ_j	−0.0859	0.1736	0.4377	0.5658	0.2683	
	1.6 (5.7)	0.7 (3.8)	2.2 (5.8)	82.8 (36.0)	12.6 (8.7)	0.1
$\omega_j^{(2)}$	0.0404	0.0464	0.0278	0.0226	0.1795	
	4.8 (6.4)	15.2 (10.4)	35.1 (17.2)	38.5 (18.2)	4.3 (6.6)	2.1
$\omega_j^{(3)}$	0.1616	0.0696	0.0185	0.0097	0.0598	
	2.0 (4.4)	11.6 (8.3)	32.3 (16.4)	42.9 (20.8)	9.0 (8.8)	2.2
$\pi_{01}^{(j)}$	0.1980	0.3880	0.5760	0.6650	0.4875	
	0.5 (3.6)	5.6 (6.4)	30.0 (16.8)	53.4 (23.3)	9.1 (9.2)	1.4
Scenario 7	(2, 60)	(4, 62)	(6, 64)	(25, 66)	(35, 68)	
δ_j	0.4491	0.4646	0.4790	0.3224	0.2083	
	41.0 (24.5)	16.8 (14.0)	29.9 (14.0)	11.9 (7.2)	0.4 (0.4)	0
$\omega_j^{(2)}$	0.0136	0.0255	0.0359	0.1717	0.2534	
	60.5 (28.8)	24.6 (14.2)	11.1 (8.3)	2.6 (5.0)	1.0 (3.5)	0.2
$\omega_j^{(3)}$	0.0091	0.0156	0.0202	0.0885	0.1192	
	45.1 (21.9)	25.8 (14.3)	20.5 (12.5)	6.4 (6.8)	1.6 (4.2)	0.6
$\pi_{01}^{(j)}$	0.5880	0.5952	0.6016	0.4950	0.4420	
	31.7 (18.0)	28.1 (15.1)	26.8 (14.4)	9.5 (7.7)	3.5 (4.6)	0.4

Table A1
(*Continued*)

	Selection percentage of dose (# of patients)					
	1	2	3	4	5	None
Scenario 8	(8, 15)	(10, 35)	(12, 52)	(45, 65)	(55, 70)	
δ_j	−0.1919	0.0647	0.2772	0.0358	−0.0799	
	0.6 (5.2)	5.8 (5.9)	72.4 (30.7)	17.6 (15.4)	0.1 (0.9)	3.5
$\omega_j^{(2)}$	0.4928	0.2063	0.1259	0.4406	0.5238	
	2.8 (6.7)	24.9 (13.4)	47.8 (21.8)	4.5 (6.0)	0.3 (2.0)	19.7
$\omega_j^{(3)}$	2.7923	0.3832	0.1162	0.2372	0.2245	
	1.8 (5.5)	25.5 (13.6)	50.4 (22.9)	3.6 (6.6)	0.4 (2.0)	18.3
$\pi_{01}^{(j)}$	0.1380	0.3150	0.4576	0.3575	0.3150	
	1.3 (4.9)	19.8 (11.5)	56.0 (24.9)	5.0 (7.1)	0.6 (2.4)	17.3
Scenario 9	(1, 1)	(2, 2)	(3, 3)	(4, 4)	(5, 5)	
δ_j	−0.3426	−0.3343	−0.3260	−0.3178	−0.3098	
	1.5 (3.2)	0 (3.0)	0 (3.0)	0 (3.0)	1.7 (4.9)	96.8
$\omega_j^{(2)}$	1	1	1	1	1	
	0 (3.1)	0 (3.2)	0 (3.4)	0 (3.6)	0 (4.0)	100
$\omega_j^{(3)}$	99.0000	49.0000	32.3333	24.0000	19.0000	
	0 (3.2)	0 (3.3)	0 (3.5)	0 (3.8)	0 (4.0)	100
$\pi_{01}^{(j)}$	0.0099	0.0196	0.0291	0.0384	0.0475	
	0 (3.2)	0 (3.3)	0 (3.5)	0 (3.9)	0 (4.2)	100
Scenario 10	(5, 40)	(15, 60)	(40, 50)	(60, 40)	(80, 30)	
δ_j	0.1624	0.3569	−0.0371	−0.4051	−0.7813	
	22.7 (15.7)	68.9 (34.0)	4.5 (7.6)	0.1 (0.8)	0 (0)	3.8
$\omega_j^{(2)}$	0.0789	0.1176	0.6667	2.2500	9.3333	
	43.6 (24.8)	48.6 (25.0)	1.3 (4.6)	0 (1.8)	0 (0.5)	6.5
$\omega_j^{(3)}$	0.1184	0.0784	0.6667	3.3750	21.7778	
	38.1 (21.6)	52.6 (27.2)	2.5 (5.4)	0 (1.9)	0 (0.5)	6.8
$\pi_{01}^{(j)}$	0.3800	0.5100	0.3000	0.16	0.0600	
	28.3 (18.2)	62.3 (30.2)	1.9 (5.3)	0.3 (2.0)	0 (0.5)	7.2
Scenario 11	(1, 10)	(5, 30)	(10, 50)	(40, 80)	(50, 60)	
δ_j	−0.2210	0.0276	0.2651	0.2271	−0.0807	
	0.1 (3.7)	0.6 (3.7)	65.2 (28.3)	33.0 (23.3)	0 (0.4)	1.1
$\omega_j^{(2)}$	0.0909	0.1228	0.1111	0.1667	0.6667	
	1.2 (4.8)	28.1 (15.6)	47.4 (21.7)	14.2 (12.0)	0.1 (2.4)	9.0
$\omega_j^{(3)}$	0.8182	0.2865	0.1111	0.0417	0.4444	
	1.0 (4.2)	22.5 (12.4)	52.9 (23.5)	14.0 (13.5)	0.1 (2.7)	9.5
$\pi_{01}^{(j)}$	0.099	0.2850	0.4500	0.4800	0.3000	
	0.5 (4.1)	16.5 (10.9)	54.3 (23.8)	17.9 (14.6)	0.4 (2.6)	10.4
Scenario 12	(1.3, 58.9)	(1.5, 66.0)	(1.7, 72.4)	(2, 78.0)	(3.5, 92.2)	
δ_j	0.4380	0.5328	0.6181	0.6919	0.8693	
	25.9 (17.7)	8.1 (11.3)	6.7 (8.4)	24.7 (13.6)	34.6 (9.0)	0
$\omega_j^{(2)}$	0.0092	0.0078	0.0066	0.0058	0.0031	
	23.8 (12.8)	11.4 (8.9)	9.3 (7.9)	8.3 (8.2)	46.8 (22.0)	0.4
$\omega_j^{(3)}$	0.0064	0.0040	0.0025	0.0016	0.0003	
	10.4 (7.9)	9.6 (8.3)	9.7 (8.5)	15.2 (10.4)	54.3 (24.4)	0.8
$\pi_{01}^{(j)}$	0.5813	0.6501	0.7117	0.7644	0.8897	
	2.3 (5.2)	4.6 (6.6)	8.1 (8.2)	14.5 (10.7)	70.4 (29.2)	0.1
Scenario 13	(2.6, 9.8)	(4.1, 29.7)	(6.4, 53.1)	(9.9, 67.9)	(41.8, 40.2)	
δ_j	−0.2321	0.0286	0.3303	0.5020	−0.1654	
	0.1 (4.0)	0.7 (3.4)	6.3 (7.0)	88.3 (39.8)	2.8 (5.0)	1.8
$\omega_j^{(2)}$	0.2457	0.1012	0.0604	0.0519	1.0684	
	0.4 (4.0)	11.3 (9.0)	35.6 (18.1)	47.0 (22.7)	0.3 (3.2)	5.4
$\omega_j^{(3)}$	2.2614	0.2395	0.0533	0.0246	1.5893	
	0.1 (3.6)	7.5 (6.9)	33.2 (17.3)	52.8 (25.5)	0.4 (3.5)	6.0
$\pi_{01}^{(j)}$	0.0954	0.2848	0.4970	0.6118	0.2340	
	0.1 (3.3)	4.4 (6.2)	29.0 (16.9)	61.3 (27.6)	0.4 (3.3)	4.8

Table A2
Simulation results for the breast cancer trial with RAD001 and docetaxel

	Selection percentage of dose (# of patients)			
	1	2	3	None
Scenario 1	(4, 15)	(8, 60)	(40, 75)	
δ_j	−0.1693	0.4120	0.1915	
	0.1 (3.1)	74.5 (19.2)	19.9 (12.1)	5.5
$\omega_j^{(2)}$	0.2361	0.0580	0.2222	
	4.2 (5.5)	72.0 (19.8)	10.2 (6.9)	13.6
$\omega_j^{(3)}$	1.3380	0.0386	0.0741	
	3.0 (4.8)	74.7 (20.7)	9.5 (7.2)	12.8
$\pi_{01}^{(j)}$	0.1440	0.5520	0.4500	
	9.4 (8.2)	63.7 (16.8)	12.6 (6.9)	14.3
Scenario 2	(10, 40)	(40, 50)	(60, 80)	
δ_j	0.1316	−0.0371	−0.0968	
	53.8 (15.7)	14.6 (10.9)	0.2 (1.5)	31.4
$\omega_j^{(2)}$	0.1667	0.6667	0.3750	
	64.6 (21.0)	5.8 (5.6)	1.5 (2.2)	28.1
$\omega_j^{(3)}$	0.2500	0.6667	0.0938	
	66.3 (20.8)	6.0 (6.3)	1.0 (2.1)	26.7
$\pi_{01}^{(j)}$	0.3600	0.3000	0.3200	
	60.9 (18.9)	9.2 (7.7)	0.4 (2.0)	29.5
Scenario 3	(1, 20)	(2, 50)	(3, 80)	
δ_j	−0.0859	0.3141	0.7127	
	0 (3.0)	0.3 (3.8)	99.6 (29.2)	0.1
$\omega_j^{(2)}$	0.0404	0.0204	0.0077	
	3.6 (4.8)	24.8 (10.1)	69.6 (20.4)	2.0
$\omega_j^{(3)}$	0.1616	0.0204	0.0019	
	2.6 (4.1)	22.1 (9.3)	73.6 (22.1)	1.7
$\pi_{01}^{(j)}$	0.1980	0.4900	0.7760	
	0 (3.0)	0 (3.1)	97.9 (29.1)	2.1
Scenario 4	(35, 10)	(45, 40)	(55, 70)	
δ_j	−0.4582	−0.2067	−0.0799	
	0.9 (3.6)	5.0 (4.7)	0.3 (1.9)	93.8
$\omega_j^{(2)}$	4.8462	1.2273	0.5238	
	0.9 (4.6)	0.8 (2.2)	0.2 (0.5)	98.1
$\omega_j^{(3)}$	43.6154	1.8409	0.2245	
	1.1 (4.4)	0.9 (2.3)	0 (0.6)	98.0
$\pi_{01}^{(j)}$	0.0650	0.2200	0.3150	
	0.6 (4.5)	1.2 (2.6)	0.1 (0.7)	98.1