Biostatistics (2016), **17**, 2, *pp.* 304–319
doi:10.1093/biostatistics/kxv045
Advance Access publication on November 9, 2015

A decision-theoretic phase I–II design for ordinal outcomes in two cycles

JUHEE LEE*

Department of Applied Mathematics and Statistics, Baskin School of Engineering, University of California, 1156 High Street, Mail Stop SOE2, Santa Cruz, CA 95064, USA

juheelee@soe.ucsc.edu

PETER F. THALL

Department of Biostatistics, University of Texas M.D. Anderson Cancer Center, Houston, TX, USA

YUAN JI

Program of Computational Genomics & Medicine, NorthShore University Health System, Evanston, IL, USA and Department of Public Health Sciences, The University of Chicago, Chicago, IL, USA

PETER MÜLLER

Department of Mathematics, University of Texas, Austin, TX, USA

Summary

This paper is motivated by a phase I–II clinical trial of a targeted agent for advanced solid tumors. We study a stylized version of this trial with the goal to determine optimal actions in each of two cycles of therapy. A design is presented that generalizes the decision-theoretic two-cycle design of Lee *and others* (2015. Bayesian dose-finding in two treatment cycles based on the joint utility of efficacy and toxicity. *Journal of the American Statistical Association*, to appear) to accommodate ordinal outcomes. Backward induction is used to jointly optimize the actions taken for each patient in each of the two cycles, with the second action accounting for the patient's cycle 1 dose and outcomes. A simulation study shows that simpler designs obtained by dichotomizing the ordinal outcomes either perform very similarly to the proposed design, or have much worse performance in some scenarios. We also compare the proposed design with the simpler approaches of optimizing the doses in each cycle separately, or ignoring the distinction between cycles 1 and 2.

Keywords: Adaptive design; Bayesian design; Decision theory; Dynamic treatment regime; Latent probit model; Ordinal outcomes; Phase I–II clinical trial.

1. Introduction and motivation

This paper is motivated by the problem of designing a dose-finding trial of a new agent for cancer patients with advanced solid tumors. The agent aims to inhibit a kinase, which regulates cell metabolism and

*To whom correspondence should be addressed.

proliferation, in the cancer cells to reduce or eradicate the disease. The agent is given orally each day of a 28-day cycle at one of five doses, 2, 4, 6, 8, or 10 mg, combined with a fixed dose of standard chemotherapy. Because both efficacy and toxicity are used for dose-finding, it is a phase I–II trial (Thall and Cook, 2004; Yin *and others*, 2006; Zhang *and others*, 2006; Thall and Nguyen, 2012). Both outcomes are 3-level ordinal variables, with toxicity defined as None/Mild (grade 0,1), Moderate (grade 2), or Severe (grade 3,4) and efficacy defined in terms of disease status compared with baseline, with possible values progressive disease (PD), stable disease (SD), or partial or complete response (PR/CR).

We study a stylized version of this trial with the more ambitious goal to determine optimal doses or actions for each patient in each of two cycles of therapy. This is a major departure from conventional dose-finding designs, which focus on choosing a dose for only the first cycle. While virtually all clinical protocols for dose-finding trials include rules for making within-patient dose adjustments in cycles after the first, this aspect usually is ignored in the trial design. In practice, each patient's doses in cycle 2, or later cycles, are chosen subjectively by the attending physician. To choose a patient's cycle 2 dose using a formal rule, it is desirable to use the patient's dose-outcome data from cycle 1, as well as data from other patients treated previously in the trial. Thus, ideally, a decision rule that is adaptive both within and between patients is needed.

Recent papers on designs accounting for multiple treatment cycles include Cheung *and others* (2014) and Lee *and others* (2015). In this paper, we build on the latter, who use a decision-theoretic approach for dose-finding in two cycles based on joint utilities of binary outcomes in each cycle. We extend the model to accommodate ordinal outcomes, and use a decision criterion that accounts for the many possible (efficacy,toxicity) outcomes in each of the two treatment cycles, including the risk-benefit trade-offs between the levels of efficacy and toxicity. In the stylized version of the trial described above, since there are 3-level ordinal toxicity and efficacy outcomes in each cycle, accounting for two cycles there are 81 possible elementary outcomes for each patient. Consequently, dose-finding is a much more complex problem than in a conventional phase I–II trial with two binary outcomes that chooses a dose for cycle 1 only.

Aside from the issue of accounting for two cycles, an important question is whether the additional complexity required to account for ordinal outcomes provides practical benefits compared with the common approach of dichotomizing efficacy and toxicity, which would allow the two-cycle design of Lee *and others* (2015) to be applied. Simulations, described in Section 4.4 of the main text, Figure 2, and Section 3 of Supplementary Material (available at *Biostatistics* online), show that reducing ordinal outcomes to binary variables produces a design that either performs very similarly to the proposed design, or has much worse performance in certain scenarios. Moreover, the behavior of the simplified design depends heavily on how one chooses to reduce the two ordinal outcomes to two binary variables.

A naive design might aim to optimize the doses given in the two cycles separately. This may be not optimal. To see this, denote a patient's toxicity outcome by Y_c and efficacy outcome by Z_c for $c = 1, 2$, and denote the current data from n patients by $\mathcal{X}_n$. We include NT = "Donottreat" as a possible action in either cycle for cases where it has been determined that no dose is acceptable, so the action d_c in each of cycles $c = 1, 2$ may be either to choose a dose or NT, that is, $d_c \in \mathcal{D} = \{\text{NT}, 1, \ldots, m\}$ with 1 and m denoting the minimum dose and the maximum dose levels, respectively. Suppose that some optimality criterion has been defined. If one derives optimal adaptive actions $d_1^\star$ for cycle 1 and $d_2^\star$ for cycle 2 separately, each based on the current data $\mathcal{X}_n$, an inherent flaw is that in choosing one $d_2^\star$ for all patients it ignores each patient's cycle 1 data. As in Lee *and others* (2015), we derive optimal decision rules $\boldsymbol{d}^\star = (d_1^\star, d_2^\star)$ with the important property that $d_2^\star = d_2^\star(d_1, Y_1, Z_1, \mathcal{X}_n)$ is a function of the first cycle decision d_1 and response Y_1, Z_1. This is implemented by applying backward induction (Bellman, 1957, etc.). The method accounts for the patient's cycle 1 dose and outcomes, as well as other patient's data, in making an optimal decision for cycle 2.

Iasonos *and others* (2011) and Van Meter *and others* (2012) studied the use of ordinal toxicity outcomes for a generalized continual reassessment method and reported that gains in performance of their

ordinal toxicity designs are not substantial in comparison to binary toxicity designs. However, the comparison looks quite different for the model-based two-cycle design for bivariate ordinal (efficacy, toxicity) outcomes that we propose in this paper. In simulations described in Section 4.4, we compare the proposed design with designs that do not properly model association between cycles. In simulations reported in Section 4.5, we show that the use of ordinal rather than binary outcomes can substantially improve design performance in our setting.

Section 2 describes the proposed decision-theoretic method for ordinal outcomes in two cycles (DTD-O2). Sections 3 and 4 include decision criteria using utilities and a simulation study. The last section concludes with a final discussion.

2. A DECISION-THEORETIC DESIGN

2.1 *Actions and optimal sequential decisions*

For notational convenience, we denote the possible levels of toxicity by $0, 1, \ldots, J-1$ and efficacy by $0, 1, \ldots, K-1$. For the motivating trial, these are $Y = 0$ for None/Mild, 1 for Moderate, and 2 for Severe, and $Z = 0$ for PD, 1 for SD, and 2 for CR/PR, so $K = J = 3$. If the adaptively chosen cycle 1 action $d_1 = \text{NT}$ for any patient, then the trial is stopped and no more patients are enrolled. Otherwise, the patient receives a dose d_1 of the agent in cycle 1. A cycle 2 action is a function mapping the cycle 1 dose and outcomes, (d_1, Y_1, Z_1), to an action in $\mathcal{D}$. For example, if the cycle 1 action d_1 produced None/Mild toxicity ($Y_1 = 0$), one possible cycle 2 action is $d_2(d_1, 0, Z_1, \mathcal{X}) = d_1 + 1$ if $Z_1 = 0$, and d_1 if $Z_1 = 1$ or 2. That is, if there was little or no toxicity but PD in cycle 1, then the action d_2 increases the dose in cycle 2, but if the patient had SD or better then it repeats the cycle 1 dose. The design thus involves an alternating sequence of decisions and observed outcomes, d_1, (Y_1, Z_1), $d_2(d_1, Y_1, Z_1)$, and (Y_2, Z_2).

We apply a Bayesian decision-theoretic paradigm to determine an optimal decision rule. First, focus on cycle 1, and temporarily ignore cycle 2. The general setup of a Bayesian decision problem involves actions d_1, observable data $\boldsymbol{y} = (Y_1, Z_1)$, parameters θ that index a sampling model $p(\boldsymbol{y} \mid \theta, d_1)$ for the data, and a prior probability model $p(\theta)$ for the parameters. We discuss specification of $\mathcal{D}$ in more detail below. A utility function $u(d_1, \theta, \boldsymbol{y})$ formalizes relative preferences for alternative actions under hypothetical outcomes $\boldsymbol{y}$ and assumed truth θ. Starting from first principles, one can then argue (Robert, 2007, Chapter 2) that a rational decision-maker chooses the action $d_1^\star$ that maximizes utility in expectation, that is

$$d_1^\star = \arg\max_{d_1} \int u(d_1, \theta, \boldsymbol{y})\, \mathrm{d}p(\boldsymbol{y}, \theta \mid d_1, \mathcal{X}_n) = \arg\max_{d_1} U_1(d_1). \tag{2.1}$$

The integral is the expected utility $U_1(d_1) = E_{y,\theta}\{u(d_1, \theta, \boldsymbol{y})\}$, with the expectation taken with respect to $p(\boldsymbol{y}, \theta \mid d_1, \mathcal{X}_n) = p(\theta \mid \mathcal{X}_n) p(\boldsymbol{y} \mid \theta, d_1)$. To simplify notation, we will henceforth suppress conditioning on $\mathcal{X}_n$ in the notation.

In the two-cycle dose-finding problem, the sequential nature of the within-patient decisions complicates the solution. In the second cycle, the utility $u(d_1, \theta, \boldsymbol{y})$ is replaced by the expected utility under optimal continuation. Denote $\boldsymbol{y}_1 = (Y_1, Z_1)$ and $\boldsymbol{y}_2 = (Y_2, Z_2)$. We get an alternating sequence of optimization and expectation

$$\begin{aligned} d_1^\star &= \arg\max_{d_1} \int \left\{ \max_{d_2} \int u(d_1, d_2, \theta, \boldsymbol{y}_1, \boldsymbol{y}_2)\, \mathrm{d}p(\boldsymbol{y}_2, \theta \mid d_1, d_2, \boldsymbol{y}_1) \right\} \mathrm{d}p(\boldsymbol{y}_1 \mid d_1) \\ &= \arg\max_{d_1} \int U_2(d_1, d_2 = d_2^\star(d_1, \boldsymbol{y}_1), \boldsymbol{y}_1)\, \mathrm{d}p(\boldsymbol{y}_1 \mid d_1) \end{aligned} \tag{2.2}$$

Table 1. *An example of elicited utilities*, $u_{\text{cycle}}(\boldsymbol{y})$

	Toxicity severity level		
Efficacy scores	Mild	Moderate	Severe
PD	25	10	0
SD	70	50	25
PR/CR	100	80	50

with the second cycle expected total utility as a function of $\boldsymbol{y}_1 = (Y_1, Z_1)$, $U_2(d_1, d_2, \boldsymbol{y}_1) = E_{y_2,\theta} u(d_1, d_2, \theta, \boldsymbol{y}_1, \boldsymbol{y}_2)$ and the optimal second cycle decision $d_2^\star(d_1, \boldsymbol{y}_1) = \arg\max_{d_2} U_2(d_1, d_2, \boldsymbol{y}_1)$. When we substitute $d_2^\star(d_1, \boldsymbol{y}_1)$ and take the expectation with respect to $\boldsymbol{y}_1$ we obtain

$$U_1(d_1) = E_{\boldsymbol{y}_1}\{U_2(d_1, d_2^\star(d_1, \boldsymbol{y}_1), \boldsymbol{y}_1)\}, \tag{2.3}$$

which is maximized to determine the optimal decision for cycle 1, $d_1^\star = \arg\max_{d_1} U_1(d_1)$. This alternating sequence of maximization and expectation, called dynamic programming, is characteristic of sequential decision problems. While it often leads to intractable computational problems (Parmigiani and Inoue, 2009, Chapter 12), in the present setting with ordinal outcomes the problem is solvable. Dynamic programming recently has been applied in other clinical trial design settings (Murphy, 2003; Zhao *and others*, 2011; Lee *and others*, 2015; Cheung *and others*, 2014).

2.2 *Utility function*

We construct a utility function

$$u(d_1, d_2, \boldsymbol{y}_1, \boldsymbol{y}_2, \theta) = \sum_{c=1,2} \lambda^{c-1} u_{\text{cycle}}(Y_c, Z_c) \tag{2.4}$$

as a sum over cycle-specific utilities $u_{\text{cycle}}(Y_c, Z_c)$, $c = 1, 2$, where $0 \leqslant \lambda \leqslant 1$ is a scale parameter. If $\lambda = 0$, then the cycle 2 utility is ignored in selecting d_1, while $\lambda = 1$ corresponds to treating utilities in the two cycles equally. Optimal decisions may change under different values of λ. Even with $\lambda = 0$, however, the importance of jointly modeling the two cycles remains in that inference on θ can be enhanced through borrowing information across cycles. For the simulations in Section 4, we used $\lambda = 0.8$. A sensitivity analysis in λ is reported in the Supplementary Materials (available at *Biostatistics* online). The utility function (2.4) focuses on the clinical outcomes and is a function of $(\boldsymbol{y}_1, \boldsymbol{y}_2) = (Y_1, Z_1, Y_2, Z_2)$ only. That is, the inference on θ does not affect utility, and we do not initially consider preferences across doses d_c. We thus drop θ and d_c from the arguments of $u_{\text{cycle}}(\cdot)$ hereafter.

In practice, numerical utilities of the $J \times K$ elementary must be elicited from the clinical collaborators, with specific numerical values reflecting physicians' relative preferences (cf. Thall and Nguyen, 2012). In our stylized illustrative trial, we fix the utilities of the best and worst possible outcomes to be $u_{\text{cycle}}(0, K-1) = 100$ and $u_{\text{cycle}}(J-1, 0) = 0$. In general, any convenient function with $u_{\text{cycle}}(j, k-1) < u_{\text{cycle}}(j, k)$ and $u_{\text{cycle}}(j, k) > u_{\text{cycle}}(j+1, k)$ that gives higher utilities to more desirable outcomes may be used. For future reference, we note that $u_{\text{cycle}}(0, 0)$ is the expected utility corresponding to NT, i.e. do not treat the patient. Table 1 shows the utilities that will be used for our simulation studies.

To reduce notation, we denote the utility $u(\boldsymbol{y}_1, \boldsymbol{y}_2)$ as a function of hypothetical outcomes $(\boldsymbol{y}_1, \boldsymbol{y}_2) = (Y_1, Z_1, Y_2, Z_2)$, and drop the arguments θ and d_c. Upper case $U_c(\cdot)$ denotes expected utility, with data $\boldsymbol{y}_{c'}$, $c' \geqslant c$ removed by marginalization and decisions $d_{c'}$, $c' > c$ substituted by maximization, as in (2.2).

In addition to the cycle index $_c$, the arguments of $U_c(\cdot)$ clarify the level of marginalization and maximization. Maximizing U_1 in (2.1) and U_2 inside the integral in (2.2) yields the optimal action pair $d^\star = (d_1^\star, d_2^\star)$, where $d_1^\star$ is either a dose or NT, $d_2^\star$ is applicable only when $d_1 \neq \text{NT}$ is a dose, and $d_2^\star(d_1, Y_1, Z_1)$ is a function of d_1 and the patient's cycle 1 outcomes, (Y_1, Z_1). Assuming that the utility function takes the additive form (2.4), we define cycle-specific expected utilities, with the expected utility for cycle 2 given by

$$\tilde{U}_2(d_2 \mid d_1, \boldsymbol{y}_1) = E_{\boldsymbol{y}_2,\theta}\{u_{\text{cycle}}(Y_2, Z_2) \mid d_1, \boldsymbol{y}_1\} = \int u_{\text{cycle}}(Y_2, Z_2)\, \mathrm{d}p(\boldsymbol{y}_2, \theta \mid d_1, d_2, \boldsymbol{y}_1). \tag{2.5}$$

Figure 1(a)–(c) illustrates $\tilde{U}_2(d_2 \mid d_1, y_1)$ under the assumed simulation truth of Scenario 3 (discussed in Section 4.2), and shows how $\tilde{U}_2(d_2 \mid d_1, \boldsymbol{y}_1)$ changes with $(d_2, \boldsymbol{y}_1)$, given $d_1 = 3$. Figure 1(d) illustrates the assumed true $U_1(d_1)$ over d_1 for the simulation scenarios discussed in Section 4.2.

Some practical guidelines of using utility functions for a design with ordinal outcomes in the two-cycle setting are provided in Section 1 of the Supplementary Material (available at *Biostatistics* online).

2.3 *Action set*

Equation (2.2) includes two maximizations to determine $d_1^\star$ and $d_2^\star$. In the discussion thus far, we have not used the particular elements of $\mathcal{D}$, and they might have been any actions. In actual dose-finding, ethical and practical constraints are motivated by the knowledge that, in general, higher doses carry a higher risk of more severe toxicity. We thus require a more restrictive action set, with additional conditions for the acceptability of a dose assignment.

The first additional criterion is that we *do not skip untried dose levels* when escalating. This rule is imposed almost invariably in actual trials with adaptive dose-finding methods. Let d_1^M denote the highest dose level among the dose levels that have been tried in cycle 1 and d_2^M the highest dose level among those that have been tried in either cycle 1 or cycle 2. The search for the optimal actions is constrained such that $1 \leqslant d_1 \leqslant d_1^M + 1$ and $1 \leqslant d_2 \leqslant d_2^M + 1$. In addition, we do not escalate a patient's dose level in cycle 2 if severe toxicity was observed in cycle 1 ($Y_1 = (J-1)$). Both restrictions are due to safety concerns.

A third safety restriction is defined implicitly in terms of the cycle-specific utility $u_{\text{cycle}}(\cdot, \cdot)$. A patient is not treated ($d_c = \text{NT}$) if there is no dose with expected utility $\geqslant u_{\text{cycle}}(0, 0)$. For d_1, the expected utility $U_1(d_1)$ is compared with the expected utility of not receiving any treatment in both cycles, $(1 + \lambda)u_{\text{cycle}}(0, 0)$ (horizontal dotted line in Figure 1(d)). Any d_1 with $U_1(d_1)$ below the line is not considered acceptable treatment. For d_2, the expected utility $U_2(d_1, d_2, \boldsymbol{y}_1)$ is similarly compared with the expected utility of $d_2 = \text{NT}$, $u_{\text{cycle}}(0, 0)$ (horizontal dotted line in Figure 1(a)–(c)), and any d_2 with $U_2(d_1, d_2, \boldsymbol{y}_1)$ below the line is not acceptable.

At any interim point in the trial, let $\mathcal{X}$ denote the current data, including dose assignments for previously enrolled patients. The three conditions together make the action sets for d_1 and d_2 dependent on $\mathcal{X}$, d_1, and $\boldsymbol{y}_1 = (Y_1, Z_1)$. We let $\mathcal{D}_1(\mathcal{X})$ and $\mathcal{D}_2(d_1, \boldsymbol{y}_1, \mathcal{X})$ denote the action sets for d_1 and d_2, respectively, that are implied by these three restrictions.

2.4 *Inference model*

Thus far, our discussion of optimal decisions has not included a particular probability model. We will assume a 4D ordinal probit model for (Y_1, Z_1, Y_2, Z_2) with a regression on doses d_1 and d_2, standardized to the domain $[0, 1]$, with $d_1 = 0$ and $d_m = 1$. Let $(u_{i,1}, v_{i,1}, u_{i,2}, v_{i,2})$ denote a vector of latent probit scores for the ith patient and let $\{\gamma_{y,j}\}$ and $\{\gamma_{z,k}\}$ denote fixed cutoffs that define $Y_{i,c} = j$ if $\gamma_{y,j-1} < u_{i,c} \leqslant \gamma_{y,j}$ and $Z_{i,c} = k$ if $\gamma_{z,k-1} < v_{i,c} \leqslant \gamma_{z,k}$, $c = 1, 2$. While varying the mean of distributions of $u_{i,c}$ and $v_{i,c}$ across

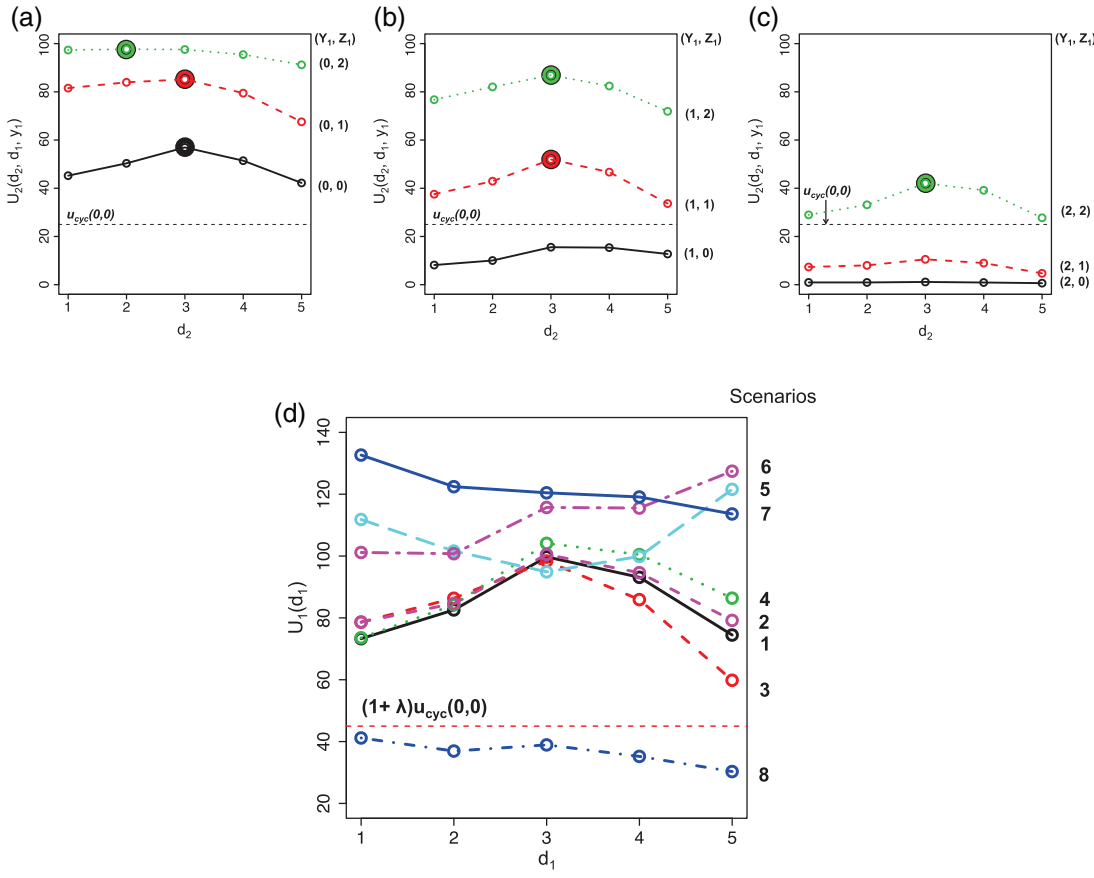

Fig. 1. (a)–(c) The true expected cycle 2 utilities of taking d_2 given $\boldsymbol{y}_1 = (Y_1, Z_1)$, $U_2(d_2, d_1, \boldsymbol{y}_1)$ with $d_1 = 3$ for scenario 3. Each panel corresponds to one of the three possible outcomes of Y_1. d_2 is acceptable only when its expected utility is greater than that of NT, $u_{\text{cycle}}(0, 0)$. $d_2^\star$ is marked with a bold circle for each $\boldsymbol{y}_1$ given $d_1 = 3$ if the corresponding expected utility is greater than $u_{\text{cycle}}(0, 0)$. If none of d_2 has an expected utility greater than $u_{\text{cycle}}(0, 0)$ for $\boldsymbol{y}_1$, $d_2^\star = \text{NT}$ and none of d_2 is marked with a bold circle. (d) Illustrates total expected utilities of d_1, $U_1(d_1)$ for the simulation scenarios assuming that the true $d_2^\star$ will be taken in cycle 2. d_1 is acceptable only when its utility is greater than that of NT in the two cycles, $(1 + \lambda)u_{\text{cycle}}(0, 0)$ (red dashed horizontal line at 45). (a) $Y_1 = 0$ (mild toxicity). (b) $Y_1 = 1$ (moderate toxicity). (c) $Y_1 = 2$ (severe toxicity). (d) $U_1(d_1)$.

cycles, the same cutoffs are used for all cycles. The $u_{i,c}$ and $v_{i,c}$ are multivariate normal probit scores,

$$(u_{i,1}, v_{i,1}, u_{i,2}, v_{i,2})' \sim \text{N}(\boldsymbol{\mu}_d, \Sigma) \quad \text{with } \Sigma = \begin{bmatrix} \sigma_Y^2 + \tau^2 & \rho\tau^2 & \tau^2 & \rho\tau^2 \\ \rho\tau^2 & \sigma_Z^2 + \tau^2 & \rho\tau^2 & \tau^2 \\ \tau^2 & \rho\tau^2 & \sigma_Y^2 + \tau^2 & \rho\tau^2 \\ \rho\tau^2 & \tau^2 & \rho\tau^2 & \sigma_Z^2 + \tau^2 \end{bmatrix}, \tag{2.6}$$

and $\boldsymbol{\mu}_d = (\mu_{u,1,d_1}, \mu_{v,1,d_1}, \mu_{u,2,d_2}, \mu_{v,2,d_2})'$. The covariance matrix implies associations across cycles and across outcomes through ρ and τ^2. Given that the ordinality of the outcomes is accounted for by the latent probit scores and fixed cutoff parameters $\{\gamma_{y,j}\}$ and $\{\gamma_{z,k}\}$, a simple yet flexible model for regression

on dose is obtained by assuming $\mu_{xcd} = \beta_{xc0} + \beta_{xc1} d_c^{\beta_{xc2}}$ with $x = u$ for toxicity and $x = v$ for efficacy. A discussion of nonlinear dose–response models is given by Bretz *and others* (2005). We assume that the toxicity and efficacy probabilities increase monotonic in dose by requiring $\beta_{xc1} > 0$ and $\beta_{xc2} > 0$. Denote $\boldsymbol{\beta}_{xc} = (\beta_{xc0}, \log \beta_{xc1}, \log \beta_{xc2})$, $x = u, v$ and $c = 1, 2$. We complete the model with a normal prior $\boldsymbol{\beta}_{xc} \sim N(\boldsymbol{\beta}_{x,c}, \Omega_{x,c})$, $x = u, v$.

3. TRIAL DESIGN

3.1 *Adaptive randomization*

Denote $\boldsymbol{d} = (d_1, d_2)$. Although, in terms of the utility-based objective function, $d^\star$ yields the best clinical outcomes for the next patient, the performance of the design, in terms of frequentist operating characteristics, can be improved by including adaptive randomization (AR) among actions giving values of the objective function near the maximum at $d^\star$. Using AR decreases the probability of getting stuck at a suboptimal d, and also has the effect of treating more patients at doses having larger utilities, on average. The problem that a "greedy" search algorithm may get stuck at suboptimal actions, and the simple solution of introducing additional randomness into the search process, are well known in the optimization literature (cf. Tokic, 2010). This has been dealt with only very recently in dose-finding (Bartroff and Lai, 2010; Azriel *and others*, 2011; Braun *and others*, 2012; Thall and Nguyen, 2012).

To implement AR, we first define ϵ_i to be a function decreasing in patient index i, and denote $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)$. We define the set of ϵ_i*-optimal doses for cycle 1* to be

$$\mathcal{D}_{i,1} = \{d_1 : |U_1(d_{1,i}^\star) - U_1(d_1)| < \epsilon_i, d_1 \in \mathcal{D}_1(\mathcal{X})\}. \tag{3.1}$$

The set, $\mathcal{D}_{i,1}$ contains doses d_1 in $\mathcal{D}_1(\mathcal{X})$ whose $U_1(d_1)$ is within ϵ_i of the maximum posterior mean utility. Similarly, we define the set of $(\epsilon_i/2)$-optimal doses for cycle 2 given $(d_{i,1}, y_{i,1})$ to be

$$\mathcal{D}_{i,2} = \{d_2 : |\tilde{U}_2(d_{i,2}^\star(d_{i,1}, \boldsymbol{y}_{i,1})|\boldsymbol{y}_{i,1}, d_{i,1}) - \tilde{U}_2(d_2 \mid d_{i,1}, \boldsymbol{y}_{i,1}) | < \epsilon_i/2,\ d_2 \in \mathcal{D}_2(d_{i,1}, \boldsymbol{y}_{i,1}, \mathcal{X})\}. \tag{3.2}$$

$\mathcal{D}_{i,2}$ in (3.1) is based on (2.5). Our design randomizes patients uniformly among doses in $\mathcal{D}_{i,1}$ for $c = 1$ and $\mathcal{D}_{i,2}$ for $c = 2$, which we call AR($\boldsymbol{\epsilon}$). Numerical values of ϵ_i depend on the range of $u_{\text{cycle}}(y, z)$, and are determined by preliminary trial simulations in which $\boldsymbol{\epsilon}$ is varied.

3.2 *Illustrative trial*

Our illustrative trial studied in the simulations is a stylized version of the phase I–II chemotherapy trial with five dose levels described in Section 1, but here accounting for two cycles of therapy. The maximum sample size is 60 patients with a cohort size of 2. Based on preliminary simulations, we set $\epsilon_i = 20$ for the first 10 patients, $\epsilon_i = 15$ for the next 10 patients, and $\epsilon_i = 10$ for the remaining 40 patients. An initial cohort of 2 patients is treated at the lowest dose level in cycle 1, their cycle 1 toxicity and efficacy outcomes are observed, the posterior of $\boldsymbol{\beta}_{xc}$, $x = u, v$ and $c = 1, 2$ is computed, and actions are taken for cycle 2 of the initial cohort. If $\mathcal{D}_{i,2} = \{\text{NT}\}$, then patient i does not receive a second cycle of treatment. If $\mathcal{D}_{i,2} \neq \{\text{NT}\}$, then AR($\boldsymbol{\epsilon}$) is used to choose an action for cycle 2 from $\mathcal{D}_{i,2} \setminus \{\text{NT}\}$. When the toxicity and efficacy outcomes are observed from cycle 2, the posterior of $\boldsymbol{\beta}_{xc}$ is updated. The second cohort is not enrolled until the first cohort has been evaluated for cycle 1. For all cohorts after the first, after the outcomes of all previous cohorts are observed, the posterior is updated, the posterior expected utility, $U_1(d_1)$ is computed using $\lambda = 0.8$, and $\mathcal{D}_1(\mathcal{X})$ is determined. Using $\mathcal{D}_1(\mathcal{X})$ and ϵ_i, we find $\mathcal{D}_{i,1}$ and search for $d_1 \in \mathcal{D}_{i,1} \setminus \{\text{NT}\}$. If $\mathcal{D}_{i,1} = \{\text{NT}\}$ for any interim $\mathcal{X}$, then $d_{i,1} = \text{NT}$, and the trial is terminated. If $\mathcal{D}_{i,1} \neq \{\text{NT}\}$, we then choose a cycle 1 dose from $\mathcal{D}_{i,1} \setminus \{\text{NT}\}$ using AR($\boldsymbol{\epsilon}$). Once the outcomes in cycle 1 are observed, the posterior

is updated. Using $(d_{i,1}, \boldsymbol{y}_{i,1}, \mathcal{X})$ and ϵ_i, $\mathcal{D}_{i,2}$ is searched. If $\mathcal{D}_{i,2}$ contains NT only, then $d_{i,2} = \text{NT}$ and no cycle 2 dose is given to patient i. Otherwise, $d_{i,2}$ is selected from $\mathcal{D}_{i,2} \setminus \{\text{NT}\}$ using AR($\boldsymbol{\epsilon}$). The toxicity and efficacy outcomes are observed from cycle 2 and the posterior of $\boldsymbol{\beta}_{xc}$ is updated. The above steps are repeated until either the trial has been stopped early or $N = 60$ has been reached. At the end of the trial, we record $d_1^\star$ as recommended first cycle dose $d_{1,\text{sel}}$ and $d_2^\star(d_1, \boldsymbol{y}_1)$ as optimal policy $d_{2,\text{sel}}(d_1, \boldsymbol{y}_1)$. If the trial is early terminated, let $d_{1,\text{sel}} = \text{NT}$ and $d_{2,\text{sel}} = \text{NT}$ for all $\boldsymbol{y}_1$.

4. Simulation study

4.1 *Designs for comparison*

Let DTD-O2 denote the proposed decision-theoretic two-cycle design. We compare DTD-O2 with three other designs. The first is obtained by reducing each 3-level efficacy and toxicity outcome to a 2-category (binary) variable by combining categories, but using the same probability model to ensure a fair comparison. The next two comparators are single cycle designs. The first, called Single Cycle Comparator 1 (SCC1), assumes no association between cycles and optimizes d_1 and d_2 separately. The second, called Single Cycle Comparator 2 (SCC2), does not distinguish between cycles and treats the two cycles identically.

For SCC1, we assume patient-specific random probit scores, independent over cycles, $(u_{i,c}, v_{i,c}) \overset{\text{indep}}{\sim} \text{N}(\boldsymbol{\mu}_{dc}, \Sigma_{11})$, where $\boldsymbol{\mu}_{dc} = (\mu_{u,c,d_c}, \mu_{v,c,d_c})$ and Σ_{11} is the 2×2 covariance matrix. We let Σ_{11} be the upper-left partition of Σ in (2.6). Owing to the independence of probit scores over cycles within a patient, SCC1 models the association between Y_c and Z_c within the same cycle only and does not assume any association between outcomes in different cycles, for example, Y_1 and Y_2. The other model specification including the regression of $\boldsymbol{\mu}_{dc}$ on the dose in Section 2.4 stays the same. For SCC2, in addition to having patient- and cycle-specific random probit scores as in SCC1, we assume that the mean dose effects are identical in the two cycles by dropping the cycle index from β in Section 2.4, i.e. setting $\mu_{x,1,d} = \mu_{x,2,d}$, $x = u, v$ for all d. For these two methods, we apply the acceptability rules in Section 2.3 and the AR rules in Section 3.1 for each cycle separately. For example, a trial is terminated if $\int u(d_1, \boldsymbol{y}_1)\, \mathrm{d}p(\boldsymbol{y}_1 \mid d_1) < u_{\text{cycle}}(0, 0)$ for all d_1 and $\mathcal{D}_{i,c}$ is defined with $\int u(d_c, \boldsymbol{y}_c)\, \mathrm{d}p(\boldsymbol{y}_c \mid d_c)$ only. Also, the no-escalation rule after $Y_1 = (J-1)$, no-skipping rule and AR similar to those implemented in the proposed method are implemented to SCC1 and SCC2.

4.2 *Simulation setup*

We simulated trials under each of 8 scenarios using each of the designs. A total of $N = 1000$ trials were simulated for each design under each scenario. The simulation scenarios were determined by fixing a set of marginal probabilities and regression coefficients on probit scores, given in Table 2 and Supplementary Material Table S1 (available at *Biostatistics* online). Each simulation scenario is specified by the marginal distributions of Y and Z. Table 2 gives the true $p_{d,j} = P(Y \leqslant j \mid d)$ and $q_{d,k} = P(Z \leqslant k \mid d)$ under each scenario. The corresponding probit scores are $\bar{\xi}_j(d) = \Phi^{-1}(p_{d,j})$ and $\bar{\eta}_k(d) = \Phi^{-1}(q_{d,k})$, where Φ is the cumulative distribution function of the standard normal distribution. To ensure a fair comparison, we intentionally define a simulation truth that is different from the assumed model used by the design methodology. The simulation model is best described as a generative model, first for Y_1, then Z_1 given Y_1, and then (Y_2, Z_2) given Y_1, Z_1.

Generating Y_1: We first generate Y_1 from the distribution specified by $P(Y_1 \leqslant j_1 \mid d_1) = \Phi(\tilde{\xi}_{1,j_1}(d_1))$, where $\tilde{\xi}_{1,j_1}(d_1) = \bar{\xi}_{j_1}(d_1)$. For later reference, we define a rescaled variable Y as $\tilde{Y} = (2Y - J - 1)/(J-1)$, which is evenly spaced in $[-1, 1]$.

Generating $Z_1 \mid Y_1$: Conditional on Y_1, we specify a distribution of Z_1 by letting

$$\phi^{-1}\{\Pr(Z_1 \leqslant k_1 \mid d_1, Y_1)\} = \bar{\eta}_{k_1}(d_1) + w_{1,1}\tilde{Y}_1$$

Table 2. *Assumed probabilities, $P(Y = y \mid d)$ and $P(Z = z \mid d)$. These marginal probabilities are used to determine probit scores, $\bar{\xi}_j(d)$ and $\bar{\eta}_k(d)$*

		Toxicity outcome			Efficacy outcome		
Scenarios	Dose	Mild	Moderate	Severe	PD	SD	PR/CR
1, 2	1	0.23	0.52	0.25	0.44	0.44	0.12
	2	0.225	0.515	0.26	0.35	0.42	0.23
3, 4	3	0.20	0.530	0.27	0.18	0.40	0.42
	4	0.18	0.40	0.42	0.10	0.445	0.455
	5	0.06	0.20	0.74	0.08	0.45	0.47
5	1	0.53	0.39	0.08	0.35	0.515	0.135
	2	0.38	0.47	0.15	0.325	0.52	0.155
	3	0.33	0.46	0.21	0.31	0.528	0.162
	4	0.315	0.455	0.23	0.225	0.505	0.27
	5	0.375	0.375	0.25	0.05	0.39	0.56
6	1	0.55	0.30	0.15	0.51	0.31	0.18
	2	0.475	0.31	0.215	0.45	0.275	0.275
	3	0.45	0.31	0.24	0.18	0.39	0.43
	4	0.44	0.31	0.25	0.15	0.40	0.45
	5	0.43	0.30	0.27	0.03	0.27	0.70
7	1	0.65	0.20	0.15	0.18	0.33	0.49
	2	0.52	0.20	0.28	0.175	0.325	0.50
	3	0.46	0.21	0.33	0.15	0.30	0.55
	4	0.37	0.27	0.36	0.125	0.25	0.625
	5	0.28	0.28	0.44	0.11	0.24	0.65
8	1	0.19	0.43	0.38	0.85	0.12	0.03
	2	0.13	0.22	0.65	0.78	0.14	0.08
	3	0.09	0.22	0.69	0.54	0.31	0.15
	4	0.03	0.23	0.74	0.43	0.39	0.18
	5	0.01	0.13	0.86	0.38	0.41	0.21

with coefficient $w_{1,1}$. Here, $w_{1,1}$ induces association between the cycle 1 outcomes, Y_1 and Z_1. A negative value of $w_{1,1}$ leads to a negative association between Y_1 and Z_1, that is, $P(Z_1 = K - 1 \mid d_1, Y_1 = j_1) \geqslant P(Z_1 = K - 1 \mid d_1, Y_1 = j_1')$, $j_1 < j_1'$. For later use, we define $\tilde{Z}$ by rescaling Z to be evenly spaced in $[-1, 1]$, similarly to $\tilde{Y}$.

Generating $Y_2 \mid Y_1, Z_1$: We generate Y_2 using

$$\phi^{-1}\{\Pr(Y_2 \leqslant j_2 \mid d_1, \boldsymbol{y}_1, d_2)\} = \bar{\xi}_{j_2}(d_2) + w_{2,1}\tilde{d}_1 + w_{2,2}r^T(d_1, Y_1) + w_{2,3}\tilde{Z}_1$$

Here, $\tilde{d}$ is a standardized dose in $[-1, 1]$. We restrict $w_{2,1}$, $w_{2,2} \leqslant 0$ and $w_{2,3} \geqslant 0$ to induce a positive association of Y_2 with d_1 and Y_1, and negative association with Z_1. Here, $r^T(d, j)$ determines how d_1 and Y_1 jointly affect Y_2. A large negative value of $r^T(d_1 = 1, J - 1)$ implies that given that $Y_1 = J - 1$ (severe toxicity) is observed at $d_1 = 1$, the probability of observing $Y_2 = j$, $j \neq 0$ greatly increases for all d_2. Similarly, observing $Y_1 = 0$ (mild toxicity) at $d_1 = 5$ greatly increases the probability of observing $Y_2 = 0$ for all d_2, implying a large positive value of $r^T(d_1 = 5, 0)$.

Table 3. *True optimal actions, $d_1^\star$ and $d_2^\star(d_1^\star, \boldsymbol{y}_1)$*

			$d_2^\star(d_1^\star, y_1)$		
			Z_1		
Scenarios	$d_1^\star$	Y_1	0	1	2
1	3	0	3	3	3
		1	3	3	3
		2	NT	NT	NT
2	3	0	3	3	3
		1	NT	3	3
		2	NT	NT	3
3	3	0	3	3	2
		1	NT	3	3
		2	NT	NT	3
4	3	0	NT	3	3
		1	NT	3	3
		2	NT	3	3
5	3	0	5	5	5
		1	5	5	5
		2	5	5	5
6	1	0	5	5	5
		1	5	5	5
		2	5	5	5
7	5	0	1	1	1
		1	1	1	1
		2	1	1	1
8	NT	0	NT	NT	NT
		1	NT	NT	NT
		2	NT	NT	NT

Generating $Z_2 \mid Y_1, Z_1, Y_2$: We use

$$\phi^{-1}\{\Pr(Z_2 \leqslant k_2 \mid d_1, Y_1, Z_1, d_2, Y_2)\} = \bar{\eta}_{k_2}^\star(d_2) + w_{3,1}\tilde{d}_1 + w_{3,2}\tilde{Y}_1 + w_{3,3}r^E(d_1, Z_1) + w_{3,4}\tilde{Y}_2,$$

where $w_{3,1}, w_{3,3} \leqslant 0$ and $w_{3,2}, w_{3,4} \geqslant 0$. Similar to r^T, r^E determines a joint effect of d_1 and Z_1 on Z_2. The detailed specification of the coefficients, r^T and r^E for each simulation scenario is described in the Supplementary Materials (available at *Biostatistics* online). Table 3 shows the optimal actions, $d_1^\star$ and $d_2^\star(d_1^\star, Y_1, Z_1)$, over two cycles under each of the 8 simulation scenarios under the simulation truth. For example, in Scenario 3, the optimal cycle 1 action is to give dose level 3, and the optimal cycle 2 action is to treat patients with $Y_1 = 0$ at $d_2 = 4$, and at $d_2 = 2$ if $Y_1 = 1$.

We calibrate the fixed hyperparameters, $\tilde{\boldsymbol{\theta}} = (\sigma_Y^2, \sigma_Z^2, \tau^2, \rho, \bar{\boldsymbol{\beta}}_{x,c}, \Omega_{x,c})$, for $x = u, v$ and $c = 1, 2$ and the cutoff points, $(\boldsymbol{\gamma}_y, \boldsymbol{\gamma}_z)$, using effective sample size (ESS), described in the Supplementary Materials (available at *Biostatistics* online). We set $\tilde{\boldsymbol{\theta}}$ and the cutoffs, $\boldsymbol{\gamma}_x$ and $\boldsymbol{\gamma}_y$, and simulate 1000 pseudo-samples of $\bar{\boldsymbol{\beta}}_{x,c}, \Omega_{x,c}, x = u, v$ and $c = 1, 2$. We then compute probabilities of interest based on the pseudo-samples, such as $\mathrm{P}(Y_c = j \mid d_c)$ and $\mathrm{P}(Z_c = k \mid d_c)$, $c = 1, 2$. For all simulations, we determined $\tilde{\boldsymbol{\theta}}$ to give each prior

ESS between 0.5 and 2, using the approximation obtained by matching moments with a Dirichlet distribution. We used the same $\tilde{\boldsymbol{\theta}}$ for SCC1 and SCC2.

4.3 *Evaluation criteria*

We evaluate design performance for the patients treated in the trial using three different summary statistics, $\bar{u}$, $\bar{U}_{\text{trt}}$, and $\bar{U}_{\text{sel}}$. Recall that in a trial we record the clinical outcomes of the n patients with their assigned doses and recommended doses for future patients, $\boldsymbol{y}_{i,c} = (Y_{i,c}, Z_{i,c})$, d_{ic}, $i = 1, \ldots, n$ and $c = 1, 2$, and $(d_{1,\text{sel}}, d_{2,\text{sel}}(d_{1,\text{sel}}, Y_1, Z_1))$, respectively. We index the N simulated replications of the trial by $\ell = 1, \ldots, N$. We define average utility for the n patients in the ℓth simulated trial in two different ways; $u^{(\ell)} = \sum_{i=1}^{n}\{u_{\text{cycle}}(\boldsymbol{y}_{i,1}) + u_{\text{cycle}}(\boldsymbol{y}_{i,2})\}/n$ and $U_{\text{trt}}^{(\ell)} = \sum_{i=1}^{n}\{E^{\text{true}}(u_{\text{cycle}}(\boldsymbol{y}) \mid d_{i,1}) + \lambda E^{\text{true}}(u_{\text{cycle}}(\boldsymbol{y}) \mid d_{i,1}, d_{i,2}, \boldsymbol{y}_{i,1})\}/n$. Note that $u^{(\ell)}$ is a function only of occurred outcomes, $(\boldsymbol{y}_{i,1}, \boldsymbol{y}_{i,2})$, whereas $U_{\text{trt}}^{(\ell)}$ depends on the true utilities of assigned doses $(d_{i,1}, d_{i,2})$. For $u^{(\ell)}$ and $U_{\text{trt}}^{(\ell)}$, $u_{\text{cycle}}(0, 0)$ is used as the utility for patients with $d_{i,c} = \text{NT}$. The empirical mean total payoffs taken over all simulated trials are

$$\bar{u} = \frac{1}{N}\sum_{\ell=1}^{N} u^{(\ell)} \quad \text{and} \quad \bar{U}_{trt} = \frac{1}{N}\sum_{\ell=1}^{N} U_{\text{trt}}^{(\ell)}.$$

One may regard $\bar{u}$ and $\bar{U}_{\text{trt}}$ as indexes of the ethical desirability of the method, given $u_{\text{cycle}}(y, z)$.

The proposed method gives an optimal action $d_{1,\text{sel}}$ for cycle 1, and policy $d_{2,\text{sel}}$ for cycle 2. We let $d_{2,\text{sel}} = \text{NT}$ for all (Y_1, Z_1) if $d_{1,\text{sel}} = \text{NT}$, so the trial is terminated early. We use $d_{1,\text{sel}}$ and $d_{2,\text{sel}}$ to evaluate performance in terms of future patient benefit. Under SCC1 and SCC2, $d_{2,\text{sel}}$ is not a function of (Y_1, Z_1). For SCC2, $d_{1,\text{sel}}$ and $d_{2,\text{sel}}$ are identical. Assuming that the simulation truth is known, we define the expected payoff in cycle 1 of giving action $d_{1,\text{sel}}$ to a future patient as $U_{1,\text{sel}}(d_{1,\text{sel}}) = E^{\text{true}}\{u_{\text{cycle}}(\boldsymbol{y}_1) \mid d_{1,\text{sel}}\}$ for $d_{1,\text{sel}} \neq \text{NT}$. That is the expected utility with respect to the assumed distribution of $\boldsymbol{y}_1$ when $d_{1,\text{sel}}$ is given. For $d_{1,\text{sel}} = \text{NT}$, let $U_{1,\text{sel}}(d_{1,\text{sel}}) = u_{\text{cycle}}(0, 0)$. This expectation is computed under the distribution of $\boldsymbol{y}_1$ given $d_{1,\text{sel}}$. If the rule $d_{2,\text{sel}}$ is used, the expected cycle 2 payoff is

$$U_{2,\text{sel}}(d_{2,\text{sel}}) = \sum_{\substack{\boldsymbol{y}_1 \in \{0,\ldots,J-1\} \\ \times\{0,\ldots,K-1\}}} E^{\text{true}}\{u_{\text{cycle}}(\boldsymbol{y}_2) \mid d_{1,\text{sel}}, d_{2,\text{sel}}(\boldsymbol{y}_1)\} p^{\text{true}}(\boldsymbol{y}_1 \mid d_{1,\text{sel}}),$$

where $E^{\text{true}}\{u_{\text{cycle}}(y_2) \mid d_{1,\text{sel}}, d_{2,\text{sel}}(\boldsymbol{y}_1)\}$ becomes $u_{\text{cycle}}(0, 0)$ if $d_{2,\text{sel}}(\boldsymbol{y}_1) = \text{NT}$. The total expected payoff to a future patient treated using the optimal regime $\boldsymbol{d}_{\text{sel}} = (d_{1,\text{sel}}, d_{2,\text{sel}})$ is defined to be $U_{\text{sel}}(\boldsymbol{d}_{\text{sel}}) = U_{1,\text{sel}}(d_{1,\text{sel}}) + \lambda U_{2,\text{sel}}(d_{2,\text{sel}})$.

4.4 *Comparison to designs with binary outcomes*

We first compare DTD-O2 with designs obtained by collapsing each trinary toxicity and efficacy outcome to a binary variable. This mimics what often is done in practice in order to apply a phase I–II design based on binary efficacy and toxicity. We use an appropriately reduced version of our assumed underlying model to ensure a fair comparison. Since this reduction is not unique, we exhaustively define binary outcomes in four different ways, binary cases 1–4, given in Section 4 of the Supplementary Material (available at *Biostatistics* online). The utilities associated with the binary outcomes are defined accordingly based on the utilities in Table 1. The results, in terms of $\bar{u}$, $\bar{U}_{\text{trt}}$, and U_{sel}, are summarized graphically in Figure 2. Scenario 8 is not included in Figure 2 because the optimal action is NT in both cycles, and in this case all designs stop the trial early with high probability, The figure shows that reducing to binary outcomes can produce designs with much worse performance than DTD-O2, while for some cases the performance may be comparable. The binary outcome design's performance also varies substantially with the particular

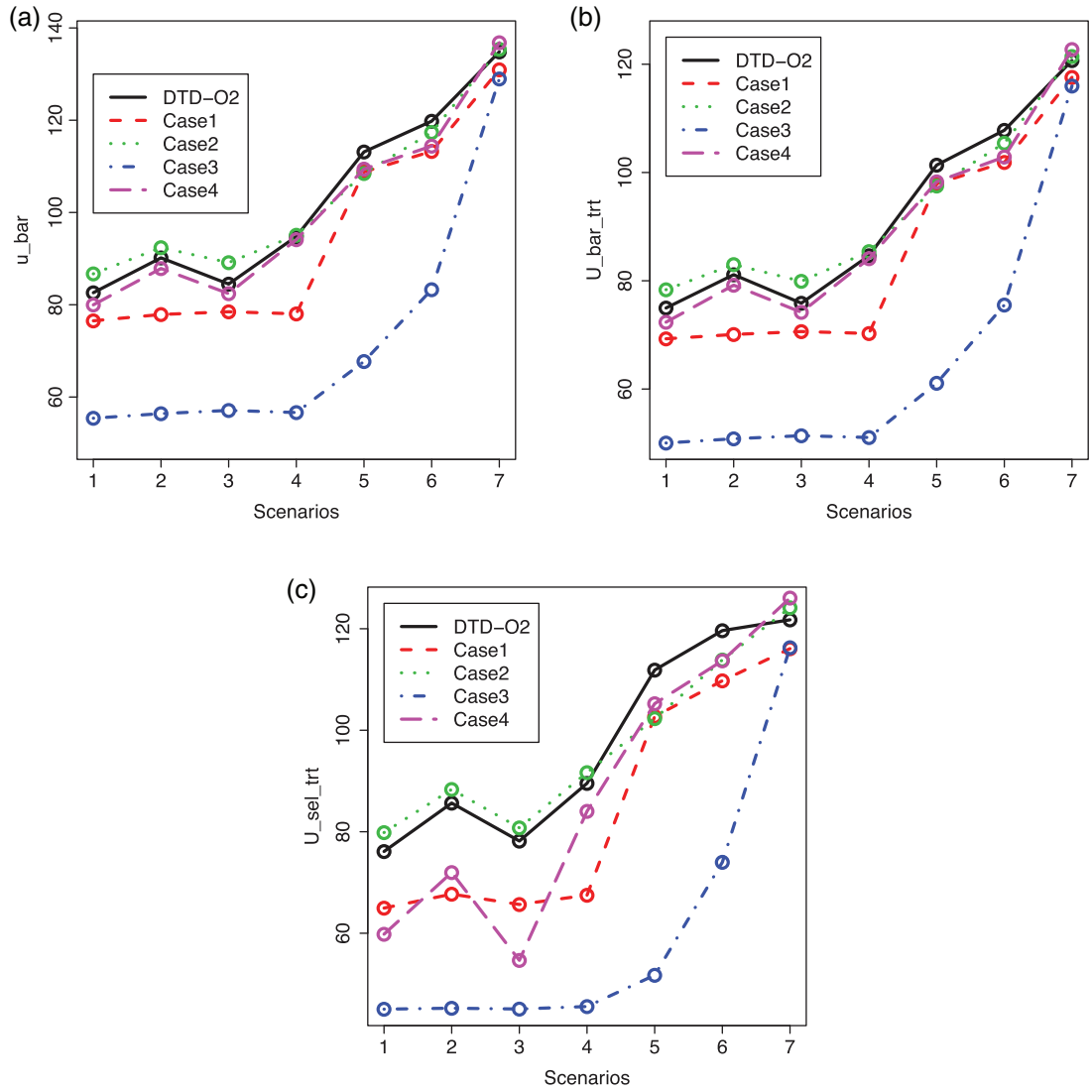

Fig. 2. Plots of $(\bar{u}, \bar{U}_{\text{trt}}, U_{\text{sel}})$ for a comparison of DTD-O2 vs. a design with binary outcomes. Here, $\bar{u}$, $\bar{U}_{\text{trt}}$, and U_{sel} represent empirical mean utilities of patients treated in the trial, true mean utilities of treatments given to patients in the trial, and true expected utilities chosen for future patients, respectively. (a) $\bar{u}$. (b) $\bar{U}_{\text{trt}}$. (c) $\bar{U}_{\text{sel}}$.

dichotomization used. Since different physicians may combine ordinal categories in different ways, the practical implication is that the additional complexity of the ordinal outcome design is worthwhile, in terms of benefit to both the patients treated in the trial and future patients.

4.5 *Comparison to single cycle designs*

The simulation results for DTD-O2, SCC1, and SCC2 are summarized in Figure 3. Scenarios 1–4 have the same marginal toxicity and efficacy probabilities, but different values of coefficients $(\boldsymbol{w})$, yielding different probit scores and different association structures of $d_1, \boldsymbol{y}_1, d_2$, and $\boldsymbol{y}_2$. Scenario 1 has large $w_{2,2}$ and $w_{3,2}$, so that the cycle 1 toxicity outcome greatly affects cycle 2 expected utilities in the simulation truth. As

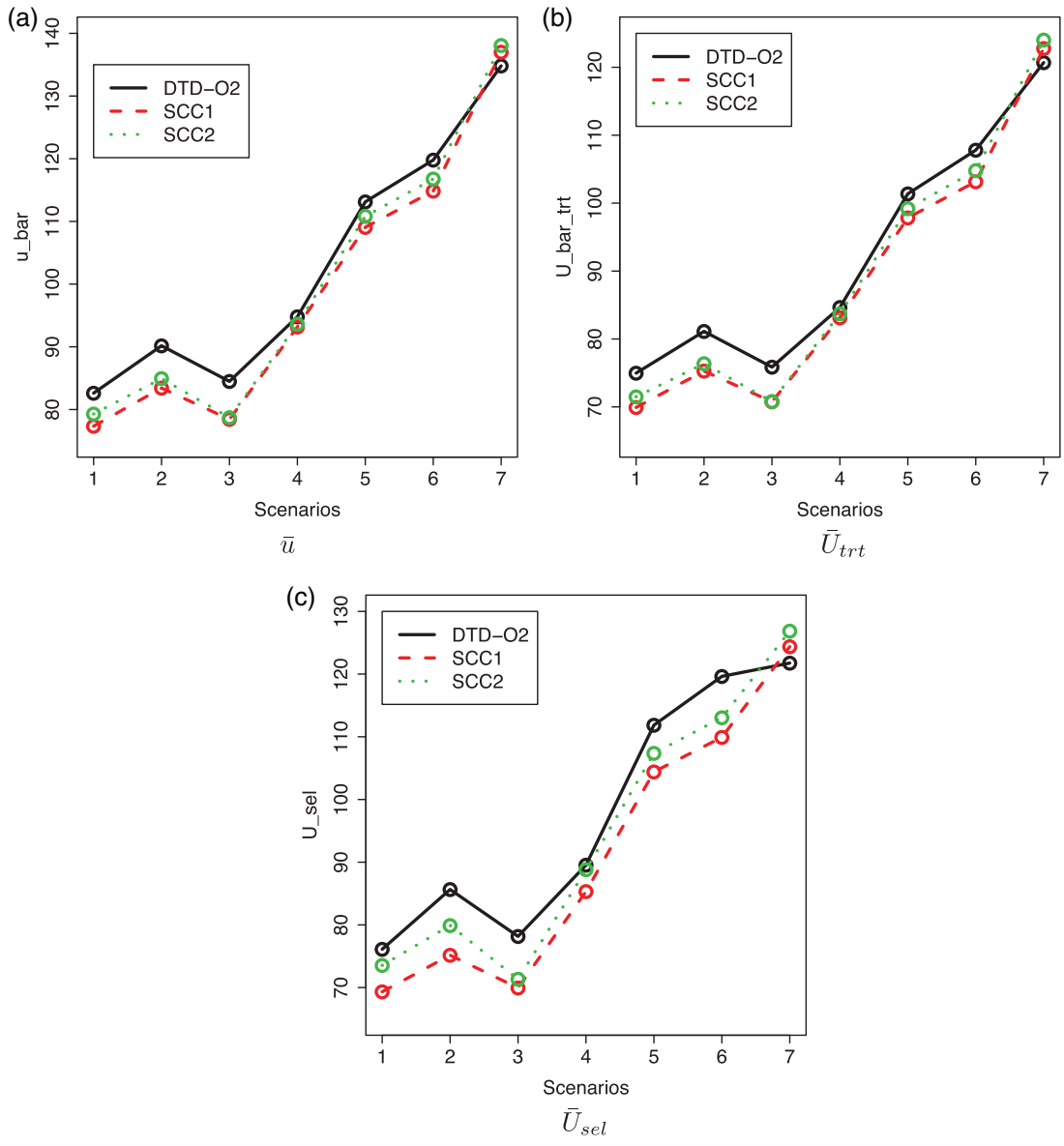

Fig. 3. Plot of $(\bar{u}, \bar{U}_{\text{trt}}, U_{\text{sel}})$ for a comparison with SCC1 and SCC2. Here, $\bar{u}$, $\bar{U}_{\text{trt}}$, and U_{sel} represent empirical mean utilities of patients treated in the trial, true mean utilities of treatments given to patients in the trial, and true expected utilities chosen for future patients, respectively. (a) $\bar{u}$. (b) $\bar{U}_{\text{trt}}$. (c) $\bar{U}_{\text{sel}}$.

shown in Table 3, the optimal action in cycle 2 after observing severe toxicity in cycle 1 is NT regardless of the cycle 1 efficacy outcome. Scenario 4 is similar to Scenario 1 but the cycle 1 efficacy outcome heavily affects the cycle 2 treatment in that all cycle 2 treatments are less desirable than NT when PD is observed in cycle 1. In Scenarios 2 and 3, the two cycle 1 outcomes jointly determine the cycle 2 treatment as shown in

the tables. Scenario 3 has larger association between Y_c and Z_c within each cycle. In Scenarios 1–4, modeling dependence across cycles improves the performance, as shown in Figure 3, where DTD-O2 is superior to SCC1 and SCC2 in terms of all the three criteria, $\bar{u}$, $\bar{U}_{\text{trt}}$, and U_{sel}. Since the only difference between DTD-O2 and SCC1 is whether the two cycles are modeled jointly or separately, the results show that the joint modeling significantly improves the performance. Differences in the performance are smaller for Scenarios 1 and 4. This may be because the true structure that one cycle 1 outcome dominates cycle 2 decisions in the scenarios is not easily accommodated under the assumed covariance structure in (2.6) and each trial gets only a small number of patients. In such a case, separate estimation for the two cycles may not be a very poor approach. In addition, the three methods are compared using $\bar{u}$ and $\bar{U}_{\text{trt}}$ based on the last 20 patients in each trial for the three designs (not shown). This comparison shows that the improvement by DTD-O2 over the other two methods becomes greater, especially for Scenarios 1 and 4. It may imply that learning takes more patients for DTD-O2 when there is a discrepancy between the truth and the model assumption.

Scenarios 5–7 have different shapes for $U_1(d_1)$ as a function of d_1. The cycle 1 utilities are U-shaped in Scenario 5, monotone increasing in Scenario 6, and monotone decreasing in Scenario 7. Very mild associations between outcomes and between cycles are assumed for these scenarios. For Scenarios 5 and 6, DTD-O2 achieves notably better performance (see Figure 3), with $\bar{u}$ and U_{sel} similar to each other for DTD-O2. This implies that DTD-O2 identifies desirable actions early in the trial, treats many of the patients with the desirable actions, and has a high probability of selecting truly optimal actions at the end of a trial. In Scenario 7, DTD-O2 shows slightly worse performance (see the rightmost of Figure 3). In the simulation truth of Scenario 7, the cycle 1 expected utility does not change much with d_1, but the cycle 2 expected utility is very sensitive to d_1, Y_1, and Z_1. This is a very challenging case for DTD-O2, and not modeling dependence between the cycles leads to better performance than incorrectly modeling in this particular scenario. Scenario 8 has no acceptable dose in either cycle. All the three methods terminate the trials with probability 1 in this case, with mean sample sizes 9.11, 8.33, and 8.29.

In all 8 scenarios, SCC2 yields better results than SCC1. This may be because $d_2^\star$ and $d_1^\star$ happen to be identical in many cases, so combining outcomes from the two cycles works well. However, the results for Scenarios 1–4 show that using each patient's cycle 1 dose and outcomes to select d_2 gives significantly superior performance in cases where there is significant dependence between the two cycles. More results are summarized using empirical toxicity and efficacy probabilities in Section 3 of Supplementary Material (available at *Biostatistics* online).

We carried out a sensitivity analysis in λ, under Scenarios 2 and 5, including the four binary outcome designs, SCC1, SCC2, and DTD-O2, for $\lambda = 0.0$, 0.4, 0.8, and 1.0. The results, given in Section 5 of Supplementary Material (available at *Biostatistics* online), show that changes in design performance with λ are very small, but $\lambda = 0$, corresponding to no use of cycle 2 utility in making a decision at cycle 1, yields higher early termination probabilities for binary outcome cases 1 and 3.

5. Discussion

We have extended the decision-theoretic two-cycle phase I–II dose-finding method in Lee *and others* (2015) to accommodate ordinal outcomes. Our simulations show that incorporating cycle 1 information into the cycle 2 treatment decision yields good performance for both patients treated in a trial and future patients. The simulations in Figure 2 show that this extension may greatly improve design performance, quantified by $\bar{u}$, $\bar{U}_{\text{trt}}$, and U_{sel}, compared with using binary toxicity and efficacy indicators. The proposed model and method also compared quite favorably with either assuming the two cycles are independent or ignoring the distinction between cycles 1 and 2.

In theory, DTD-O2 could be extended to more than two cycles. For this to be tractable, additional modeling assumptions may required to control the number of parameters, since decisions must be made

based on small sample sizes. Two possible approaches are to model dependence among cycles as a function of distance between cycles, or to make a Markovian assumption.

SUPPLEMENTARY MATERIAL

Supplementary material is available at http://biostatistics.oxfordjournals.org.

ACKNOWLEDGEMENTS

We specifically acknowledge the assistance of Lorenzo Pesce (University of Chicago). *Conflict of Interest*: None declared.

FUNDING

Y.J. research is supported in part by NIH R01 CA132897. P.F.T. research was supported in part by NIH R01 CA 83932. P.M. research was supported in part by NIH 1-R01-CA157458-01A1. This research was supported in part by NIH through resources provided by the Computation Institute and the Biological Sciences Division of the University of Chicago and Argonne National Laboratory, under grant S10 RR029030-01.

REFERENCES

AZRIEL, D., MANDEL, M. AND RINOTT, Y. (2011). The treatment versus experimentation dilemma in dose finding studies. *Journal of Statistical Planning and Inference* **141**(8), 2759–2768.

BARTROFF, J. AND LAI, T. L. (2010). Approximate dynamic programming and its applications to the design of phase I cancer trials. *Statistical Science* **25**(5), 245–257.

BELLMAN, R. (1957) *Dynamic Programming*, 1 edition. Princeton, NJ, USA: Princeton University Press.

BRAUN, T. M., KANG, S. AND TAYLOR, J. M. G. (2012). A phase I/II trial design when response is unobserved in subjects with dose-limiting toxicity. *Statistical Methods in Medical Research*. 0962280212464541.

BRETZ, F., PINHEIRO, J. C. AND BRANSON, M. (2005). Combining multiple comparisons and modeling techniques in dose–response studies. *Biometrics* **61**(3), 738–748.

CHEUNG, Y. K., CHAKRABORTY, B. AND DAVIDSON, K. W. (2014). Sequential multiple assignment randomized trial (smart) with adaptive randomization for quality improvement in depression treatment program. *Biometrics* **71**(2), 450–459.

IASONOS, A., ZOHAR, S. AND O'QUIGLEY, J. (2011). Incorporating lower grade toxicity information into dose finding designs. *Clinical Trials* **8**(4), 370–379.

LEE, J., THALL, P. F., JI, Y. AND MÜLLER, P. (2015). Bayesian dose-finding in two treatment cycles based on the joint utility of efficacy and toxicity. *Journal of the American Statistical Association* **110**(510), 711–722.

MURPHY, S. A. (2003). Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **65**(2), 331–355.

PARMIGIANI, G. AND INOUE, L. (2009) *Decision Theory: Principles and Approaches*. New York: Wiley.

ROBERT, C. P. (2007) *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*, 2nd edition. Berlin: Springer.

THALL, P. F. AND COOK, J. D. (2004). Dose-finding based on efficacy–toxicity trade-offs. *Biometrics* **60**(3), 684–693.

Thall, P. F and Nguyen, H. Q. (2012). Adaptive randomization to improve utility-based dose-finding with bivariate ordinal outcomes. *Journal of Biopharmaceutical Statistics* **22**(4), 785–801.

Tokic, M. (2010). Adaptive ε-greedy exploration in reinforcement learning based on value differences. In: *KI 2010: Advances in Artificial Intelligence*. Berlin: Springer, pp. 203–210.

Van Meter, E. M., Garrett-Mayer, E. and Bandyopadhyay, D. (2012). Dose-finding clinical trial design for ordinal toxicity grades using the continuation ratio model: an extension of the continual reassessment method. *Clinical Trials* **9**(3), 303–313.

Yin, G., Li, Y. and Ji, Y. (2006). Bayesian dose-finding in phase i/ii clinical trials using toxicity and efficacy odds ratios. *Biometrics* **62**(3), 777–787.

Zhang, W., Sargent, D. J. and Mandrekar, S. (2006). An adaptive dose-finding design incorporating both toxicity and efficacy. *Statistics in Medicine* **25**(14), 2365–2383.

Zhao, Y., Zeng, D., Socinski, M. A. and Kosorok, M. R. (2011). Reinforcement learning strategies for clinical trials in nonsmall cell lung cancer. *Biometrics* **67**(4), 1422–1433.

[*Received March 3, 2015; revised October 14, 2015; accepted for publication October 15, 2015*]