Biometrics 64, 1126–1136
December 2008
DOI: 10.1111/j.1541-0420.2008.01009.x

Patient-Specific Dose Finding Based on Bivariate Outcomes and Covariates

Peter F. Thall,[1,*] Hoang Q. Nguyen,[1] and Elihu H. Estey[2]

Departments of [1]Biostatistics and [2]Leukemia, The University of Texas,
M.D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, Texas 77030, U.S.A.
[*]*email:* rex@mdanderson.org

Summary. A Bayesian sequential dose-finding procedure based on bivariate (efficacy, toxicity) outcomes that accounts for patient covariates and dose-covariate interactions is presented. Historical data are used to obtain an informative prior on covariate main effects, with uninformative priors assumed for all dose effect parameters. Elicited limits on the probabilities of efficacy and toxicity for each of a representative set of covariate vectors are used to construct bounding functions that determine the acceptability of each dose for each patient. Elicited outcome probability pairs that are equally desirable for a reference patient are used to define two different posterior criteria, either of which may be used to select an optimal covariate-specific dose for each patient. Because the dose selection criteria are covariate specific, different patients may receive different doses at the same point in the trial, and the set of eligible patients may change adaptively during the trial. The method is illustrated by a dose-finding trial in acute leukemia, including a simulation study.

Key words: Adaptive design; Bayesian design; Biologic agents; Individualized treatment; Phase I clinical trial; Phase I/II clinical trial.

1. Introduction

Two useful extensions of methods for sequentially adaptive dose finding based on toxicity (T) in phase I clinical trials have been proposed in recent years. The first extension accounts for the effects of patient prognostic covariates, $\mathbf{Z} = (Z_1, \ldots, Z_q)$, as well as dose, x, on the probability of toxicity, $\pi_T(x, \mathbf{Z})$. This provides a basis for "patient-specific" dosing (Babb and Rogatko, 2001), also called "individualized" dosing (Ratain et al., 1996; Cheng et al., 2004), wherein the dose is chosen for each patient based on his/her $\mathbf{Z}$ vector (Mick and Ratain, 1993; Wijesinha and Piantadosi, 1995; Piantadosi and Liu, 1996; O'Quigley, Shen, and Gamst, 1999). Methods for finding tolerable doses within ordinal prognostic subgroups have been proposed by O'Quigley and Paoletti (2003), Yuan and Chappell (2004), and Ivanova and Wang (2006). For example, accounting for $\mathbf{Z}$ addresses the well-known problem that a patient's risk of toxicity from a cytotoxic agent increases with the patient's age (Estey et al., 1989). The second extension utilizes efficacy (E) as well as T to choose doses. A wide variety of approaches to this problem have been proposed (O'Quigley, Hughes, and Fenton, 2001; Braun, 2002; Ivanova, 2003; Thall and Cook, 2004; Bekele and Shen, 2005). The general objective is to choose x to obtain a desirably large value of $\pi_E(x)$ while also controlling $\pi_T(x)$. Trials using such methods are often called "phase I/II" because they combine elements of conventional phase I and phase II trials.

In this article, we present a new family of methods that combine these two extensions by basing dose finding on both E and T while also accounting for each patient's covariates. Our models and dose-finding procedures extend the methodology of Thall and Cook (2004) by accounting for covariate effects. As we will show, constructing a method that does this reliably involves much more than simply including $\mathbf{Z}$ in the underlying regression model.

Let $\pi_k(x, \mathbf{Z}, \boldsymbol{\theta})$ denote the marginal probability of outcome $k = E$ or T for a patient with covariates $\mathbf{Z}$ treated with dose x, where $\boldsymbol{\theta}$ is the model parameter vector. Our approach focuses on the probability pair, $\boldsymbol{\pi}(x, \mathbf{Z}, \boldsymbol{\theta}) = \{\pi_E(x, \mathbf{Z}, \boldsymbol{\theta}), \pi_T(x, \mathbf{Z}, \boldsymbol{\theta})\}$, which we will denote by $\boldsymbol{\pi}$ for brevity when no meaning is lost. We require an informative prior on covariate effect parameters, obtained from a preliminary fit of historical data. The physician must provide a lower limit on π_E and an upper limit on π_T for each of a set of covariate vectors ranging over the domain of $\mathbf{Z}$. These limits are used to construct bounding functions that vary with $\mathbf{Z}$, one function for π_E, and one for π_T, which together determine the acceptability of each x for each patient. For a reference $\mathbf{Z}^*$, the physician also must specify several equally desirable target values of $\boldsymbol{\pi}$ that are used to construct two different types of posterior criteria, either of which may be used to compare doses.

Because each patient's selected dose depends on his/her $\mathbf{Z}$ vector, different patients may receive different doses at the same point in the trial, and the trial's entry criteria may change dynamically as the data accumulate. For example, favorable interim outcomes may expand the entry criteria to include patients for whom entry was not permitted initially. On completion of the trial, rather than choosing one dose for

© 2008, The International Biometric Society

all patients, the method uses the trial's final data to select covariate-specific doses for future patients.

In Section 2, we present a family of regression models and explain how priors may be established. In Section 3, we define the criteria for determining the acceptability and desirability of each x for each $\mathbf{Z}$, and provide rules for trial conduct. Numerical methods are discussed in Section 4. In Section 5, we describe application of the method to an acute leukemia trial, including a simulation study, and we close with a discussion in Section 6.

2. Probability Models

2.1 *Bivariate Regression Models*

We focus on the case of bivariate binary outcomes $\mathbf{Y} = (Y_E, Y_T)$, where Y_k is the indicator of the event $k = E, T$. Denoting suitably normalized doses of the experimental agent by $x_1 < x_2 < \cdots < x_J$, the data from the first n patients in the trial take the form $\mathcal{D}_n = \{(\mathbf{Y}_i, \mathbf{Z}_i, x_{[i]}), i = 1, \ldots, n\}$, where $x_{[i]}$ denotes the ith patient's dose. Similarly, we denote the data from n_H historical patients by $\mathcal{H} = \{(\mathbf{Y}_i, \mathbf{Z}_i, \tau_{[i]}), i = 1, \ldots, n_H\}$, where $\{\tau_1, \ldots, \tau_m\}$ are the historical treatments and $\tau_{[i]}$ denotes the ith patient's treatment. For convenience, we denote either a given dose or historical treatment by unsubscripted τ, so that the joint outcome probabilities of a patient with covariates $\mathbf{Z}$ treated with τ are $\pi_{a,b}(\tau, \mathbf{Z}, \boldsymbol{\theta}) = \Pr(Y_E = a, Y_T = b \mid \tau, \mathbf{Z}, \boldsymbol{\theta})$, for $a, b \in \{0, 1\}$, and the marginal probabilities are $\pi_E(\tau, \mathbf{Z}, \boldsymbol{\theta}) = \Pr(Y_E = 1 \mid \tau, \mathbf{Z}, \boldsymbol{\theta}) = \pi_{1,1}(\tau, \mathbf{Z}, \boldsymbol{\theta}) + \pi_{1,0}(\tau, \mathbf{Z}, \boldsymbol{\theta})$ and $\pi_T(\tau, \mathbf{Z}, \boldsymbol{\theta}) = \Pr(Y_T = 1 \mid \tau, \mathbf{Z}, \boldsymbol{\theta}) = \pi_{1,1}(\tau, \mathbf{Z}, \boldsymbol{\theta}) + \pi_{0,1}(\tau, \mathbf{Z}, \boldsymbol{\theta})$.

Denote the linear terms $\eta_k = \mathrm{g}(\pi_k)$ for $k = E, T$, where g is a suitable link function. Our modeling strategy will be to determine the $\pi_{a,b}$'s in terms of the marginals $\pi_E = g^{-1}(\eta_E)$ and $\pi_T = g^{-1}(\eta_T)$ and one association parameter, ψ, so that

$$\pi_{a,b} = \pi_{a,b}(\pi_E, \pi_T, \psi) = \pi_{a,b}\{g^{-1}(\eta_E), \quad g^{-1}(\eta_T), \psi\}. \quad (1)$$

There are many models of this form, including the Gumbel (cf. Prentice, 1988),

$$\pi_{a,b} = \pi_E^a(1-\pi_E)^{1-a}\pi_T^b(1-\pi_T)^{1-b} + (-1)^{a+b}\psi\pi_E(1-\pi_E)\pi_T(1-\pi_T), \quad a, b = 0, 1, \quad (2)$$

where $-1 \leqslant \psi \leqslant 1$, which was used by Thall and Cook (2004) in the simpler case without $\mathbf{Z}$. Copulas (cf. Joe, 1997) provide a general class of models for constructing bivariate distributions with given marginals. A tractable model is given by the Gaussian copula

$$C_\psi(u, v) = \Phi_\psi(\Phi^{-1}(u), \Phi^{-1}(v)), \quad 0 \leqslant u, v \leqslant 1, \quad (3)$$

where Φ_ψ denotes the bivariate standard normal cumulative distribution function (cdf) with correlation ψ and unsubscripted Φ denotes the univariate standard normal cdf. In the present setting, (3) gives

$$\pi_{0,0} = \Phi_\psi(\Phi^{-1}(1-\pi_E), \Phi^{-1}(1-\pi_T)), \quad (4)$$

which determines the remaining $\pi_{a,b}$'s to be $\pi_{1,0} = 1 - \pi_E - \pi_{0,0}$, $\pi_{0,1} = 1 - \pi_T - \pi_{0,0}$, and $\pi_{1,1} = \pi_E + \pi_T + \pi_{0,0} - 1$. Under the probit link $\pi_k = \Phi(\eta_k)$, the model (4) takes the simpler form $\pi_{0,0} = \Phi_\psi(-\eta_E, -\eta_T)$. In practice, the choice of a particular form for (1) is motivated by models giving a good fit to $\mathcal{H}$ and the need for tractability, because the model must be fit many times to simulate and conduct the trial.

Given a particular bivariate model for $\pi_{a,b}\{g^{-1}(\eta_E), g^{-1}(\eta_T), \psi\}$, we focus on parametric forms for $\eta_E(\tau, \mathbf{Z}, \boldsymbol{\theta})$ and $\eta_T(\tau, \mathbf{Z}, \boldsymbol{\theta})$ that provide a basis for using historical data about covariate effects while also accounting for the joint effects of x and $\mathbf{Z}$ on π_E and π_T based on the current data, $\mathcal{D}_n$, from the trial. To do this, we assume the general linear form

$$\eta_k(\tau, \mathbf{Z}, \boldsymbol{\theta}) = \boldsymbol{\beta}_k\mathbf{Z} + \sum_{j=1}^{m}(\mu_{k,j} + \boldsymbol{\xi}_{k,j}\mathbf{Z})\mathbf{1}(\tau = \tau_j) + \{f_k(x, \boldsymbol{\alpha}_k) + \boldsymbol{\gamma}_k\mathbf{Z}\}\mathbf{1}(\tau = x), \quad (5)$$

for $k = E, T$, where $\mathbf{1}(A)$ is the indicator of the event A, linear combinations of vectors are denoted by juxtaposition, for example, $\boldsymbol{\beta}_k\mathbf{Z} = \beta_{k,1}Z_1 + \cdots + \beta_{k,q}Z_q$, and $f_E(x, \boldsymbol{\alpha}_E)$ and $f_T(x, \boldsymbol{\alpha}_T)$ characterize the main dose effects on π_E and π_T. For each k, the covariate main effects are $\boldsymbol{\beta}_k = (\beta_{k,1}, \ldots, \beta_{k,q})$, interactions between $\mathbf{Z}$ and historical treatment τ_j are $\boldsymbol{\xi}_{k,j} = (\xi_{k,j,1}, \ldots, \xi_{k,j,q})$ and, similarly, $\boldsymbol{\gamma}_k = (\gamma_{k,1}, \ldots, \gamma_{k,q})$ are dose–covariate interactions. For the historical data, the general expression (5) takes the form

$$\eta_k(\tau_j, \mathbf{Z}, \boldsymbol{\theta}) = \mu_{k,j} + \boldsymbol{\beta}_k\mathbf{Z} + \boldsymbol{\xi}_{k,j}\mathbf{Z}, \quad \text{for } j = 1, \ldots, m \text{ and } k = E, T, \quad (6)$$

where $\boldsymbol{\mu}_k = (\mu_{k,1}, \ldots, \mu_{k,m})$ are the historical main treatment effects. For the data obtained during the trial, (5) takes the form

$$\eta_k(x, \mathbf{Z}, \boldsymbol{\theta}) = f_k(x, \boldsymbol{\alpha}_k) + \boldsymbol{\beta}_k\mathbf{Z} + x\boldsymbol{\gamma}_k\mathbf{Z}, \quad \text{for } k = E, T. \quad (7)$$

The particular forms of f_E and f_T should be chosen to reflect the application at hand. For a trial of a cytotoxic agent, $f_k(x, \boldsymbol{\alpha}_k) = \alpha_{k,0} + \alpha_{k,1}x$ with $\Pr(\alpha_{k,1} > 0) = 1$ yields $\pi_k(x, \mathbf{Z}, \boldsymbol{\theta})$ increasing in x. The more flexible quadratic function $f_k(x, \boldsymbol{\alpha}_k) = \alpha_{k,0} + \alpha_{k,1}x + \alpha_{k,2}x^2$ allows π_k to be nonmonotone in x, which may be appropriate for biologic agents, because with increasing dose $\pi_E(x, \mathbf{Z}, \boldsymbol{\theta})$ may reach a plateau and possibly decrease thereafter. A nonmonotone $f_T(x, \boldsymbol{\alpha}_T)$ also may be needed, depending on how toxicity is defined, because new agents sometimes have effects not predicted by preclinical data. For example, in a trial of an antigraft-versus-host disease agent in allogeneic stem cell transplantation, if "toxicity" includes systemic infection ("sepsis"), and if higher doses of the agent help resolve sepsis, then $\pi_T(x, \mathbf{Z}, \boldsymbol{\theta})$ may decrease with x.

2.2 *Establishing Priors*

Denoting $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_E, \boldsymbol{\alpha}_T)$, $\boldsymbol{\mu} = (\boldsymbol{\mu}_E, \boldsymbol{\mu}_T)$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_E, \boldsymbol{\beta}_T)$, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_E, \boldsymbol{\gamma}_T)$, and $\boldsymbol{\xi} = (\boldsymbol{\xi}_{E,1}, \boldsymbol{\xi}_{T,1}, \ldots, \boldsymbol{\xi}_{E,k}, \boldsymbol{\xi}_{T,k})$, the vector of all model parameters is $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\beta}, \boldsymbol{\xi}, \psi, \boldsymbol{\alpha}, \boldsymbol{\gamma})$. A preliminary fit of $\mathcal{H}$ yields informative distributions on $\boldsymbol{\mu}$, $\boldsymbol{\beta}$, $\boldsymbol{\xi}$, and ψ. For the purpose of dose finding, $\boldsymbol{\mu}$ and $\boldsymbol{\xi}$ are nuisance parameters, but the informative marginal posterior $p(\boldsymbol{\beta}, \psi \mid \mathcal{H}) = \int p(\boldsymbol{\mu}, \boldsymbol{\beta}, \psi \mid \mathcal{H})\, d\boldsymbol{\mu}$ plays a key role in our method, because it is the prior on $(\boldsymbol{\beta}, \psi)$ at the start of the trial. In contrast, because $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ account for effects of the experimental agent, their priors should be noninformative. For prior means, we set $\mathrm{E}(\boldsymbol{\gamma}) = \mathbf{0}$, and obtain prior $\mathrm{E}(\boldsymbol{\alpha})$ by eliciting means of

$\pi_E(x_j, \mathbf{Z}^*, \boldsymbol{\theta})$ and $\pi_T(x_j, \mathbf{Z}^*, \boldsymbol{\theta})$ at two or more values of x_j and solving for $\mathrm{E}(\boldsymbol{\alpha})$. If the number of elicited means is larger than $\dim(\boldsymbol{\alpha})$, least squares (LS) may be used to solve for $\mathrm{E}(\boldsymbol{\alpha})$, as in Thall and Cook (2004).

To determine prior variances, the following two-step heuristic may be used. For the first step, assume that all entries of $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ have the same prior variance, say σ_0^2, and calibrate σ_0^2 to control the effective sample size (ESS) of the prior $p\{\pi_k(x_j, \mathbf{Z}^*, \boldsymbol{\theta})\}$ for all combinations of $k = E, T$, and x_j. A practical method for doing this is to match the first two moments of each $\pi_k(x_j, \mathbf{Z}^*, \boldsymbol{\theta})$ with a beta$(a_{k,j}, b_{k,j})$ and approximate the ESS of $p\{\pi_k(x_j, \mathbf{Z}^*, \boldsymbol{\theta})\}$ by

$$a_{k,j} + b_{k,j} = \frac{\mathrm{E}\{\pi_k(x_j, \mathbf{Z}^*, \boldsymbol{\theta})\}\{1 - E\{\pi_k(x_j, \mathbf{Z}^*, \boldsymbol{\theta})\}\}}{\mathrm{var}\{\pi_k(x_j, \mathbf{Z}^*, \boldsymbol{\theta})\}} - 1.$$

To reflect limited prior knowledge about dose effects and allow $p(\boldsymbol{\beta}, \psi \mid \mathcal{H})$ and the trial's data, rather than the prior on $(\boldsymbol{\alpha}, \boldsymbol{\gamma})$, to dominate the method's decisions, all $a_{k,j} + b_{k,j}$ should be small. One may use σ_0^2 as a tuning parameter to ensure this. However, a priori, the model also must control $\sigma^2(\alpha_{k,2}x^2) = \mathrm{var}(\alpha_{k,2}x^2)$ relative to $\sigma^2(\alpha_{k,1}x) = \mathrm{var}(\alpha_{k,1}x)$ for each k in order to avoid the quadratic effect $\alpha_{k,2}x^2$ being too large relative to the first-order effect $\alpha_{k,1}x$ and thus misrepresenting the true form of f_k. In terms of the standard deviations (SDs), first fixing $\sigma(\alpha_{k,1}) = \sigma_0$, one may examine the effects of values of the ratio $\sigma(\alpha_{k,2}x^2)/\sigma(\alpha_{k,1}x) = x\sigma(\alpha_{k,2})/\sigma(\alpha_{k,1}) = x\sigma(\alpha_{k,2})/\sigma_0$ in the range 0.1–1.0 on the design's operating characteristics (OCs), as described in Section 5.4, and choose a value of $\sigma(\alpha_{k,2})$ that ensures good OCs as well as reasonable prior ESS values.

3. Patient-Specific Dose Finding

At any interim point in the trial when a patient's dose must be chosen, the likelihood for the current trial data is the product

$$\mathcal{L}(\mathcal{D}_n \mid \boldsymbol{\theta}) = \prod_{i=1}^{n} \prod_{a=0}^{1} \prod_{b=0}^{1} \{\pi_{a,b}(x_{[i]}, \mathbf{Z}_i, \boldsymbol{\theta})\}^{\mathbf{1}\{\mathbf{Y}_i=(a,b)\}}.$$

During the trial, quantities computed from the successive posteriors

$$p(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\beta}, \psi \mid \mathcal{H} \cup \mathcal{D}_n) \propto \mathcal{L}(\mathcal{D}_n \mid \boldsymbol{\theta}) p(\boldsymbol{\alpha}, \boldsymbol{\gamma}) p(\boldsymbol{\beta}, \psi \mid \mathcal{H}) \quad (8)$$

are used as a basis for choosing doses adaptively. To construct a dose-finding method that accounts for both x and $\mathbf{Z}$, we first define two different kinds of criteria. The first criterion determines whether a given x is *acceptable* for given $\mathbf{Z}$. The second criterion quantifies the *desirability* of each $\boldsymbol{\pi} = (\pi_E, \pi_T)$ in terms of a subjective tradeoff between π_E and π_T, which provides a basis for comparing acceptable doses, and thus selecting the best dose for each patient. Both criteria require covariate-specific information elicited from the physician.

3.1 *Constructing Bounding Functions for Dose Acceptability*

First, specify a representative set of covariate vectors, $\{\mathbf{Z}^{(1)}, \ldots, \mathbf{Z}^{(K)}\}$, varying over the domain of $\mathbf{Z}$. For each $\mathbf{Z}^{(j)}$, elicit the smallest probability of efficacy, $\underline{\pi}_E^{(j)}$, and the largest probability of toxicity, $\bar{\pi}_T^{(j)}$, that the physician wishes to allow for a patient with covariates $\mathbf{Z}^{(j)}$. These $2K$ elicited limits are used to construct functions of $\mathbf{Z}$ that provide a lower bound on $\pi_E(x, \mathbf{Z}, \boldsymbol{\theta})$ and an upper bound on $\pi_T(x, \mathbf{Z}, \boldsymbol{\theta})$, which are used to determine whether each x is acceptable for a patient with covariates $\mathbf{Z}$. In general, $K > q$, but K must be small enough so that elicitation is practical. Denote $\zeta_k(\mathbf{Z}) = \mathrm{E}(\boldsymbol{\beta}_k \mid \mathcal{H})\mathbf{Z}$. To construct the bounding function for $\pi_E(x, \mathbf{Z}, \boldsymbol{\theta})$, we treat the K pairs $\{\zeta_E(\mathbf{Z}^{(1)}), \underline{\pi}_E^{(1)}\}, \ldots, \{\zeta_E(\mathbf{Z}^{(K)}), \underline{\pi}_E^{(K)}\}$ of estimated linear terms and corresponding elicited lower bounds on π_E like regression data, fit a curve to these points using LS with $\zeta_E(\mathbf{Z}^{(j)})$ the predictor and $\underline{\pi}_E^{(j)}$ the outcome variable, and denote the LS estimate by $\hat{\underline{\pi}}_E(\zeta_E)$. In practice, a linear function $\hat{\underline{\pi}}_E(\zeta_E) = a + b\zeta_E$ or quadratic $\hat{\underline{\pi}}_E(\zeta_E) = a + b\zeta_E + c\zeta_E^2$ works well. Other functions may be used, provided that $0 \leqslant \hat{\underline{\pi}}_E(\zeta_E) \leqslant 1$ on the domain of ζ_E values considered. We define the *efficacy lower bounding function* to be $\underline{\pi}_E(\mathbf{Z}) = \hat{\underline{\pi}}_E \circ \zeta_E(\mathbf{Z})$. For example, if the fitted LS curve is $\hat{\underline{\pi}}_E(\zeta_E) = \hat{a} + \hat{b}\zeta_E + \hat{c}\zeta_E^2$ then $\underline{\pi}_E(\mathbf{Z}) = \hat{a} + \hat{b}\zeta_E(\mathbf{Z}) + \hat{c}\{\zeta_E(\mathbf{Z})\}^2$. Similarly, we fit a curve to the K pairs $\{\zeta_T(\mathbf{Z}^{(1)}), \bar{\pi}_T^{(1)}\}, \ldots, \{\zeta_T(\mathbf{Z}^{(K)}), \bar{\pi}_T^{(K)}\}$, obtain an LS estimate $\hat{\bar{\pi}}_T(\zeta_T)$, and define the *toxicity upper bounding function* to be $\bar{\pi}_T(\mathbf{Z}) = \hat{\bar{\pi}}_T \circ \zeta_T(\mathbf{Z})$. In this way, we utilize $\mathrm{E}(\boldsymbol{\beta}_E \mid \mathcal{H})$ and $\mathrm{E}(\boldsymbol{\beta}_T \mid \mathcal{H})$ to define the composite functions

$$\mathbf{Z} \xrightarrow{\zeta_E} \mathcal{R}^1 \xrightarrow{\hat{\underline{\pi}}_E} [0,1] \quad \text{and} \quad \mathbf{Z} \xrightarrow{\zeta_T} \mathcal{R}^1 \xrightarrow{\hat{\bar{\pi}}_T} [0,1] \quad (9)$$

from the q-dimensional covariate space to the one-dimensional real domain $\mathcal{R}^1$ of $\boldsymbol{\beta}_E\mathbf{Z}$ and $\boldsymbol{\beta}_T\mathbf{Z}$, and then to the probability domain of π_E and π_T. When constructing these functions, it is useful to plot the scattergram of $(\zeta_E^{(1)}, \underline{\pi}_E^{(1)}), \ldots, (\zeta_E^{(K)}, \underline{\pi}_E^{(K)})$ with its fitted function $\hat{\underline{\pi}}_E(\zeta_E)$, and likewise plot $(\zeta_T^{(1)}, \bar{\pi}_T^{(1)}), \ldots, (\zeta_T^{(K)}, \bar{\pi}_T^{(K)})$ with $\hat{\bar{\pi}}_T(\zeta_T)$. Because the goal is to obtain bounds on $\pi_E(x, \mathbf{Z}, \boldsymbol{\theta})$ and $\pi_T(x, \mathbf{Z}, \boldsymbol{\theta})$ as $\mathbf{Z}$ is varied, which fulfils the physician's requirements, these plots allow the physician to adjust any $\underline{\pi}_E^{(j)}$ or $\bar{\pi}_T^{(j)}$ values if desired. Figure 1a and b illustrate the scattergrams and bounding functions for the acute myelogenous leukemia (AML) trial design, details of which will be provided in Section 5. The following definition uses $\underline{\pi}_E(\mathbf{Z})$ and $\bar{\pi}_T(\mathbf{Z})$ to provide criteria for deciding adaptively which doses may be given to each new patient.

DEFINITION 1: *Given current trial data $\mathcal{D}_n$ and fixed cutoffs p_T and p_E, the set $\mathcal{A}_n(\mathbf{Z})$ of acceptable doses for a patient with covariates $\mathbf{Z}$ consists of all $x \in \{x_1, \ldots, x_J\}$ such that*

$$\Pr\{\pi_E(x, \mathbf{Z}, \boldsymbol{\theta}) < \underline{\pi}_E(\mathbf{Z}) \mid \mathcal{D}_n \cup \mathcal{H}\} < p_E \quad (10)$$

and

$$\Pr\{\pi_T(x, \mathbf{Z}, \boldsymbol{\theta}) > \bar{\pi}_T(\mathbf{Z}) \mid \mathcal{D}_n \cup \mathcal{H}\} < p_T. \quad (11)$$

The cutoffs p_T and p_E should be calibrated to obtain good OCs, with values around 0.90 typically giving good designs. For a patient with covariates $\mathbf{Z}$, (10) says that x is not likely to be inefficacious. Similarly, (11) says that x is not likely to be too toxic. If $\mathcal{A}_n(\mathbf{Z}) = \phi$, then no dose is acceptable for that patient. If $\mathcal{A}_n(\mathbf{Z})$ consists of a single dose, then that dose is used by default. If $\mathcal{A}_n(\mathbf{Z})$ contains more than one dose, however, then a criterion for choosing a dose from $\mathcal{A}_n(\mathbf{Z})$ is needed.

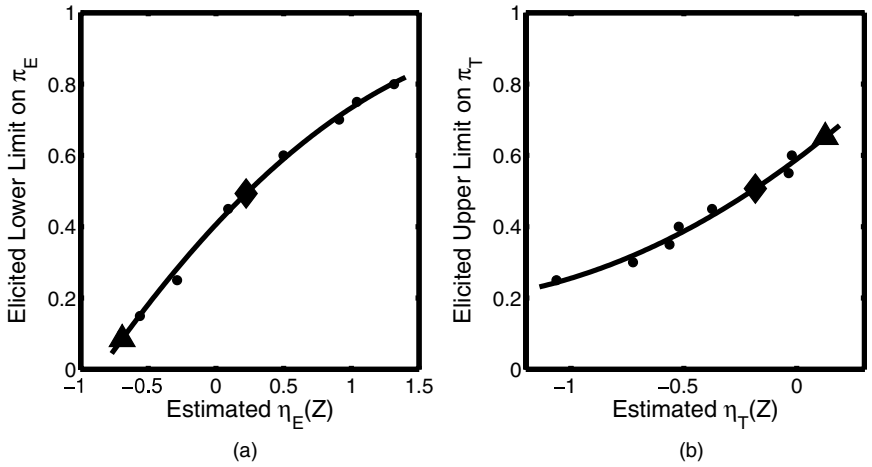

Figure 1. Bounding functions for determining dose acceptability as a function of $\mathbf{Z}$. Each circle represents a pair $(\zeta_E(\mathbf{Z}^{(j)}), \underline{\pi}_E^{(j)})$ in the left-side plot and a pair $(\zeta_T(\mathbf{Z}^{(j)}), \bar{\pi}_T^{(j)})$ in the right-side plot. In both figures the triangle corresponds to a 58-year-old-patient with poor cytogenetics, and the diamond corresponds to a 48-year-old-patient with intermediate cytogenetics.

3.2 *Dose Desirability*

To select one dose from $\mathcal{A}_n(\mathbf{Z})$, we will define criteria for comparing doses in terms of a *target contour*, $\mathcal{C}$, in the two-dimensional set of $\boldsymbol{\pi}$ values. One may construct $\mathcal{C}$ in several ways. A general method (cf. Thall, Wooten, and Shpall, 2006) is to elicit several target probability pairs $\boldsymbol{\pi}_1^*, \ldots, \boldsymbol{\pi}_c^*$ that the physician considers equally desirable for a reference patient with $\mathbf{Z} = \mathbf{Z}^*$ and fit a smooth curve to the elicited points so that, considering π_T as a function of π_E, $d\pi_T/d\pi_E > 0$, that is, this function is strictly increasing on the relevant domain of π_E values. One may define $\mathcal{C}$ to be the fitted curve. An alternative method for constructing $\mathcal{C}$ (Thall, Cook, and Estey, 2006) is to elicit exactly $c = 3$ target pairs, with $\boldsymbol{\pi}_1^*$ in the interior of $[0,1]^2$, $\boldsymbol{\pi}_2^* = (\pi_{E,2}^*, 0)$ with $\pi_{E,2}^*$ the smallest value of π_E that makes $\boldsymbol{\pi}_2^*$ as desirable as $\boldsymbol{\pi}_1^*$, and $\boldsymbol{\pi}_3^* = (1, \pi_{T,3}^*)$ with $\pi_{T,3}^*$ the largest value of π_T that makes $\boldsymbol{\pi}_3^*$ as desirable as $\boldsymbol{\pi}_1^*$ and $\boldsymbol{\pi}_2^*$. Under the admissibility conditions $\pi_{E,2}^* < \boldsymbol{\pi}_{E,1}^* < 1$ and $0 < \pi_{T,1}^* < \pi_{T,3}^*$, the L^p norm distance from $\boldsymbol{\pi}$ to the ideal point (1,0) with the axes scaled by $\pi_{E,2}^*$ and $\pi_{T,3}^*$ is

$$\|\boldsymbol{\pi} - (1,0)\|_p = \left\{ \left(\frac{\pi_E - 1}{\pi_{E,2}^* - 1} \right)^p + \left(\frac{\pi_T - 0}{\pi_{T,3}^* - 0} \right)^p \right\}^{1/p}.$$

Because, given d, there is a unique $p > 0$ such that $\mathcal{C}_d = \{\boldsymbol{\pi} : \|\boldsymbol{\pi} - (1,0)\|_p = d\}$ passes through $\boldsymbol{\pi}_1^*$, $\boldsymbol{\pi}_2^*$, and $\boldsymbol{\pi}_3^*$, one may define $\mathcal{C} = \mathcal{C}_d$.

Given $\mathcal{C}$ and trial data $\mathcal{D}_n$, the desirability of a dose x for a patient with covariates $\mathbf{Z}$ may be defined in several ways. We will consider two definitions. The first uses posterior probabilities of the set $\mathcal{P}_\mathcal{C} = \{\boldsymbol{\pi} : \pi_E \geqslant \pi'_E \text{ and } \pi_T \leqslant \pi'_T \ \exists\ \boldsymbol{\pi}' \in \mathcal{C}\}$ of $\boldsymbol{\pi}$ pairs at least as desirable as a pair on $\mathcal{C}$.

Definition 2: *Given observed data $\mathcal{D}_n$, the posterior probability (PP) desirability of dose x for a patient with covariates $\mathbf{Z}$ is* $\delta_n^{(PP)}(x, \mathbf{Z}) = \Pr\{\boldsymbol{\pi}(x, \mathbf{Z}, \boldsymbol{\theta}) \in \mathcal{P}_\mathcal{C} \mid \mathcal{H} \cup \mathcal{D}_n\}$.

For the second, alternative definition, given $\boldsymbol{\pi}$ we denote by $\boldsymbol{\pi}_\mathcal{C}$ the point where the straight line segment $L_{\boldsymbol{\pi}}$ in $[0,1]^2$ passing through $\boldsymbol{\pi}$ and the ideal point (1,0) intersects $\mathcal{C}$. Denoting Euclidean distance by $\|\cdot\|$, we define the *geometric desirability* of $\boldsymbol{\pi}$ to be

$$r_\mathcal{C}(\boldsymbol{\pi}) = \exp\left\{ -\frac{\|\boldsymbol{\pi} - (1,0)\|}{\|\boldsymbol{\pi}_\mathcal{C} - (1,0)\|} \right\}.$$

As $\boldsymbol{\pi}$ moves away from (1,0) along any $L_{\boldsymbol{\pi}}, r_\mathcal{C}(\boldsymbol{\pi})$ decreases from its maximum value $r_\mathcal{C}\{(1,0)\} = 1$, and $r_\mathcal{C}(\boldsymbol{\pi}) = e^{-1}$ for all $\boldsymbol{\pi} \in \mathcal{C}$. Thus, the geometric desirability $r_\mathcal{C}(\boldsymbol{\pi})$ determined by $\mathcal{C}$ may be used to compare $\boldsymbol{\pi}$ pairs. Our second definition evaluates this function at the posterior mean of $\boldsymbol{\pi}(x, \mathbf{Z}, \boldsymbol{\theta})$.

Definition 3: *Given observed data $\mathcal{D}_n$, the posterior mean (PM) desirability of dose x for a patient with covariates $\mathbf{Z}$ is* $\delta_n^{(PM)}(x, \mathbf{Z}) = r_\mathcal{C}(\mathrm{E}\{\boldsymbol{\pi}(x, \mathbf{Z}, \boldsymbol{\theta}) \mid \mathcal{H} \cup \mathcal{D}_n\})$.

Using either criterion $\delta_n^{(PP)}(x, \mathbf{Z})$ or $\delta_n^{(PM)}(x, \mathbf{Z})$, the best acceptable dose for a patient with covariates $\mathbf{Z}$ is the x maximizing the desirability. In general, the desirability of x given $\mathbf{Z}$ is obtained as a composite function

$$(x, \mathbf{Z}) \longrightarrow \boldsymbol{\pi}(x, \mathbf{Z}, \boldsymbol{\theta}) \longrightarrow \delta_n(x, \mathbf{Z}), \tag{12}$$

where the second function exploits the Bayesian model and geometry of $\mathcal{C}$, using either $r_\mathcal{C}$ or $\mathcal{P}_\mathcal{C}$, by averaging over $\boldsymbol{\theta}$.

3.3 *Trial Design and Conduct*

When planning the trial, the historical data first must be analyzed to choose a model and compute the posterior $p(\boldsymbol{\beta}, \psi \mid \mathcal{H})$, and the priors on $\boldsymbol{\alpha}$ and $\boldsymbol{\delta}$ must be determined. Additionally, the doses to be studied, maximum sample size, N, the set $\{\mathbf{Z}^{(1)}, \ldots, \mathbf{Z}^{(K)}\}$, of representative covariate vectors, and acceptability and desirability criteria all must be established. During the trial, when a new patient with covariates $\mathbf{Z}$ is enrolled, $\mathcal{A}_n(\mathbf{Z})$ is computed based on the current posterior and the following decision rules are applied. Let $\delta_n(x, \mathbf{Z})$ denote the desirability criterion being used.

1. If $\mathcal{A}_n(\mathbf{Z}) = \phi$, do not treat the patient on protocol.

2. If $\mathcal{A}_n(\mathbf{Z}) \neq \phi$, treat the patient using the dose x that maximizes $\delta_n(x, \mathbf{Z})$.
3. If $\mathcal{A}_n(\mathbf{Z}^{(j)}) = \phi$ for all $j = 1, \ldots, K$, stop the trial and declare the new agent unacceptable at any dose.
4. After the trial has been completed with final data $\mathcal{D}_N$, use the acceptability and desirability criteria based on $p(\boldsymbol{\theta} \mid \mathcal{H} \cup \mathcal{D}_N)$ to select doses for future patients.

If no dose is acceptable for a patient with covariates $\mathbf{Z}$, under rule (1) the patient must either be given a different treatment off protocol or may have such poor prognosis that no treatment is reasonable. This provides a formal basis for decisions that often must be made subjectively by the physician. Because the acceptability criteria (10) and (11) depend on $\mathbf{Z}$, x, and $\mathcal{D}_n$, whether a patient with a particular $\mathbf{Z}$ has any acceptable dose and thus may be enrolled may change adaptively over the course of the trial. Rule (3) is a combined futility and safety stopping rule, because if no dose is both acceptably safe and acceptably efficacious for any $\mathbf{Z}^{(1)}, \ldots, \mathbf{Z}^{(K)}$ in the representative covariate set, then the trial is stopped.

4. Numerical Methods

Markov chain Monte Carlo (MCMC) with Gibbs sampling (Robert and Cassella, 1999) was used to compute all posterior quantities. To generate each MCMC series of parameter vectors during the trial, for each update of the posterior an approximate posterior mode was first determined by sampling iteratively around the mode of the previous posterior. MCMC convergence was monitored by comparing the Monte Carlo standard error (MCSE) to the SDs of $\pi_E(x, \mathbf{Z}^*, \boldsymbol{\theta})$ and of $\pi_T(x, \mathbf{Z}^*, \boldsymbol{\theta})$ evaluated at the highest and lowest doses, with sufficiently small values of MCSE/sd$\{\pi_E(x, \mathbf{Z}^*, \boldsymbol{\theta})\}$ and MCSE/sd$\{\pi_T(x, \mathbf{Z}^*, \boldsymbol{\theta})\}$ indicating convergence.

5. Application

5.1 *A Dose-Finding Trial in Acute Leukemia*

We illustrate the method with a trial combining an experimental chemo-protective agent (CPA) with the standard drugs idarubicin (IDA) and cytosine arabinoside (ara-C) in patients under age 60 with untreated AML. The therapeutic goal is to produce a complete remission (CR), defined by recovery of circulating blood cells and immature cells in the bone marrow to normal levels, a necessary condition for survival. Although the CR rate may be increased by higher doses of IDA, this also increases the risk of mouth ulcers ("mucositis") and desquamatation of the intestinal lining, which increase the risk of sepsis, severe diarrhea, and death. The CPA is postulated to decrease the risk of IDA-induced mucositis and diarrhea, which would allow higher IDA doses and thus, hopefully, yield higher CR rates. In this trial, the ara-C and CPA doses are fixed at 1.5 grams/m^2 of ara-C and 2.4 mg/kg of CPA, both given daily on days 1–4. Using the covariate-specific dose selection method, each patient will be given one of five IDA doses: 12 (the standard dose), 15, 18, 21, or 24 mg/m^2 daily for 3 days. Efficacy is defined as the event that the patient is alive and in CR 42 days (6 weeks) after beginning treatment. Toxicity is defined as severe (NIH grade 3 or 4) mucositis, diarrhea, pneumonia, or sepsis within 42 days. For example, a patient who dies before day 42 has outcome $(Y_E, Y_T) = (0,1)$, whereas a patient who achieves the efficacy outcome but suffers a nonfatal toxicity has $(Y_E, Y_T) = (1,1)$. For a patient with no toxicity, $(Y_E, Y_T) = (1,0)$ if efficacy is achieved and $(Y_E, Y_T) = (0,0)$ if not.

It is well known that the CR rate with chemotherapy in AML varies greatly with the patient's age and type of cytogenetic abnormality found in the leukemia cells. Three cytogenetic ("Cyto") prognostic subgroups can be distinguished: an inversion of chromosome 16 or a translocation between chromosomes 8 and 21 (Cyto = good), abnormalities of chromosomes 5 and/or 7 (Cyto = poor), with all other cytogenetic findings comprising an intermediate group. To account for the effects of these covariates, we defined $Z_1 = 0.01(\text{age} - 45)$ for numerical stability, with $Z_2 = \mathbf{1}(\text{Cyto} = \text{good})$ and $Z_3 = \mathbf{1}(\text{Cyto} = \text{poor})$, so that $q = 3$.

5.2 *Analysis of Historical Data*

We began by analyzing historical data from 693 untreated AML patients with age < 60, given chemotherapy in the Department of Leukemia at M.D. Anderson Cancer Center (MDACC) during the period from January 2000 to December 2004. These patients received one of three different historical treatments: ara-C + IDA ± other drugs, ara-C + fludarabine ± other drugs, and ara-C ± other drugs not including either IDA or fludarabine. The definitions of E and T in the historical data are identical to those given above for the planned trial. Frequencies of the four elementary outcomes, within each of the nine representative prognostic subgroups, are given in Table 1. We fit a total of nine models to the historical data, obtained from three different bivariate distributions $\pi_{a,b}\{\pi_E, \pi_T, \psi\}$ and three different forms for the linear terms $\eta_k(\tau_j, \mathbf{Z}, \boldsymbol{\theta})$. The distributions were a Gaussian copula with probit link, a Gumbel model with logit link, and a Gumbel with complementary log-log (c-l-l) link. The three linear terms were (1) $\mu_{k,j} + \boldsymbol{\beta}_k\mathbf{Z} + \boldsymbol{\xi}_{k,j}\mathbf{Z}$, which includes historical treatment–covariate interactions, (2) $\mu_{k,j} + \boldsymbol{\beta}_k\mathbf{Z}$, which assumes homogeneous covariate effects, and (3) $\mu_k + \boldsymbol{\beta}_k\mathbf{Z}$, which assumes both homogeneous covariate effects and no difference between historical treatments. For all fits of the historical data, we assumed the following noninformative prior. Because $-1 \leqslant \psi \leqslant 1$ for all models, we assumed $(\psi + 1)/2$ followed a beta(1,1) prior, and all other model parameters followed independent normal priors with mean 0 and variance 144. To select a model from these

Table 1
Frequency (row %) of each elementary outcome for each of nine prognostic subgroups in the historical data

		(Y_E, Y_T)				
Age	Cyto	(1,0)	(1,1)	(0,0)	(0,1)	Total
14–41	Good	42 (81)	6 (12)	4 (8)	0 (0)	52
14–41	Intermediate	63 (46)	30 (22)	27 (20)	17 (12)	137
14–41	Poor	7 (25)	4 (14)	11 (39)	6 (21)	28
42–52	Good	16 (59)	6 (22)	3 (11)	2 (7)	27
42–52	Intermediate	64 (37)	35 (20)	36 (21)	38 (22)	173
42–52	Poor	7 (15)	6 (13)	16 (33)	19 (40)	48
53–59	Good	4 (44)	3 (33)	2 (22)	0 (0)	9
53–59	Intermediate	45 (30)	40 (26)	36 (24)	31 (20)	152
53–59	Poor	9 (13)	8 (12)	23 (34)	27 (40)	67

Table 2

Fit of the historical data under the bivariate Gaussian copula model with probit link. Cytogenetic abnormality (Cyto) is classified as prognostically good if I-16 or t(8:21), poor if −5/−7, and otherwise intermediate.

	Posterior values based on the historical data			
	Effects on Pr(efficacy)		Effects on Pr(toxicity)	
Variable	Mean_{sd}	$\Pr(\beta > 0)$	Mean_{sd}	$\Pr(\beta > 0)$
Intercept	$0.262_{.060}$	–	$-0.232_{.059}$	–
0.01 (Age = 45)	$-1.309_{.461}$	0.002	$1.619_{.458}$	1.000
Cyto = good	$0.818_{.186}$	1.000	$-0.542_{.170}$	0.000
Cyto = poor	$-0.785_{.127}$	0.000	$0.148_{.121}$	0.889
Association	$-0.097_{.063}$	0.062		

nine possibilities, we used the Akaike (AIK), Bayes (BIC) and Deviance information criteria (DIC; Spiegelhalter et al., 2002). Denoting the maximum likelihood estimator (MLE) by $\hat{\boldsymbol{\theta}}_{MLE}$, number of model parameters by p, and sample size by n, these are AIK $= -2\log\{\mathcal{L}(\mathcal{H} \mid \hat{\boldsymbol{\theta}}_{MLE})\} + 2p$, BIC $= -2\log\{\mathcal{L}(\mathcal{H} \mid \hat{\boldsymbol{\theta}}_{MLE})\} + p\,\log(n)$, and DIC $= p_D + \bar{D}$, where $p_D = \bar{D} - D(\bar{\boldsymbol{\theta}})$ is the effective number of parameters, $\bar{D}$ is the PM deviance, and $D(\bar{\boldsymbol{\theta}})$ is the deviance evaluated at the PM $\bar{\boldsymbol{\theta}} = \text{E}(\boldsymbol{\theta} \mid \mathcal{H})$. For each of the three bivariate distributions, each of the three criteria was by far the smallest for the completely homogeneous linear terms $\eta_k(\tau_j, \mathbf{Z}, \boldsymbol{\theta}) = \mu_k + \boldsymbol{\beta}_k\mathbf{Z}$, indicating that there was no substantive difference among the historical treatment effects on either π_E or π_T. Next comparing the three bivariate models fit with this linear term, all three criteria were the smallest for the Gaussian copula with probit link. Numerical details of these analyses are given in Web Appendix Supplementary Table 1. Consequently, we assumed this model to construct a trial design. The fitted model is summarized in Table 2, which shows the strong effects of age and cytogenetics on the outcomes of AML patients undergoing chemotherapy. For each of the nine representative $\mathbf{Z}$ vectors, Table 3 gives the PMs and SDs of the $\pi_k(\mathbf{Z}, \boldsymbol{\theta})$'s and $\eta_k(\mathbf{Z}, \boldsymbol{\theta})$'s, and the corresponding desirabilities $r_{\mathcal{C}}(\boldsymbol{\pi})$ evaluated at the PMs $\boldsymbol{\pi} = E\{\boldsymbol{\pi}(\mathbf{Z}^{(j)}, \boldsymbol{\theta}) \mid \mathcal{H}\}$ for $j = 1, \ldots, K$. The fit of this data set illustrates how the values of $r_{\mathcal{C}}(\boldsymbol{\pi})$ may be used as a one-dimensional index for comparing the representative prognostic covariates $\{\mathbf{Z}^{(1)}, \ldots, \mathbf{Z}^{(K)}\}$ in terms of their associated historical PM-CR and toxicity probabilities.

5.3 *Trial Design*

A maximum of 60 patients will be enrolled in the trial. The target contour $\mathcal{C}$ was obtained by fitting a quadratic to the four target pairs $(0.75, 0.50)$, $(0.30, 0)$, $(1, 0.60)$, $(0.40, 0.12)$ corresponding to a patient with $\mathbf{Z}^*$ $=(48, \text{intermediate})$. These targets were considered equally desirable improvements over the historical mean $\text{E}\{\boldsymbol{\pi}(\mathbf{Z}^*, \boldsymbol{\theta}) \mid \mathcal{H}\} = (0.59, 0.43)$. This yielded the target curve $\pi_T = -\,0.5605 + 2.1226\pi_E - 0.9591\pi_E^2$, and $\mathcal{C}$ was defined to be the set of $\boldsymbol{\pi}$ satisfying this equation for $0.30 \leqslant \pi_E \leqslant 1$, where the fitted curve is strictly increasing. Based on this $\mathcal{C}$, the PM criterion $\delta_n^{(PM)}(x, \mathbf{Z})$ will be used to quantify desirability. The trial acceptability bounding functions were constructed using the elicited covariate-specific acceptability bounds given in the last two columns of Table 3, and are given by the quadratics

$$\underline{\pi}_E(\mathbf{Z}) = 0.4063 + 0.4078\zeta_E(\mathbf{Z}) - 0.0806\{\zeta_E(\mathbf{Z})\}^2$$

and

$$\bar{\pi}_T(\mathbf{Z}) = 0.5890 + 0.4739\zeta_T(\mathbf{Z}) + 0.1403\{\zeta_T(\mathbf{Z})\}^2,$$

where $\zeta_E(\mathbf{Z})$ and $\zeta_T(\mathbf{Z})$ are obtained from the fitted values given in Table 3. The bounding functions are plotted in Figure 1, where for illustration the bounds $\underline{\pi}_E(58, \text{poor}) = 0.10$ and $\bar{\pi}_T(58, \text{poor}) = 0.65$ for a 58-year-old patient with poor Cyto are denoted by a triangle, and the bounds $\underline{\pi}_E(48, \text{intermediate}) = 0.50$ and $\bar{\pi}_T(48, \text{intermediate}) = 0.50$ for a 48-year-old patient with intermediate Cyto are denoted by a diamond. Using these bounding functions, the acceptability criteria (10) and (11) will be applied with $p_E = p_T = 0.90$.

Based on the fits of the historical data, the assumed model underlying the method will be the Gaussian copula with probit link, and to ensure flexibility both dose-outcome functions will be assumed to follow a quadratic form,

Table 3

Posterior mean outcome probabilities and linear terms for E and T under the bivariate Gaussian copula model with probit link, and desirability values evaluated at the posterior means of $\boldsymbol{\pi}(\mathbf{Z}, \boldsymbol{\theta})$, based on the historical data, and the corresponding elicited bounds $\underline{\pi}_E(\mathbf{Z})$ on $\pi_E(\mathbf{Z}, \boldsymbol{\theta})$ and $\bar{\pi}_T(\mathbf{Z})$ on $\pi_T(\mathbf{Z}, \boldsymbol{\theta})$, for each of nine representative covariate vectors

		Posterior mean_{sd}					Elicited bounds	
Age	Cyto	$\pi_E(\mathbf{Z}, \boldsymbol{\theta})_{\text{sd}}$	$\pi_T(\mathbf{Z}, \boldsymbol{\theta})_{\text{sd}}$	$\eta_E(\mathbf{Z}, \boldsymbol{\theta})_{\text{sd}}$	$\eta_T(\mathbf{Z}, \boldsymbol{\theta})_{\text{sd}}$	$r_{\mathcal{C}}(\boldsymbol{\pi})$	$\underline{\pi}_E(\mathbf{Z})$	$\bar{\pi}_T(\mathbf{Z})$
27	Good	$0.902_{.031}$	$0.147_{.038}$	$1.316_{.184}$	$-1.065_{.166}$	0.73	0.80	0.25
27	Intermediate	$0.690_{.037}$	$0.301_{.036}$	$0.497_{.105}$	$-0.523_{.105}$	0.45	0.60	0.40
27	Poor	$0.388_{.056}$	$0.355_{.053}$	$-0.288_{.147}$	$-0.375_{.144}$	0.28	0.25	0.45
48	Good	$0.847_{.041}$	$0.237_{.049}$	$1.041_{.178}$	$-0.725_{.161}$	0.60	0.75	0.30
48	Intermediate	$0.588_{.023}$	$0.427_{.024}$	$0.222_{.060}$	$-0.183_{.060}$	0.33	0.50	0.50
48	Poor	$0.288_{.038}$	$0.486_{.042}$	$-0.563_{.111}$	$-0.035_{.105}$	0.21	0.15	0.55
58	Good	$0.814_{.051}$	$0.290_{.060}$	$0.910_{.194}$	$-0.563_{.177}$	0.53	0.70	0.35
58	Intermediate	$0.536_{.032}$	$0.492_{.032}$	$0.091_{.082}$	$-0.021_{.081}$	0.29	0.45	0.60
58	Poor	$0.246_{.038}$	$0.550_{.045}$	$-0.693_{.121}$	$0.127_{.114}$	0.18	0.10	0.65

$f_k(x, \boldsymbol{\alpha}_k) = \alpha_{k,0} + \alpha_{k,1}x + \alpha_{k,2}x^2$, for $k = E, T$. The elicited prior means of $\boldsymbol{\pi}(x_j, \mathbf{Z}^*)$ for the five IDA dose levels were the historical mean $(0.59, 0.43)$ for the standard dose 12 mg/m^2, and the respective pairs for 15, 18, 21, 24 mg/m^2 were $(0.70, 0.50)$, $(0.75, 0.55)$, $(0.85, 0.70)$, $(0.90, 0.90)$. Using the normalized values $x = -2, -1, 0, 1, 2$ for the five IDA doses in the model, a LS fit of the linear terms yielded prior means (0.773, 0.262, 0.0077) for $(\alpha_{E,0}, \alpha_{E,1}, \alpha_{E,2})$ and (0.098, 0.344, 0.1025) for $(\alpha_{T,0}, \alpha_{T,1}, \alpha_{T,2})$. The prior means of all interaction parameters $\boldsymbol{\gamma}_E$ and $\boldsymbol{\gamma}_T$ were set equal to 0. Using the methods described in Section 2 with a maximum ESS value of 1.3, each parameter in $\boldsymbol{\alpha}_E$, $\boldsymbol{\alpha}_T$, $\boldsymbol{\gamma}_E$, and $\boldsymbol{\gamma}_T$ was given prior SD 1.33, except for the quadratic coefficients $\alpha_{E,2}$ and $\alpha_{T,2}$, which were set equal to $0.2 \times 1.33 = 0.266$.

5.4 *Simulation Study*

We studied the design's behavior under a wide variety of dose-outcome scenarios. Each scenario was specified in terms of fixed values of the 10 marginal probabilities $\{\pi_E(x_j, \mathbf{Z}^*), \pi_T(x_j, \mathbf{Z}^*), j = 1, \ldots, 5\}$ at reference $\mathbf{Z}^*$ and fixed interaction parameters $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_E, \boldsymbol{\gamma}_T)$. In each simulated trial, each patient's $\mathbf{Z}$ vector was sampled from $\mathcal{H}$. To obtain fixed values of $\pi_{a,b}(x_j, \mathbf{Z})$ for each sampled $\mathbf{Z}$ to use as a basis for simulating the outcome $[\mathbf{Y} \mid \mathbf{Z}]$, for each k and x_j we solved the equation $g\{\pi_k(x_j, \mathbf{Z}^*)\} = \eta_k(x_j, \mathbf{Z}^*) = \boldsymbol{\beta}_k\mathbf{Z}^* + f_k(x_j) + x_j\,\boldsymbol{\gamma}_k\mathbf{Z}^*$ for $f_k(x_j)$, substituting the historical means $\mathrm{E}(\boldsymbol{\beta}_k \mid \mathcal{H})$ for $\boldsymbol{\beta}_k$. This in turn yielded the fixed marginals $\pi_k(x_j, \mathbf{Z}) = g^{-1}\{\boldsymbol{\beta}_k\mathbf{Z} + f_k(x_j) + x_j\,\boldsymbol{\gamma}_k\mathbf{Z}\}$ for the given $\mathbf{Z}$, and the bivariate probabilities $\pi_{a,b}(x_j, \mathbf{Z})$ were then obtained under the Gaussian copula with probit link by substituting the historical mean $\mathrm{E}(\psi \mid \mathcal{H}) = -0.097$ for ψ.

To assess the comparative advantage of accounting for $\mathbf{Z}$ using our proposed model and method, under each scenario we also simulated the trial using (1) our proposed method but assuming the reduced model with no dose–covariate interactions, $\boldsymbol{\gamma} \equiv \mathbf{0}$, and (2) a greatly simplified version of our method that ignores $\mathbf{Z}$ completely. For this method, the model is reduced by assuming $\boldsymbol{\beta} \equiv \mathbf{0}$ and $\boldsymbol{\gamma} \equiv \mathbf{0}$ so that $\eta_k(x, \mathbf{Z}, \boldsymbol{\theta}) = f_k(x, \boldsymbol{\alpha}_k)$, using the same acceptability bounds $\underline{\pi}_E(\mathbf{Z}^*)$ and $\bar{\pi}_T(\mathbf{Z}^*)$ for all patients, so that $\mathcal{A}_n(\mathbf{Z}^*)$ is used regardless of the patient's $\mathbf{Z}$ vector and, similarly, the dose desirability criterion is based on the simplified model $\boldsymbol{\pi}(x, \boldsymbol{\theta})$ that ignores $\mathbf{Z}$. For brevity, we denote these three methods by $\mathbf{Z}$ − INT, $\mathbf{Z}$ − No INT, and No $\mathbf{Z}$, respectively.

To evaluate how well each method selected x_j for a patient with covariates $\mathbf{Z}$ under each scenario, in addition to tabulating selection probabilities for each x_j and $\mathbf{Z}^{(r)}$, we computed the following statistic. Given $\mathbf{Z}$ and D_n, unless $\mathcal{A}_n(\mathbf{Z}) = \phi$, let $x(\mathbf{Z}, D_n)$ denote the covariate-specific dose selected by the method. One may compute the true value of $\boldsymbol{\pi}(x(\mathbf{Z}, D_n), \mathbf{Z})$ under the scenario and thus the geometric desirability $r_{\mathcal{C}}\{\boldsymbol{\pi}(x(\mathbf{Z}, D_n), \mathbf{Z})\}$ at the selected dose for the given $\mathbf{Z}$. Thus, a simple criterion for comparing $x(\mathbf{Z}, D_n)$ to the best possible dose that could have been selected for a patient with covariates $\mathbf{Z}$ is

$$\rho(\mathbf{Z}) = \frac{r_{\mathcal{C}}\{\boldsymbol{\pi}(x(\mathbf{Z}, D_n), \mathbf{Z})\}}{\max\limits_{j=1,\ldots,5} r_{\mathcal{C}}\{\boldsymbol{\pi}(x_j, \mathbf{Z})\}}, \tag{13}$$

the ratio of the geometric desirability of the selected dose to maximum desirability that could have been achieved. This takes on values between 0 and 1, with $\rho(\mathbf{Z}) = 1$ if the best possible dose was selected.

Simulation results for patients with $\mathbf{Z} =$ (27, good), (48, intermediate), and (58, poor) are summarized in Table 4. Under simulation scenario 1, the geometric desirability $r_{\mathcal{C}}(\pi_E, \pi_T)$ increases with dose for each $\mathbf{Z}$, and for each dose the numerical values of $r_{\mathcal{C}}(\pi_E, \pi_T)$ change dramatically with $\mathbf{Z}$, similar to what was seen in the historical data. Scenario 2 is obtained from scenario 1 by adding dose-covariate interactions, $\gamma_{E,2} = \gamma_{E,\text{good}} = 0.6$, $\gamma_{E,3} = \gamma_{E,\text{poor}} = -0.8$, $\gamma_{T,2} = \gamma_{T,\text{good}} = -0.6$, and $\gamma_{T,3} = \gamma_{T,\text{poor}} = 0.8$, with no interactions between dose and patient age. This says that, with higher x, the interactions lead to larger $\pi_E(x, \boldsymbol{Z}, \boldsymbol{\theta})$ and smaller $\pi_T(x, \boldsymbol{Z}, \boldsymbol{\theta})$ for patients with good cytogenetics, whereas the interactions have the opposite effect on patients with poor cytogenetics. The particular numerical values of the interaction parameters were chosen to produce effects of reasonable magnitudes on the probability domain. This is shown by Figure 2, which illustrates the effect of patient prognosis on π_E and π_T under scenario 2, and shows that higher doses are more desirable for patients with good or intermediate cytogenetics but lower doses are more desirable for patients with poor cytogenetics. This scenario reflects what might be anticipated to occur in actual treatment of AML. In scenario 3, $\pi_T(x, \boldsymbol{Z}, \boldsymbol{\theta})$ and hence $r_{\mathcal{C}}(\pi_E, \pi_T)$ is nonmonotone in dose, with $r_{\mathcal{C}}(\pi_E, \pi_T)$ largest at the middle-dose level. Scenario 4 is obtained from scenario 3 by adding the same interaction parameters as assumed in scenario 2.

The simulation results show that the method based on the full model including dose–covariate interactions does a very reliable job of selecting the most desirable doses within patient prognostic subgroups. The results for scenarios 2 and 4 are quite striking, because they show that the method very reliably detects treatment–covariate interactions and selects the doses that are most desirable within each subgroup. For example, under scenario 4, the higher doses that are most desirable for (27, good) patients are selected with high probability for those patients, the middle doses that are most desirable for (48, intermediate) patients are selected with high probability for those patients, and the lower doses that are most desirable for (58, poor) patients are selected with high probability for those patients. In terms of the desirability of the selected dose relative to the best possible dose, the values of ρ for $\mathbf{Z}$ − INT, $\mathbf{Z}$ − No INT, and No $\mathbf{Z}$ are virtually identical under the no-interaction scenarios 1 and 3, which reflects the fact that the dose acceptability limits change with patient prognosis (Figure 1). In contrast, under the dose–covariate interaction scenarios 2 and 4, the $\mathbf{Z}$ − No INT method that assumes a model ignoring interactions has ρ values substantially lower than those for $\mathbf{Z}$ − INT, and the ρ values for the No $\mathbf{Z}$ method are extremely low for (58, poor) patients. Thus, either ignoring treatment–covariate interactions or ignoring covariates entirely produces a method with greatly inferior properties.

Additional simulations under other scenarios are summarized in Web Appendix Supplementary Table 2. In particular, the additional simulations show that the method

Table 4

Simulation results for the AML trial. Dose-covariate interactions $\boldsymbol{\gamma} = (\gamma_{E,1}, \gamma_{E,2}, \gamma_{E,3}, \gamma_{T,1}, \gamma_{T,2}, \gamma_{T,3})$ *correspond to* $\boldsymbol{Z} =$ (*age, good cyto, poor cyto*)

		Dose level						
Scenario 1	$\boldsymbol{\gamma} = (0,0,0,0,0,0)$	1	2	3	4	5	None	ρ
27, good	π_E, π_T	0.57, 0.04	0.70, 0.06	0.89, 0.07	0.96, 0.08	0.97, 0.09		
	$r_{\mathcal{C}}(\pi_E, \pi_T)$	0.52	0.61	0.79	0.84	0.84		
Z – INT	% Selected	1	4	53	29	13	0	0.95
Z – NO INT	% Selected	0	2	35	34	29	0	0.97
NO **Z**	% Selected	0	1	37	33	25	4	0.97
48, intermediate	π_E, π_T	0.18, 0.20	0.28, 0.24	0.55, 0.28	0.74, 0.31	0.79, 0.33		
	$r_{\mathcal{C}}(\pi_E, \pi_T)$	0.25	0.28	0.38	0.47	0.49		
Z – INT	% Selected	0	2	32	37	29	0	0.91
Z – NO INT	% Selected	0	1	31	32	35	0	0.92
NO **Z**	% Selected	0	1	37	33	25	4	0.90
58, Poor	π_E, π_T	0.03, .30	0.07, 0.35	0.21, 0.40	0.39, 0.43	0.46, 0.45		
	$r_{\mathcal{C}}(\pi_E, \pi_T)$	0.18	0.18	0.21	0.26	0.27		
Z – INT	% Selected	10	11	14	23	42	0	0.88
Z – NO INT	% Selected	4	9	19	25	43	0	0.90
NO **Z**	% Selected	0	1	37	33	25	4	0.89
Scenario 2	$\boldsymbol{\gamma} = (0,.6,-0.8,0,-0.6,0.8)$	1	2	3	4	5	None	ρ
27, good	π_E, π_T	0.15, 0.30	0.46, 0.16	0.89, 0.07	0.99, 0.02	1.00, 0.01		
	$r_{\mathcal{C}}(\pi_E, \pi_T)$	0.22	0.39	0.79	0.95	0.99		
Z – INT	% Selected	0	0	29	40	31	0	0.93
Z – NO INT	% Selected	7	15	51	16	10	1	0.74
NO **Z**	% Selected	1	4	50	26	14	5	0.85
48, intermediate	π_E, π_T	0.18, 0.20	0.28, 0.24	0.55, 0.28	0.74, 0.31	0.79, 0.33		
	$r_{\mathcal{C}}(\pi_E, \pi_T)$	0.25	0.28	0.38	0.47	0.49		
Z – INT	% Selected	1	3	43	29	23	0	0.88
Z – NO INT	% Selected	7	14	47	17	13	2	0.79
NO **Z**	% Selected	1	4	50	26	14	5	0.86
58, Poor	π_E, π_T	0.41, 0.02	0.24, 0.12	0.21, 0.40	0.14, 0.73	0.04, 0.93		
	$r_{\mathcal{C}}(\pi_E, \pi_T)$	0.42	0.30	0.21	0.13	0.09		
Z – INT	% Selected	74	21	4	1	0	0	0.91
Z – NO INT	% Selected	26	26	23	11	14	0	0.62
NO **Z**	% Selected	1	4	50	26	14	5	0.41
		Dose level						
Scenario 3	$\boldsymbol{\gamma} = (0,0,0,0,0,0)$	1	2	3	4	5	None	
27, good	π_E, π_T	0.43, 0.03	0.76, 0.07	0.94, 0.12	0.90, 0.17	0.65, 0.31		
	$r_{\mathcal{C}}(\pi_E, \pi_T)$	0.43	0.66	0.79	0.70	0.42		
Z – INT	% Selected	0	8	82	10	0	0	0.97
Z – NO INT	% Selected	0	12	81	5	1	1	0.97
NO **Z**	% Selected	0	5	79	8	0	8	0.98
48, intermediate	π_E, π_T	0.10, 0.15	0.35, 0.28	0.68, 0.38	0.58, 0.47	0.24, 0.65		
	$r_{\mathcal{C}}(\pi_E, \pi_T)$	0.24	0.29	0.40	0.31	0.16		
Z – INT	% Selected	0	9	82	8	0	1	0.96
Z – NO INT	% Selected	0	11	81	6	0	2	0.95
NO **Z**	% Selected	0	5	79	8	0	8	0.96
58, poor	π_E, π_T	0.01, 0.23	0.10, 0.40	0.33, 0.50	0.24, 0.60	0.05, 0.76		
	$r_{\mathcal{C}}(\pi_E, \pi_T)$	0.19	0.18	0.22	0.17	0.11		
Z – INT	% Selected	13	32	42	9	4	0	0.89
Z – NO INT	% Selected	8	26	55	8	3	0	0.91
NO **Z**	% Selected	0	5	79	8	0	8	0.97
Scenario 4	$\boldsymbol{\gamma} = (0,0.6,-0.8,0,-0.6,0.8)$	1	2	3	4	5	None	
27, good	π_E, π_T	0.08, 0.24	0.54, 0.20	0.94, 0.12	0.97, .06	0.94, 0.04		
	$r_{\mathcal{C}}(\pi_E, \pi_T)$	0.21	0.42	0.79	0.89	0.88		
Z – INT	% Selected	0	0	46	46	8	0	0.95
Z – NO INT	% Selected	2	43	54	1	0	0	0.70
NO **Z**	% Selected	0	11	78	3	1	7	0.85
48, intermediate	π_E, π_T	0.10, 0.15	0.35, 0.28	0.68, 0.38	0.58, 0.47	0.24, 0.65		
	$r_{\mathcal{C}}(\pi_E, \pi_T)$	0.24	0.29	0.40	0.31	0.16		
Z – INT	% Selected	0	14	77	7	1	1	0.94
Z – NO INT	% Selected	3	36	57	1	1	2	0.88
NO **Z**	% Selected	0	11	78	3	1	7	0.95
58, poor	π_E, π_T	0.28, 0.01	0.31, 0.14	0.33, 0.50	0.07, 0.85	0.00, 0.99		
	$r_{\mathcal{C}}(\pi_E, \pi_T)$	0.35	0.32	0.22	0.10	0.07		
Z – INT	% Selected	38	58	4	0	0	0	0.93
Z – NO INT	% Selected	31	37	27	2	3	0	0.83
NO **Z**	% Selected	0	11	78	3	1	7	0.64

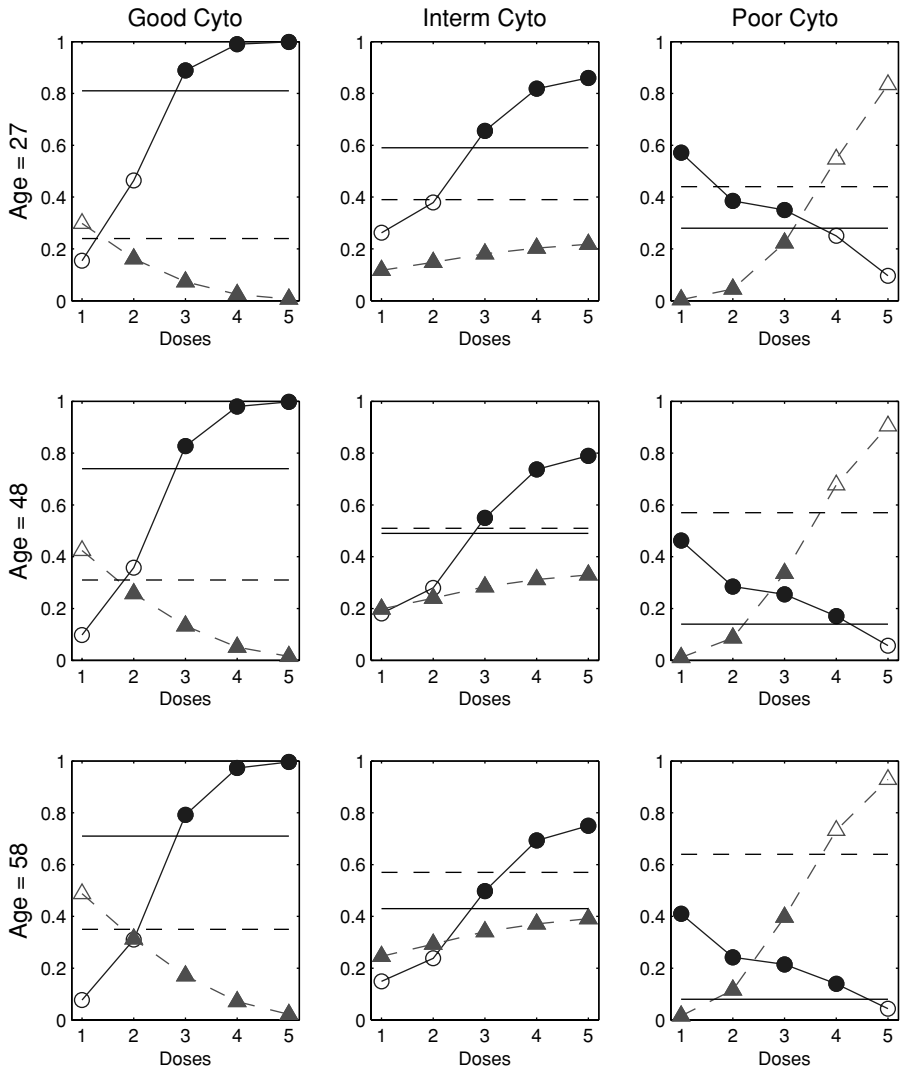

Figure 2. Simulation scenario 2 in terms of the fixed marginal outcome probabilities $\pi_E(x_j, \mathbf{Z})$ (shown as circles) and $\pi_T(x_j, \mathbf{Z})$ (shown as triangles) for nine different patient covariate $\mathbf{Z}$ vectors. Solid (open) points correspond to acceptable (unacceptable) doses.

terminates accrual with no dose selected with high probability in subgroups where toxicity is excessive for all doses. Simulations (not shown) using the PP desirability criterion $\delta_n^{(PP)}(x, \mathbf{Z})$ in place of $\delta_n^{(PM)}(x, \mathbf{Z})$ showed that the PM-based method has slightly better performance than the PP-based method in the cases studied.

6. Discussion

We have proposed a dose-finding method suitable for phase I–II clinical trials that chooses doses based on efficacy and toxicity while accounting for each patient's prognostic covariates (Figure 3). Our proposed method is very complex, and it requires a substantial effort on the part of both the statistician and the physicians planning the trial. Our simulation study shows that this complexity is justified. In terms of the probabilities of selecting acceptably safe and efficacious doses that are most desirable for each patient, our method provides substantial improvements over simpler versions that either ignore dose–covariate interactions (scenarios 2 and 4, Table 4) or that ignore covariates entirely (scenarios 1 and 3, Table 4). A general implication is that established dose-finding methods that ignore patient heterogeneity run a high risk of assigning inferior doses to particular patient subgroups. Given the complexity of our method, however, when entry criteria ensure that patients are reasonably homogeneous, it may be more appropriate to use simpler methods not adjusting patient covariates.

The method is computationally intensive and requires specialized software. To facilitate application of the method, computer programs needed for implementation are available from the second author on request.

7. Supplementary Materials

Web Appendices referenced in Section 5 are available under the Paper Information link at the *Biometrics* website `http://www.biometrics.tibs.org`.

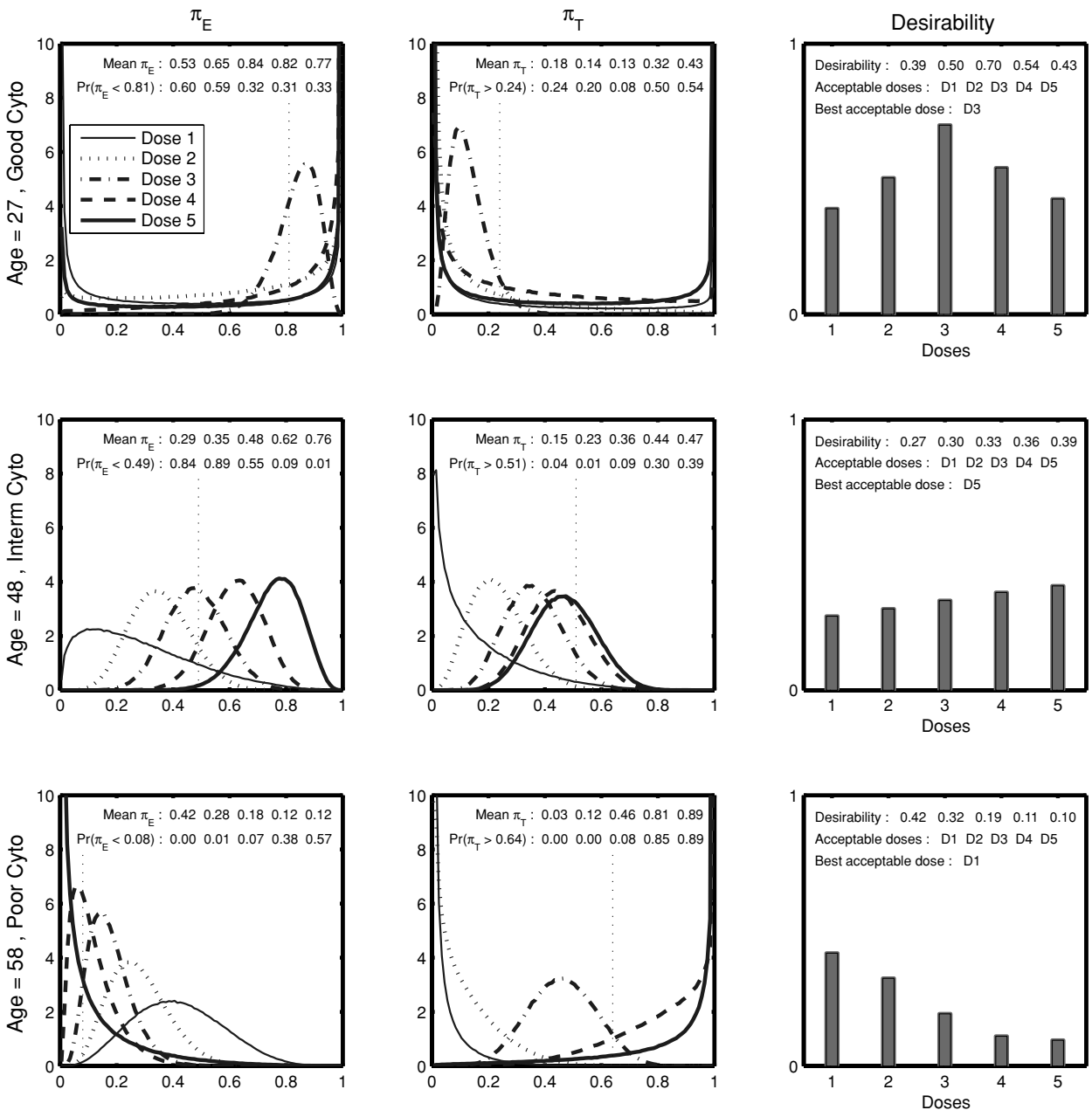

Figure 3. Illustration of the proposed method by the outcome of a single trial in terms of final posteriors of $\pi_E(x_j, \mathbf{Z})$ and $\pi_T(x_j, \mathbf{Z})$ and the desirability values $r_{\mathcal{C}}\{\boldsymbol{\pi}(x_j, \mathbf{Z})\}$ for doses $x_1, \ldots, x_5$, for $\mathbf{Z} = (27, \text{good})$, $(48, \text{interm})$, and $(58, \text{poor})$.

Acknowledgements

This research was partially supported by NIH grant RO1 CA-83932. The authors thank an associate editor and two referees for their helpful comments and suggestions.

References

Babb, J. S. and Rogatko, A. (2001). Patient specific dosing in a phase I cancer trial. *Statistics in Medicine* **20,** 2079–2090.

Bekele, B. N. and Shen, Y. (2004). A Bayesian approach to jointly modeling toxicity and biomarker expression in a phase I/II dose-finding trial. *Biometrics* **60,** 343–354.

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

Cheng, J. D., Babb, J. S., Langer, S., Aamdal, S., Robert, F., Engelhardt, L. P., Fernberg, O., Schiller, J., Forsberg, G., Alpaugh, R. K., Weiner, L. M., and Rogatko, A. (2004). Individualized patient dosing in phase I clinical trials: The role of escalation with overdose control in PNU-214936. *Journal of Clinical Oncology* **22,** 602–609.

Estey, E., Smith, T. L., Keating, M. J., McCredie, K. B., Gehan, E. A., and Freireich, E. J. (1989). Prediction of survival during induction therapy in patients with newly diagnosed acute myeloblastic leukemia. *Leukemia* **3,** 257–263.

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

Ivanova, A. and Wang, K. (2006). Bivariate isotonic design for dose-finding with ordered groups. *Biometrics* **25,** 2018–2026.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. New York: Chapman and Hall.

Mick, R. and Ratain, M. J. (1993). Model-guided determination of maximum tolerated dose in phase I clinical trials: Evidence of increased precision. *Journal of the National Cancer Institute* **85,** 217–223.

O'Quigley, J. and Paoletti, X. (2003). Continual reassessment method for ordered groups. *Biometrics* **59,** 430–440.

O'Quigley, J., Shen, L. Z., and Gamst, A. (1999). Two-sample continual reassessment method. *Journal of Biopharmaceutical Statistics* **9,** 17–44.

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

Piantadosi, S. and Liu, G. (1996). Improved designs for dose-escalation studies using pharmacokinetic measurements. *Statistics in Medicine* **15,** 1605–1618.

Prentice, R. L. (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics* **44,** 1033–1048.

Ratain, M. J., Mick, R., Janisch, L., Berezin, F., Schilsky, R. L., Vogelzang, N. J., and Kut, M. (1996). Individualized dosing of amonafide based on a pharmacodynamic model incorporating acetylator phenotype and gender. *Pharmacogenetics*, **6,** 93–101.

Robert, C. P. and Cassella, G. (1999). *Monte Carlo Statistical Methods*. New York: Springer.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B* **64,** 583–639.

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

Thall, P. F., Cook, J. D., and Estey, E. H. (2006). Adaptive dose selection using efficacy-toxicity trade-offs: Illustrations and practical considerations. *Journal of Biopharmaceutical Statistics* **16,** 623–638.

Thall, P. F., Wooten, L. H., and Shpall, E. J. (2006). A geometric approach to comparing treatments for rapidly fatal diseases. *Biometrics* **62,** 193–201.

Wijesinha, M. C. and Piantadosi, S. (1995). Dose-response models with covariates. *Biometrics* **51,** 977–987.

Yuan, Z. and Chappell, R. (2004). Isotonic designs for phase I cancer clinical trials with multiple risk groups. *Clinical Trials* **1,** 499–508.

Received June 2007. *Revised December* 2007.
Accepted December 2007.