Dose-Finding Based On Multiple Ordinal Toxicities in Phase I Oncology Trials

B. Nebiyou Bekele[1] and Peter F. Thall

Department of Biostatistics, Box 447

The University of Texas M.D. Anderson Cancer Center

1515 Holcombe Boulevard, Houston, Texas 77030, USA

[1]*E–mail: nbekele@odin.mdacc.tmc.edu*

October 30, 2002

SUMMARY. Nearly all statistical methods for dose-finding in phase I clinical trials require that toxicity be summarized as a single binary variable. In most phase I oncology settings, however, the patient is at risk of several qualitatively different toxicities, each occurring at several possible levels of severity. Moreover, the different toxicities often are not of equal clinical importance, even if they occur at the same nominal level of severity. To apply established dose-finding methods, it is common practice to first dichotomize each type of toxicity at a particular severity level and then define "toxicity" as the maximum of these indicators. We propose a new statistical framework for dose-finding based on a vector of qualitatively different, ordinal-valued toxicities. The underlying probability model specifies how the severity of each type of toxicity varies with dose, and a vector of correlated Gaussian latent variables is used to induce association among the different toxicities. Numerical weights characterizing the importance of each level of each type of toxicity are elicited from the physicians planning the trial, and they are used to define a one-dimensional total toxicity burden that is the basis for the dose-finding algorithm. The method is illustrated by application to a phase I trial of gemcitabine in soft tissue sarcoma.

KEY WORDS: Adaptive design; Bayesian inference; Latent variables; Markov Chain Monte Carlo

1. Introduction

The primary goal of a phase I clinical trial of a new chemotherapeutic agent in oncology is to determine a dose having acceptable toxicity. Because patient safety is a central concern in such trials, typically patients are treated in successive cohorts, with the dose for each cohort chosen adaptively utilizing the data from previous patients in the trial. Thus, a phase I design must provide both an algorithm for sequentially assigning doses to patient cohorts and a rule for selecting a dose, usually called the "maximum tolerated dose" (MTD), at the end of the trial. For convenience, we will refer to these two related statistical problems together as "dose-finding." Ethically, the phase I dose-finding problem is difficult because doses must be assigned to patients based on very little data, and this problem is especially acute early in the trial. The problem is scientifically difficult because any ethically reasonable algorithm must de-escalate when a dose is found to be unacceptably toxic. Consequently, because a limited number of toxicities are permitted to occur, little or no data are available on doses having high toxicity probabilities. Numerous statistical designs for phase I trials have been proposed (Storer, 1989; O'Quigley, Pepe and Fisher, 1990; Durham and Fluornoy, 1994; Whitehead, 1997; Babb, Rogatko and Zacks, 1998; Piantadosi, Fisher and Grossman, 1998; Gasparini and Eisele, 2000). Each of these approaches characterizes toxicity as a binary variable, with the underlying statistical model and the algorithm for trial conduct based on the probability of toxicity as a function of dose.

Given the paucity of data in phase I, it is especially important that statistical dose-finding methods use as much of the available information as possible. We are motivated by three closely related problems, all pertaining to the manner in which "toxicity" is defined before specification of a model and method for dose-finding in phase I. The first problem is that, in virtually all phase I oncology settings, the patient is at risk of several qualitatively different toxicities, rather than only one. A typical protocol for a phase I oncology trial includes a list of possible toxicities that must be monitored. This list often includes transient conditions such as fatigue, nausea and vomiting, myelosuppression (low blood cell count, associated with suppression of normal bone marrow function), thrombocytopenia (low platelet

count), fever, infection, dysfunction of specific organs, and irreversible toxicities such as permanent organ damage or death. For a variety of reasons related to the biochemistry of the particular chemotherapy or the biology of the disease, in general the toxicities do not occur independently. For example, myelosuppression often leads to infection or fever.

The second problem is that a given toxicity generally has several levels of severity. Phase I trial protocols routinely include detailed definitions of the grades of each type of toxicity, typically coded in terms of integer values varying from 0 (no toxicity of that type) to 4 (the most severe level, such as permanent damage to a particular organ). In order to accommodate statistical paradigms that require toxicity to be defined as a single binary variable, most phase I protocols define "toxicity" as the occurrence of any of several listed toxicities at grade 3 or 4. While it is reasonable in many phase I settings to reduce the ordinal scale of given type of toxicity to the binary variable for which grades 0, 1 or 2 are "no toxicity" while grades 3 or 4 are "toxicity," this common practice may discard potentially useful information. For example, if several patients experience a grade 2 toxicity of a given type at a dose level d, then a typical method based on the above binary variable would escalate to d+1 as if no toxicities had occurred at d. Clearly, a probability model that distinguishes between grades 0, 1 and 2, rather than combining them as the event "no toxicity," should provide a more reliable basis for predicting the jump from grade 2 to grade 3 or higher as the dose is increased from d to d+1.

The third problem is that qualitatively different toxicities often are not equally important clinically, even if they occur at the same nominal level of severity. Assuming for simplicity that each of several different types of toxicity has been defined as a binary variable, the further data reduction of defining "toxicity" as the maximum of these indicators implicitly assumes that the different toxicities are exchangeable, hence equally important. For example, this definition does not distinguish between a patient with grade 3 fatigue and a patient who has suffered complete kidney failure.

In this paper, we propose a statistical framework for dose-finding that addresses all of the problems described above. We characterize patient outcome as a vector of qualitatively

different, ordinal-valued toxicities. The underlying probability model specifies the manner in which the severity of each type of toxicity varies with dose, and a vector of correlated latent variables is used to induce association among the different toxicities. Our dose-finding method relies on numerical weights characterizing the importance of each level of each type of toxicity. These weights must be elicited from the physicians planning the trial. The weights provide a basis for defining a one-dimensional total toxicity burden for each patient that is the basis for the dose-finding algorithm. This method of dimension reduction quantifies the physicians' experiences in dealing with multiple toxicities in the clinic. Our application illustrates how the dose-finding algorithm reflects the physicians' clinical goals and standards more closely than do conventional methods that begin by defining a single binary toxicity as the maximum of several different toxicity severity indicators.

The remainder of the paper is organized as follows. We present the probability model in Section 2. Trial conduct is described in Section 3, including algorithms for eliciting toxicity severity scores and a target toxicity burden, and for assigning doses to successive patient cohorts. In Section 4, we describe numerical methods for computing posteriors and decision criteria. We illustrate the methodology with an application to a clinical trial in soft tissue sarcoma in Section 5. Section 6 illustrates how the method may be applied in the special case of one ordinal toxicity, and we conclude with a discussion in Section 7.

2. Multiple Toxicity Model

Let $\mathbf{Y} = (Y_1, \cdots, Y_J)$ denote the vector of ordinal toxicity variables. The j^{th} type of toxicity, Y_j, takes on one of the $C_j + 1$ values $y_{j,0}, y_{j,1}, \ldots, y_{j,C_j}$, where $y_{j,k}$ is the k^{th} most severe level for $k = 0, \ldots, C_j$, and $y_{j,0}$ = "no toxicity of type j." For example, if Y_j has the five possible values $y_{j,0}$ and $y_{j,k}$ = "grade k toxicity of type j" for k=1, 2, 3 or 4, then $C_j = 4$. Binary Y_j corresponds to $C_j = 1$.

We will apply the multivariate ordinal probit model of Chen and Dey (2000), extended to allow the entries of $\mathbf{Y}$ to have different numbers of ordinal categories. This class of models, which relies on a vector of correlated latent Gaussian variables to induce association among binary, categorical, or ordinal Y_j's, was developed by Albert and Chib (1993), and Chib and

Greenberg (1998). To improve numerical stability, we first replace each raw dose d by $x = \log(d/d*)$, where d^* is at least as large as the maximum raw dose value, and we will refer to x as the "dose." We model association among the Y_j's by introducing the vector $\mathbf{Z}^{J\times 1} = (Z_1, \cdots, Z_J)$ of correlated latent variables, which is assumed to be multivariate normal with mean $\mathbf{X}\boldsymbol{\beta}$, all variances equal to 1 and correlation matrix $\boldsymbol{\Omega}$. Here $\mathbf{X}^{J\times 2J}$ is the dose matrix

$$\mathbf{X} \;=\; \begin{bmatrix} 1\ x & 0\ 0 & \dots & 0\ 0 \\ 0\ 0 & 1\ x & \dots & 0\ 0 \\ \vdots & \vdots & & \vdots \\ 0\ 0 & 0\ 0 & \cdots & 1\ x \end{bmatrix}$$

and $\boldsymbol{\beta}^{2J\times 1} = (\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_J)$ with each $\boldsymbol{\beta}_j^{2\times 1} = (\beta_{j,0}, \beta_{j,1})$. The latent variable vector $\mathbf{Z}$ determines the observed outcome vector $\mathbf{Y}$ via the conditions

$$Y_j \;=\; y_{j,k} \text{ if } \gamma_{j,k} \le Z_j < \gamma_{j,k+1} \text{ for } k = 0, 1, \ldots, C_j \text{ and } j = 1, \ldots, J,$$

where $\boldsymbol{\gamma}_j^{C_j\times 1} = (\gamma_{j,1}, \cdots, \gamma_{j,C_j})$ is a vector of model parameters satisfying the ordering constraint $-\infty = \gamma_{j,0} < \gamma_{j,1} < \cdots < \gamma_{j,C_j} < \gamma_{j,C_{j+1}} = \gamma_{max}$ where γ_{max} is some fixed positive quantity (we use 10). We denote $A_{j,k} = (\gamma_{j,k},\ \gamma_{j,k+1}]$ and let $\boldsymbol{\gamma}^{C_+\times 1} = (\boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_J)$ be the vector of all cut-off parameters, where $C_+ = C_1 + \cdots + C_J$. The requirement that the variance-covariance matrix $\boldsymbol{\Omega}$ of $\mathbf{Z}$ be its correlation matrix is necessary to ensure identifiability of the posterior distributions, which also requires that $\gamma_{j,1} \equiv 0$. Since $\gamma_{j,0} = -\infty$, $\gamma_{j,C_j+1} = \gamma_{max}$, and $\gamma_{j,1} \equiv 0$, if $C_j > 1$ there are only $C_j - 1$ random cut-point parameters. Thus, although $\boldsymbol{\gamma}$ has C_+ entries, it actually contains only $\Sigma_{j=1}^{J}\, \mathbf{1}(C_j > 1)(C_j - 1)$ random parameters, where $\mathbf{1}(A)$ is the indicator of the event A. The marginal probability distribution of Y_j for a patient treated with dose x thus takes the form

$$\pi_{j,k}(x) \;=\; \Pr(Y_j = y_{j,k} \mid x) \;=\; \Phi\{\gamma_{j,k+1} - (\beta_{j,0} + \beta_{j,1}x)\}, \tag{1}$$

where Φ is the standard normal cdf. Denote $\boldsymbol{\pi}_j(x, \boldsymbol{\theta}) = (\pi_{j,1}(x, \boldsymbol{\theta}), \ldots, \pi_{j,C_j}(x, \boldsymbol{\theta}))$ for each $j = 1, \ldots, J$ and $\boldsymbol{\pi}(x, \boldsymbol{\theta}) = (\boldsymbol{\pi}_1(x, \boldsymbol{\theta}), \ldots, \boldsymbol{\pi}_J(x, \boldsymbol{\theta}))$. Let $\phi_{\mathbf{W}}(\cdot \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote the pdf of a multivariate normal random vector $\mathbf{W}$ with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix

$\boldsymbol{\Sigma}$. For a given vector $\mathbf{k} = (k_1, \ldots, k_J)$ of toxicity severity levels, denote the corresponding outcome of $\mathbf{Y}$ by $\mathbf{y}(\mathbf{k}) = (y_{1,k_1}, \ldots, y_{J,k_J})$, and the corresponding J-dimensional set of $\mathbf{Z}$ values by $\mathbf{A}(\mathbf{k}, \boldsymbol{\gamma}) = A_{1,k_1} \times \cdots \times A_{J,k_J}$. The likelihood of a single patient is given by

$$\mathcal{L}(\mathbf{Y} \mid \boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\Omega}, x) \;=\; \prod_{k_1=0}^{C_1} \cdots \prod_{k_J=0}^{C_J} \left\{ \int_{\mathbf{A}(\mathbf{k},\boldsymbol{\gamma})} \phi_{\mathbf{Z}}(\mathbf{z} \mid \mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Omega})\, \mathbf{dz} \right\}^{\mathbf{1}[\mathbf{Y}=\mathbf{y}(\mathbf{k})]}, \tag{2}$$

which shows how $\mathbf{Z}$ induces association among the elements of $\mathbf{Y}$ through the correlation matrix $\boldsymbol{\Omega}$. Denoting the dose assigned to the i^{th} patient by $x_{(i)}$ and the corresponding matrix by $\mathbf{X}_i$, the likelihood for n patients is obtained by substituting $\mathbf{Y} = \mathbf{Y}_i$, $x = x_{(i)}$ and $\mathbf{X} = \mathbf{X}_i$ in expression (2) and taking the product over $i = 1, \cdots, n$. We will denote the $J(J-1)/2$ unique off-diagonal elements of $\boldsymbol{\Omega}$ by $\boldsymbol{\rho} = (\rho_{1,2}, \rho_{1,3}, \cdots, \rho_{J-1,J})$ and let $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\rho})$, where $\boldsymbol{\gamma}$ contains only the random cut-point parameters, as described above.

We assume that $\boldsymbol{\beta} \sim N_{2J}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, *a priori*, subject to the constraint that $\Pr(\beta_{j,1} > 0) = 1$ for all $j = 1, \ldots, J$. That is, we abuse notation in that the prior of $\boldsymbol{\beta}$ is a $2J$-variate normal with J of its values truncated at 0, but $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ correspond to the untruncated $2J$-variate normal. This constraint ensures that $\Pr(Y_j > y_{j,k} \mid x) = 1 - \Phi\{\gamma_{j,k} - (\beta_{j,0} + \beta_{j,1}x)\}$ increases with x for each j and $k > 1$, which is necessary for the model to make sense. For each j with $C_j > 1$, we will assume that the random parameters $\{\gamma_{j,2}, \ldots, \gamma_{j,C_j}\}$ follow independent uninformative priors, with each $g(\boldsymbol{\gamma}_j) \propto 1$, subject to the constraint $0 < \gamma_{j,2} < \gamma_{j,3} < \cdots < \gamma_{j,C_j} < \gamma_{max}$. For the correlation parameters, we assume that the elements of $\boldsymbol{\rho}$ are iid $N(0, 1000)$ truncated to the support $[-1, +1]$. We denote the prior of $\boldsymbol{\theta}$ by $f(\boldsymbol{\theta})$.

3. Toxicity Weight Elicitation and Dose-Finding

If doses are to be chosen based on multivariate toxicity data, inevitably some form of dimension reduction must be carried out. The conventional approach is to first reduce each Y_j to $\mathbf{1}(Y_j > y_{j,k})$ for a toxicity level $y_{j,k}$ of type j considered by the physicians to be dose limiting, and then define "toxicity" to be the maximum of the J indicators. As described in Section 1, this conventional approach suffers from several pathological properties. The following proposed alternative approach does away with these pathologies by incorporating medical knowledge into the dimension reduction. We begin by eliciting positive-valued

numerical weights characterizing the importance of each severity level of each type of toxicity from the physicians planning the trial. For each $j = 1, \ldots, J$, let $w_{j,k}$ be the elicited weight of toxicity type j occurring at severity level $y_{j,k}$, and denote $\mathbf{w}_j = (w_{j,1}, \cdots, w_{j,C_j})$ and $\mathbf{w}^{C_+ - J \times 1} = (\mathbf{w}_1, \ldots, \mathbf{w}_J)$. The j^{th} weight vector must satisfy the obvious requirement that $0 = w_{j,0} < w_{j,1} < w_{j,2} < \cdots < w_{j,C_j}$. If the physicians consider two consecutive levels of Y_j to have the same weight, then these two levels should be combined. For example, if they give grades 3 and 4 of a given toxicity the same weight then these should not be considered separate levels. The numerical values of the $w_{j,k}$'s may be on any positive domain, since the method is invariant to the weights' multiplicative scale. In our application, we found it convenient to elicit the weights on the interval $(0, 10)$.

We define the *severity weight* of Y_j for a patient treated at dose x to be the random variable W_j taking on the value $w_{j,k}$ with probability $\pi_{j,k}(x, \boldsymbol{\theta})$. This replaces the observed ordinal variable Y_j with the weight-valued random variable W_j that puts the severity category probabilities of Y_j on the corresponding elicited weights. We define the *total toxicity burden* to be the sum $\mathrm{TTB} = \sum_{j=1}^{J} W_j$, and base the dose-finding algorithm on the posterior expected TTB at each dose:

$$\psi(\mathbf{w}, x, data) \;=\; \mathrm{E}\{\, \mathrm{E}\,(\mathrm{TTB} \mid x, \boldsymbol{\theta}) \;\mid data\} \;=\; \sum_{j=1}^{J} \sum_{k=1}^{C_j} w_{j,k}\, \mathrm{E}\{\pi_{j,k}(x, \boldsymbol{\theta}) \mid data\}. \qquad (3)$$

The dose-finding algorithm requires a fixed target TTB value, ψ^*, which must be elicited from the physicians to reflect their clinical decision-making. Each successive cohort's dose is that having posterior mean $\psi(\mathbf{w}, x, data)$ closest to ψ^*. Using the posterior median or mode of $\sum_{j=1}^{J} W_j$ rather than the mean are also reasonable alternatives.

The severity weights and the target TTB may be obtained as follows. This must be done in close collaboration with the physicians, and we strongly advise against a statistician specifying either the weights or the target TTB independently. After asking the physicians to specify the particular toxicities to be monitored and their severity levels, ask them to specify a numerical severity weight for each level of each toxicity. This establishes $\mathbf{w}_1, \ldots, \mathbf{w}_J$. Next, identify a set of hypothetical J-variate toxicity outcomes for each of the patients in m hypothetical cohorts. If the planned cohort size is c, these may be denoted $(\mathbf{y}^*_{1,1}, \ldots, \mathbf{y}^*_{1,c}), \ldots, (\mathbf{y}^*_{m,1}, \ldots, \mathbf{y}^*_{m,c})$.

These outcomes should be chosen so that they cover a reasonably wide range of realistic possibilities. Let $\mathbf{w}^*_{r,l}$ be the severity weight vector corresponding to $\mathbf{y}^*_{r,l}$, for $r = 1, \ldots, m$ and $l = 1, \ldots, c$. The mean total toxicity burden of the r^{th} hypothetical cohort is

$$\overline{\mathrm{TTB}}_r^{\,*} \;=\; \frac{1}{c} \sum_{l=1}^{c} \sum_{j=1}^{J} w^*_{r,l,j} \;.$$

Denote the ordered mean hypothetical TTB values by $\overline{\mathrm{TTB}}^{\,*}_{(1)} \leq \cdots \leq \overline{\mathrm{TTB}}^{\,*}_{(m)}$. Next, for each hypothetical cohort, $r = 1, \ldots, m$, ask the physicians to determine whether observation of $\mathbf{y}^*_{r,1}, \ldots, \mathbf{y}^*_{r,c}$ would cause them to repeat the same dose (D_r = Repeat), escalate the dose (D_r = Escalate), or de-escalate the dose (D_r = De-escalate) for the next cohort. Denote the vector of decisions in the above order of increasing TTB by $D_{(1)} \leq D_{(2)}, \ldots \leq D_{(m)}$. An admissible sequence of decisions ordered in this way is defined as one consisting of a string of escalations, followed by a string of repeats, followed by a string of de-escalations. If the m decisions are not admissible then, in collaboration with the physicians, modify the hypothetical outcomes, elicited decisions, weights, or possibly other portions of the underlying structure as appropriate. Once an admissible set of decisions is obtained, define the target TTB to be the mean of the elicited $\overline{\mathrm{TTB}}_r^{\,*}$ values for which the physicians' decision was to repeat the same dose. Formally,

$$\psi^* \;=\; \frac{\sum_{r=1}^{m} \overline{\mathrm{TTB}}_r^{\,*}\, \mathbf{1}(D_r = \mathrm{Repeat})}{\sum_{r=1}^{m} \mathbf{1}(D_r = \mathrm{Repeat})} \;.$$

4. Computing

We will follow the computational framework developed by Albert and Chib (1993) for one polytomous outcome, extended by Chib and Greenberg (1998) to accommodate correlated binary outcomes and by Chen and Dey (2000) to the correlated ordinal case. Denote the outcome indices of the i^{th} patient by $\mathbf{k}_i = (k_{i,1}, \ldots, k_{i,J})$, for $i = 1, \ldots, n$, and write $\mathbf{Z}^{(n)} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_n)$ and $\mathbf{Y}^{(n)} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_n)$. Since $\Pr\{\mathbf{Y}_i = \mathbf{y}(\mathbf{k}_i) \mid \mathbf{Z}_i, \boldsymbol{\theta}\} = \mathbf{1}\{\mathbf{Z}_i \in \mathbf{A}(\mathbf{k}_i, \boldsymbol{\gamma})\}$, by Bayes' theorem the joint posterior of the latent variables and parameters is given by

$$f(\mathbf{Z}^{(n)}, \boldsymbol{\theta} \mid \mathbf{Y}^{(n)}) \;=\; \prod_{i=1}^{n} f(\mathbf{Z}_i, \boldsymbol{\theta} \mid \mathbf{Y}_i = \mathbf{y}(\mathbf{k}_i),\, x_i) \propto f(\boldsymbol{\theta}) \prod_{i=1}^{n} \mathbf{1}\{\mathbf{Z}_i \in \mathbf{A}(\mathbf{k}_i, \boldsymbol{\gamma})\}\, f(\mathbf{Z}_i \mid \boldsymbol{\theta},\, x_i). \tag{4}$$

Under this representation, the latent variables will be used to ease the computational burden of computing the posterior $f(\boldsymbol{\theta} \mid \mathbf{Y})$. By using MCMC methods, values of $(\mathbf{Z}^{(n)}, \boldsymbol{\theta})$ generated from (4) will yield the desired posterior.

The following algorithm is similar to those given by Chen and Dey (2000) and Cowles (1996). Let $\mathbf{Z}_j^{(n)} = (Z_{j,1}, \ldots, Z_{j,n})$ be the vector of n independent latent variables associated with the j^{th} toxicity, and let $\mathbf{Z}_{-j}^{(n)}$ denote the subvector of $\mathbf{Z}^{(n)}$ obtained by deleting $\mathbf{Z}_j^{(n)}$, that is, by deleting $Z_{i,j}$ from $\mathbf{Z}_i$ for each $i = 1, \ldots, n$. The MCMC proceeds as follows:

Step 1. For each $j = 1, \ldots, J$, beginning with $f(\boldsymbol{\gamma}_j, \mathbf{Z}_j^{(n)} \mid \mathbf{Z}_{-j}^{(n)}, \mathbf{Y}^{(n)}, \boldsymbol{\beta}, \boldsymbol{\rho})$, integrate out $\mathbf{Z}_j^{(n)}$ to obtain $f(\boldsymbol{\gamma}_j \mid \mathbf{Z}_{-j}^{(n)}, \mathbf{Y}^{(n)}, \boldsymbol{\beta}, \boldsymbol{\rho})$, generate $\boldsymbol{\gamma}_j$ from this distribution, and generate $\mathbf{Z}_j^{(n)}$ from $f(\mathbf{Z}_j^{(n)} \mid \mathbf{Z}_{-j}^{(n)}, \boldsymbol{\gamma}_j, \mathbf{Y}^{(n)}, \boldsymbol{\beta}, \boldsymbol{\rho})$.

Step 2. Generate $\boldsymbol{\beta}$ from $f(\boldsymbol{\beta} \mid \mathbf{Z}^{(n)}, \mathbf{Y}^{(n)}, \boldsymbol{\gamma}, \boldsymbol{\rho})$.

Step 3. Generate $\boldsymbol{\rho}$ from $f(\boldsymbol{\rho} \mid \mathbf{Z}^{(n)}, \mathbf{Y}^{(n)}, \boldsymbol{\gamma}, \beta)$.

Step 1 utilizes the fact that $f(\boldsymbol{\gamma}_j \mid \mathbf{Z}_{-j}^{(n)}, \mathbf{Y}^{(n)}, \boldsymbol{\beta}, \boldsymbol{\rho}) = f(\boldsymbol{\gamma}_j \mid \mathbf{Z}_{-j}^{(n)}, \boldsymbol{\gamma}_{-j}, \mathbf{Y}^{(n)}, \boldsymbol{\beta}, \boldsymbol{\rho})$ due to the conditional independence of $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_J$. It alternates between this distribution and $f(\mathbf{Z}_j^{(n)} \mid \mathbf{Z}_{-j}^{(n)}, \boldsymbol{\gamma}_j, \mathbf{Y}^{(n)}, \boldsymbol{\beta}, \boldsymbol{\rho})$ because these are much more tractable than $f(\boldsymbol{\gamma}_j \mid \mathbf{Z}^{(n)}, \mathbf{Y}^{(n)}, \boldsymbol{\beta}, \boldsymbol{\rho})$. Because this MCMC approach generates observations from the full conditional $f(\boldsymbol{\theta}, \mathbf{Z}^{(n)} \mid \mathbf{Y}^{(n)})$, the posterior $f(\boldsymbol{\theta} \mid \mathbf{Y}^{(n)})$ is obtained as a natural consequence. Details of the above steps are given in an Appendix.

5. Application

We illustrate the methodology by describing application to a phase I trial of pre-surgical gemcitabine (G) with external beam radiation (EBR) for patients with soft tissue sarcoma. The trial was activated in January, 2002, and is ongoing at this writing. The three physicians conducting the trial, P. Pisters, M. Ballo, and S. Patel, together specified the toxicities and severity weights summarized in Table 3. The process of eliciting this information took a total of four sessions, during which the physicians successively modified the toxicities, their categories, and the category severity weights. These modifications were based on the observed

behavior of the algorithm and the fact that this process requires careful consideration of the clinical importance of each severity level of each type of toxicity. The elicitation process also was attenuated, in part, because the model, methodology, and computer code still were being developed. This experience motivated the algorithm for eliciting the toxicity severity weights and the total toxicity burden that we have described above, in Section 3. We expect the elicitation process to take considerably less time in future applications of the method.

Each patient in the trial receives a fixed dose of EBR (50 cGy) and one of ten doses of gemcitabine, 100, 200, ..., or 1000 mg/m^2. Patients will be treated in cohorts at successively chosen dose levels, with the first cohort treated at 400 mg/m^2. As in many phase I trials, the physicians imposed the extra safety measure that no untried dose may be skipped when escalating. Because it may take up to 6 weeks to evaluate each of the five types of toxicity in a given patient, the cohort size is allowed to vary between 3 and 4, as follows. If the first 3 patients in a cohort have had all of their toxicities evaluated before a 4th patient is accrued, then that cohort is considered complete and a new cohort is treated at the next chosen dose. When at least 36 patients have been accrued, the trial will stop. Using the elicitation method described earlier, the target per patient TTB was determined to be $\psi^* = 3.04$.

As a basis for comparison, for this trial conventional methods typically would define one binary "toxicity" as the maximum of the indicators $\mathbf{1}$(Myelosuppression grade ≥ 3), $\mathbf{1}$(Dermatitis grade $= 4$), $\mathbf{1}$(Liver toxicity grade ≥ 3), $\mathbf{1}$(Nausea/vomiting grade ≥ 3), $\mathbf{1}$(Fatigue grade $= 4$). For example, a conventional method would consider a patient with grade 3 dermatitis and grade 2 liver toxicity (TTB=4.5) to have "no toxicity" and a patient with grade 4 fatigue (TTB=1) to have "toxicity," and furthermore would not distinguish the latter patient from a patient with grade 4 myelosuppression with fever, grade 4 dermatitis, and grade 4 liver toxicity (TTB=18). Consequently, the proposed algorithm based on the TTB with target 3.04 for $\psi(\mathbf{w}, x, data)$ makes more sensible decisions. For example, if 3 of 4 patients treated at 400 mg/m^2 have either grade 4 fatigue or grade 3 myelosuppression without fever and one also has grade 3 nausea, for TTB values $\{0, 1, 1, 2.5\}$ and empirical mean TTB = 1.125, then the algorithm would escalate to 500 mg/m^2, whereas a conventional method would score 3

toxicities in the 4 patients and de-escalate to a lower dose.

An illustration of how the algorithm behaves in practice is given in Table 2, with the corresponding TTB values $\psi(\mathbf{w}, x, data_n)$ after n patients' data have been observed given in Table 3 for n = 4, 8,..., 36, that is, after each cohort. The first two cohorts' outcomes are actual data from the trial, whereas the data for all subsequent cohorts are hypothetical. After the first cohort, although $\psi(\mathbf{w}, 700, data_4) = 3.24$ is closest to the target 3.04, since no untried dose may be skipped when escalating the second cohort is treated at 500. Incorporating the second cohort's data, since $\psi(\mathbf{w}, 600, data_8) = 2.85$ is closest to 3.04 the third cohort is treated at 600. The next value $\psi(\mathbf{w}, 700, data_{12}) = 3.22$ determines that the fourth cohort receives dose 700, and the trial subsequently de-escalates to 600 for two cohorts and then returns to 700 as the dose for the final three cohorts, with $\psi(\mathbf{w}, 700, data_{36}) = 2.97$ determining 700 to be the MTD. Note that any conventional method based on a typical binary toxicity would have scored 3 toxicities in the first cohort of 4 patients and thus de-escalated to 300.

To assess average behavior of the method, we performed a simulation study of the sarcoma trial. Due to the inherent complexity of patient outcome, specifying a reasonably representative set of possible dose-toxicity probabilities to study is not straightforward. In order to obtain a manageable set of dose-toxicity scenarios for the simulation study, we considered only cases where the target TTB occurred at 200, 500, or 800 mg/m^2, and we categorized the main source of toxicity as being either those having high severity (HS) weights, of 5 or greater, or low severity (LS) weights, of 2 or smaller. The remaining toxicities, specifically grade 3 dermatitis (w=2.5) and grade 3 liver toxicity (w=3), were considered intermediate and were included in either group. Thus, we studied a total of six different scenarios. Each scenario was characterized by $10\,C_+ = 130$ fixed probabilities $p_{j,0,d}, p_{j,1,d}, ..., p_{j,C_j,d}$, for $j = 1, ..., 5$ and d = 100, ..., 1000, where $p_{j,k,d}$ is the fixed probability of the jth toxicity occurring at its kth severity level with dose d. These probabilities were chosen nonparametrically, and not determined by the parametric model underlying the method. Figure 1 summarizes the six simulation scenarios graphically in terms of the TTB as a function of dose.

Association among the elements of each simulated toxicity vector $(Y_1, \ldots, Y_5)$ was induced

by first generating observations from a vector $\mathbf{Z}^{5\times1}$ of standard normal random variables having a specified correlation matrix, then defining the vector $\mathbf{U}^{5\times1} = (\Phi(Z_1), \ldots, \Phi(Z_5))$ of correlated uniform(0,1) random variates, and then, denoting $P_{j,k,d} = \sum_{r=0}^{k} p_{j,r,d}$ for $k = 0, \ldots, C_j$ and $P_{j,-1,d} = 0$, defining $Y_j = y_{j,k}$ if $P_{j,k-1,d} \leq U_j < P_{j,k,d}$. The correlations were elicited from the physicians in terms of the latent variables, $\mathbf{Z}$, underlying the toxicities, as follows. The only toxicities that the physicians considered correlated *a priori* were fatigue and nausea/vomiting. We asked the physicians the following question: "For two randomly chosen patients, if you observe that the first patient is more fatigued than the second, what is the probability of observing more severe nausea/vomiting in the first patient than the second?" The physicians' response was that they would assign a probability between 0.55 and 0.60 to this event. Denoting the latent variables corresponding to fatigue and nausea/vomiting for the two patients by $Z_{i,F}$ and $Z_{i,N}$ for $i = 1,2$ and $q = \Pr(Z_{1,F} > Z_{2,F} \mid Z_{1,N} > Z_{2,N})$, if we assume q is symmetric in that $q = \Pr(Z_{1,F} < Z_{2,F} \mid Z_{1,N} < Z_{2,N})$, then this probability is related to Kendall's τ via $\tau = 2q - 1$. Furthermore, τ and the Pearson's correlation ρ between Z_F and Z_N satisfy the relationship $\rho = \sin(\tau\,\pi/2)$ (Kruskal, 1958). Thus, $0.55 < q < 0.60$ implies that $.15 < \rho < .31$. For the simulations, we used the average $\rho = 0.23$.

For the simulation study, we evaluated the MCMC algorithm's performance using standard convergence diagnostics. It was determined that a burn-in of 1,000 and a chain of length 30,000, retaining every 15th sample, provided adequate convergence. Although the posterior sample size is constrained by computing resources due to the need for many replications in the simulation study, in the actual trial conduct we base all inferences on a much larger MCMC posterior sample size.

The trial was simulated 1000 times under each scenario. Table 3 summarizes the results. Under scenarios 1 and 2, because the target TTB of 3.04 is achieved at 200 mg/m^2, the starting dose of 400 mg/m^2 is unacceptably toxic. Under scenario 1, most of the toxicity is due to LS events such as grade 2 liver failure, grade 3 or 4 fatigue, or nausea. In contrast, most of the toxicity burden under scenario 2 is due to HS toxicities such as myelosuppression with fever or grade 4 liver toxicity. Under either of these two scenarios, the method chooses

the correct dose 200 mg/m^2 over 90% of the time, and on average treats 22 of 36 patients at this dose. The algorithm thus appears to perform well regardless of whether most of the toxicity burden arises from low or high severity toxicities. For scenarios 3 and 4, the target TTB is achieved at 500 mg/m^2, with most of the TTB due to LS toxicities under under scenario 3 and HS toxicities under scenario 4. Again, the method is insensitive to the source of the toxicities, with an 85% to 87% correct selection rate and most of the 36 patients treated at or near the selected MTD. For each of scenarios 5 and 6, where the target TTB occurs at 800 mg/m^2, the correct selection probability is about 80%, slightly lower than the other cases. This is due primarily to the "do-not-skip" rule, which requires that at least one cohort be treated at each dose level when escalating. Consequently, at least 16 of the 36 patients must be treated at doses below 800 mg/m^2, which reduces the number of patients available for evaluation at the higher dose levels.

We also examined the sensitivity of the method to varying cohort size, sample size, starting dose, and correlation among the toxicities. The additional simulations examining the effects of cohort size and sample size were conducted under scenario 4, which has target TTB achieved at 500 mg/m^2 from mostly HS toxicities. For cohort sizes 1, 2, 3, 4, and 5, with starting dose 400 and sample size 36 as previously, the respective correct selection percentages were 89.0, 85.5, 88.8, 86.9 and 86.4. Since the range of these values is well within what would be expected from simulation variation, the method thus appears to be insensitive to cohort size. For sample sizes 28, 32, 36, 40, 44, with the cohort size fixed at 4 and the starting dose 400 mg/m^2, the correct selection percentages were 82.5, 84.2, 86.9, 88.4, and 89.2. Thus, as might be expected, the reliability of the method improves with larger sample size. We examined the effect of changing the starting dose from 400 mg/m^2 to 100 mg/m^2 under scenarios 2, 4, and 6, where the target TTB is achieved at 200 mg/ m^2, 500 mg/m^2, and 800 mg/m^2, respectively. In these cases the correct selection percentages were 91.7% when the target is 200 mg/m^2, 83.5% when the target is 500 mg/m^2 and 42.5% when the target is 800 mg/m^2. The comparatively low value in the last case is as expected, since the "do not skip rule" with starting dose 100 mg/m^2 requires that 28 of the 36 patients be treated at doses below 800

mg/m^2, which leaves at most two cohorts to treat at the correct dose. If this rule is dropped, the correct selection percentage in this case is 79.7%. However, the trade-off between larger correct selection probabilities and the extra measure of safety afforded by the "do-not-skip" rule is an important ethical consideration and most clinical oncologists would prefer including this rule.

Finally, we examined the effect of higher correlation among the toxicities on the correct selection rate. For this purpose, the correlations were changed such that there was high correlation between fatigue and nausea/vomiting (0.60), low correlation between fatigue and dermatitis (0.20) and moderate correlation between fatigue and myelosuppression (0.40). Under scenarios 2, 4, and 6, with this correlation structure the correct selection percentages were 91.9% when the target TTB was achieved at 200 mg/ m^2, 86.1% at 500 mg/ m^2, and 79.1% at 800 mg/ m^2. Since these are nearly identical to the values in Table 4 obtained with the original correlation structure, it appears that this degree of association among the toxicities does not alter the method's behavior, on average.

6. One Ordinal-Valued Toxicity

Because most dose-finding methods discussed in the literature and used to conduct phase I trials are based on one binary toxicity, it is worthwhile to explain how our method works in the case of one ordinal-valued toxicity, Y. Here, Y takes on one of $C+1$ ordinal severity values $y_0, y_1, \ldots, y_C$, there is one latent variable $Z \sim N(\beta_0 + \beta_1 x, 1)$ with $(Y = y_k) = (\gamma_k \leq Z < \gamma_{k+1})$, and $\pi_k(x, \boldsymbol{\theta}) = \Pr(Y = y_k \mid x, \boldsymbol{\theta}) = \Phi\{\gamma_{k+1} - (\beta_0 + \beta_1 x)\}$ for $k = 0, \ldots, C$, where $0 = \gamma_1 < \gamma_2 < \cdots < \gamma_C$. Only one vector of increasing toxicity severity weights, $(w_1, \ldots, w_k)$, is elicited, $TTB = W$ where W is the single random variable with $\Pr(W = w_k) = \pi_k(x, \boldsymbol{\theta})$, and ψ^* is the elicited target for $\mathrm{E}(W \mid data) = \sum_{k=1}^{C} w_k \, E\{\pi_k(x, \boldsymbol{\theta}) \mid data\}$. For example, suppose that a single ordinal toxicity is defined in terms of grades 0, 1, 2, 3, 4 with elicited severity weights 0, 1, 2, 3, 6. Suppose further that, for a given dose x, $\boldsymbol{\pi}^{(a)}(x) = (.50, .10, .10, .20, .10)$ and $\boldsymbol{\pi}^{(b)}(x) = (.10, .10, .50, .10, .20)$. Both of these probability vectors yield the same conventionally used probability $\Pr(Y \geq 3) = .30$ of "severe" (grade 3 or 4) toxicity, whereas $\boldsymbol{\pi}^{(a)}(x)$ has $\mathrm{E}^{(a)}(TTB) = 1.5$ while $\boldsymbol{\pi}^{(b)}(x)$ has $\mathrm{E}^{(b)}(TTB) = 2.6$.

This illustrates the fact, even if only one toxicity is considered, accounting for multiple toxicity levels by using their probabilities and elicited toxicity severity weights provides a more informative way to evaluate toxicity.

7. Discussion

We have described a statistical framework for dose-finding that accommodates multiple types of ordinal toxicities for which the severity levels have varying clinical importance. The proposed method requires considerably more effort to implement than conventional dose-finding methods based on a single binary toxicity. This includes close interaction between the physicians and statisticians to establish the toxicities, their severity weights, and the target TTB, as well as a subsequent simulation study of the design to establish its operating characteristics and, if appropriate, calibrate the design parameters. We feel that this effort is well warranted by the advantages that the method provides over more conventional dose-finding procedures. By providing a more complete and realistic accounting of adverse treatment effects, the method is able to make much more informed decisions than methods that reduce the multivariate ordinal toxicity outcome to a single binary variable. Our application to the soft-tissue sarcoma trial illustrates the method's flexibility, and the simulation study shows that on average the algorithm performs quite well under a wide variety of circumstances.

Still, such a thorough accounting of the complexities of patient outcome may not appeal to some clinical collaborators, since their participation in the design process requires considerably more work than is the case with conventional methods. We have noted that the method is much easier to implement in the case of one ordinal toxicity, and this may serve as a bridge to more complex settings as the process of physician-statistician collaboration evolves in a given clinic.

Because the statistical framework is, at its heart, a regression model, it can easily incorporate covariates. Accounting for heterogeneity in this way would provide a basis for patient-specific dosing, as discussed by Wijesinha and Piantadosi (1995), Babb and Rogatko (2001), and Legedza and Ibrahim (2001) in the single binary toxicity case. A rather different but equally important extension would be to incorporate efficacy outcomes, such as the

degree of tumor shrinkage. This would be similar in spirit to the methods proposed by Thall and Russell (1998), O'Quigley, Hughes and Fenton (2001), or Braun (2002). These currently are topics for future study.

Acknowledgement

Dr. Thall's work was partially supported by NIH grant RO1-CA-83932.

References

Albert, J.H. and Chib, S. (1993) Bayesian analysis of binary and polytomous response data. *Journal of the American Statistical Association* **88**, 669-679.

Babb, J., Rogatko, A. and Zacks, S. (1998). Cancer phase I clinical trials: Efficient dose escalation with overdose control. *Statistics in Medicine* **17**, 1103–1120.

Babb, J., Rogatko, A. (2001). Patient-specific dosing in cancer phase I clinical trials. *Statistics in Medicine* **20**, 2079-2090.

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.

Chen, M-H. and Dey, D.K. (2000) Bayesian analysis for correlated ordinal data models. In *Generalized Linear Models: A Bayesian Perspective*, D.K. Dey, S.K. Ghosh and B.K. Mallick, (eds),133-157. New York: Marcel Dekker.

Cheung, Y.K. and Chappell, R. (2000) Sequential designs for phase I clinical trials with late-onset toxicities. *Biometrics* **56**, 1177-1182.

Chib, S. and Greenberg, E. (1998) Analysis of multivariate probit models. *Biometrika* **85**, 347-361.

Cowles, M.K. (1996) Accelerating Monte Carlo Markov chain convergence for cumulative-link generalized linear models. *Statistics and Computing* **6**, 101-111.

Durham, S.D. and Fluornoy, N. (1994) Random walks for quantile estimation. In Statistical Decision Theory and Related Topics, S. Gupta and J.O. Berger (eds), 467-476. New York: Springer-Verlag.

Gasparini, M. and Eisele, J. (2000) A curve-free method for phase I clinical trials. *Biometrics* **56**, 609-615. Correction **57** ,659-660.

Kruskal W. (1958). Ordinal Measures of Association. *Journal of the American Statistical Association* **53**, 814-861.

Legedza, A. and Ibrahim, J. (2001) Heterogeneity in phase I clinical trials: prior elicitation and computation using the continual reassessment method. *Statistics in Medicine* **20**, 867-882.

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

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

Piantadosi, S., Fisher, J.D. and Grossman, S. (1998). Practical implementation of a modified continual reassessment method for dose–finding trials. *Cancer Chemotherapy and Pharmacology* **41**, 429-436.

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

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

Whitehead, J. (1997) Bayesian decision procedures with applications to dose finding studies. *International Journal of Pharmaceutical Medicine* **11**, 201-208.

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

Appendix

To generate the cut-point parameters $\boldsymbol{\gamma}_j$ from $f(\boldsymbol{\gamma}_j \mid \mathbf{Z}_{-j}^{(n)}, \mathbf{Y}^{(n)}, \boldsymbol{\beta}, \boldsymbol{\rho})$ in Step 1, first note that, due to independence among patients,

$$\begin{aligned} f(\mathbf{Z}_j^{(n)}, \boldsymbol{\gamma}_j \mid \mathbf{Y}^{(n)}, \mathbf{Z}_{-j}^{(n)}, \boldsymbol{\beta}, \boldsymbol{\rho}) &= \prod_{i=1}^{n} \Big\{ f(Z_{i,j},\, \boldsymbol{\gamma}_j \mid \mathbf{Y}_i,\, \mathbf{Z}_{i,-j},\, \boldsymbol{\beta}, \boldsymbol{\rho}) \Big\} \\ &\propto \prod_{i=1}^{n} \mathbf{1}(Z_{i,j} \in A_{j,k_{i,j}})\, \phi(Z_{i,j} \mid m_{i,j}, \nu_{i,j}), \end{aligned}$$

where $m_{i,j}$ and $\nu_{i,j}$ are the location and scale parameters for the truncated normal,

$$m_{i,j} = \beta_{j,0} + x_i \beta_{j,1} + \boldsymbol{\rho}_{(j,-j)}\, \boldsymbol{\rho}_{(-j,-j)}^{-1}\, (\mathbf{Z}_{i,-j} - (\mathbf{X}\boldsymbol{\beta})_{-j})$$

and

$$\nu_{i,j} = 1 - \boldsymbol{\rho}_{(j,-j)}\, \boldsymbol{\rho}_{(-j,-j)}^{-1}\, [\boldsymbol{\rho}_{(j,-j)}]^t,$$

denoting $\mathrm{corr}(Z_{i,j}, \mathbf{Z}_{i,-j}) = \boldsymbol{\rho}_{(j,-j)}^{(1\times J-1)}$, and $\mathrm{corr}(\mathbf{Z}_{i,-j}, \mathbf{Z}_{i,-j}) = \boldsymbol{\rho}_{(-j,-j)}^{(J-1\times J-1)}$, where $(\mathbf{X}\boldsymbol{\beta})_{-j}^{(J-1\times 1)}$ is $\mathbf{X}\boldsymbol{\beta}$ with the j^{th} element deleted. Denoting $S_r = \{i : Z_{i,j} \in A_{j,r}\}$ for $r = 0, \ldots, C_j$, we can write

$$\prod_{i=1}^{n} \mathbf{1}(Z_{i,j} \in A_{j,k_{i,j}})\, \phi(Z_{i,j} \mid m_{i,j}, \nu_{i,j}) = \prod_{r=0}^{C_j} \prod_{i \in S_r} \mathbf{1}(Z_{i,j} \in A_{i,j}) \phi(Z_{i,j} \mid m_{i,j}, \nu_{i,j})$$

and, integrating out $Z_{i,j}$,

$$f(\boldsymbol{\gamma}_j \mid \mathbf{Y}^{(n)}, \mathbf{Z}_{-j}^{(n)}, \boldsymbol{\beta}, \boldsymbol{\rho}) = \prod_{r=0}^{C_j} \prod_{i \in S_r} \left\{ \Phi\left(\frac{\gamma_{j,r+1} - m_{i,j}}{\nu_{i,j}^{1/2}} \right) - \Phi\left(\frac{\gamma_{j,r} - m_{i,j}}{\nu_{i,j}^{1/2}} \right) \right\}.$$

For Step 2, by independence

$$f(\mathbf{Z}_j \mid \mathbf{Z}_{-j}, \mathbf{Y}^{(n)}, \boldsymbol{\beta}, \boldsymbol{\rho}, \boldsymbol{\gamma}) = \prod_{i=1}^{n} f(\mathbf{Z}_{i,j} \mid \mathbf{Z}_{i,-j}, \mathbf{Y}^{(n)}, \boldsymbol{\beta}, \boldsymbol{\rho}, \boldsymbol{\gamma}).$$

Values may be obtained from this full conditional distribution by noting that it is the truncated normal with mean $m_{i,j}$ and variance $\nu_{i,j}$ as given above.

Step 3 is complicated by the requirement that $\beta_{j,1} > 0$. The full conditional distribution of $\boldsymbol{\beta}$ is $2J$-dimensional normal with mean $\tilde{\boldsymbol{\mu}}$ and covariance matrix $\tilde{\boldsymbol{\Sigma}}$ given by

$$\tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{\Sigma}}\Big(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \sum_{i=1}^{n} \mathbf{X}_i' \Omega^{-1} \mathbf{Z}_i \Big) \text{ and } \tilde{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}^{-1} + \sum_{i=1}^{n} \mathbf{X}_i' \Omega^{-1} \mathbf{X}_i.$$

Partition $\boldsymbol{\beta}$ into the intercepts $\boldsymbol{\beta}^{(0)} = (\beta_{1,0}, \ldots, \beta_{J,0})$ and truncated slopes $\boldsymbol{\beta}^{(1)} = (\beta_{1,1}, \ldots, \beta_{J,1})$. Let $\boldsymbol{\beta}_{-(j,1)}$ be the $2J-1$ dimensional subvector of $\boldsymbol{\beta}$ obtained by deleting $\beta_{j,1}$, and define $\tilde{\mu}_{-(j,1)}$ similarly. To generate the elements of $\boldsymbol{\beta}^{(1)}$, we will sample from the univariate full conditional of each truncated $\beta_{j,1}$. Denote the $(2J-1) \times (2J-1)$ submatrix of $\tilde{\Sigma}$ that is the variance-covariance matrix of $\boldsymbol{\beta}_{-(j,1)}$ by $\tilde{\Sigma}_{-[(j,1),(j,1)]}$, let $\sigma_{j,1} = \text{var}(\beta_{j,1})$, and let $\sigma_{[-(j,1),(j,1)]}$ the vector of $2J-1$ covariances of $\beta_{j,1}$ with $\boldsymbol{\beta}_{-(j,1)}$. Re-arrange the rows and columns of $\tilde{\Sigma}$ so that

$$\tilde{\Sigma} \quad = \quad \begin{bmatrix} \tilde{\Sigma}_{-[(j,1),(j,1)]} & \sigma_{[-(j,1),(j,1)]} \\ \sigma_{[-(j,1),(j,1)]} & \sigma_{j,1} \end{bmatrix}$$

Sample $[\beta_{j,1} \mid \beta_{-(j,1)}, \mathbf{Z}^{(n)}, \mathbf{Y}^{(n)}, \boldsymbol{\gamma}, \boldsymbol{\rho}]$, the j^{th} truncated slope parameter, from its univariate truncated normal, which has location parameter $\tilde{\mu}_{(j,1)} + \sigma_{[-(j,1),(j,1)]}{}' \tilde{\Sigma}^{-1}_{-[(j,1),(j,1)]} (\boldsymbol{\beta}_{-(j,1)} - \tilde{\mu}_{-(j,1)})$ and scale parameter $\sigma_{(j,1)} - \sigma_{[-(j,1),(j,1)]}{}' \tilde{\Sigma}^{-1}_{-[(j,1),(j,1)]} \sigma_{[-(j,1),(j,1)]}$.

Next, we sample from the full conditional distribution of the intercepts $\boldsymbol{\beta}^{(0)}$. Re-arrange the rows and columns of $\tilde{\Sigma}$ so that

$$\tilde{\Sigma} \quad = \quad \begin{bmatrix} \tilde{\Sigma}_{1,1} & \tilde{\Sigma}_{1,0} \\ \tilde{\Sigma}_{0,1} & \tilde{\Sigma}_{0,0} \end{bmatrix}$$

where $\tilde{\Sigma}_{1,1}$ is the variance-covariance matrix of $\boldsymbol{\beta}^{(1)}$, $\tilde{\Sigma}_{0,0}$ is the covariance matrix of $\boldsymbol{\beta}^{(0)}$, and $\tilde{\Sigma}_{1,0}$ is the cross-covariance matrix of $\boldsymbol{\beta}^{(1)}$ with $\boldsymbol{\beta}^{(0)}$. Sample $\boldsymbol{\beta}^{(0)}$ from the conditional J-variate normal distribution of $[\boldsymbol{\beta}^{(0)} \mid \boldsymbol{\beta}^{(1)}, \mathbf{Z}^{(n)}, \mathbf{Y}^{(n)}, \boldsymbol{\gamma}, \boldsymbol{\rho}]$, which has mean vector $\tilde{\boldsymbol{\mu}}^{(0)} + \tilde{\Sigma}_{0,1} \tilde{\Sigma}^{-1}_{1,1} (\boldsymbol{\beta}^{(1)} - \tilde{\boldsymbol{\mu}}^{(1)})$ and variance-covariance matrix $\tilde{\Sigma}_{(0,0)}$ - $\tilde{\Sigma}_{0,1} \tilde{\Sigma}^{-1}_{1,1} \tilde{\Sigma}_{1,0}$.

Finally, using the fact that $f(\boldsymbol{\rho} \mid \mathbf{Z}^{(n)}, \mathbf{Y}^{(n)}, \boldsymbol{\gamma}, \boldsymbol{\beta}) \;\propto f(\boldsymbol{\rho}) \prod_{i=1}^{n} \phi(\mathbf{Z}_i \mid \mathbf{X}_i \boldsymbol{\beta}, \boldsymbol{\rho})$, we use an independence chain Metropolis-Hastings step to sample $\boldsymbol{\rho}$, following the suggestion of Chib and Greenberg (1998) by applying the algorithm to blocks of size 1 on the off-diagonal of $\boldsymbol{\Omega}$.

Table 1. *Toxicities and severity weights in the sarcoma trial*

Type of Toxicity	Grade	Severity Weight
Myelosuppression w/o fever	3	1.0
	4	1.5
Myelosuppression with fever	3	5.0
	4	6.0
Dermatitis	3	2.5
	4	6.0
Liver	2	2.0
	3	3.0
	4	6.0
Nausea/Vomiting	3	1.5
	4	2.0
Fatigue	3	0.5
	4	1.0

Table 2. *Case-by-case illustration of the method in the sarcoma trial*

Patient	Dose	Myelosupp.	Dermatitis	Liver	Fatigue	Nausea	TTB
1	400	Gr. 3 w/o Fev.	Gr. 3	None	None	Gr. 3	5.0
2	400	Gr. 3 w/o Fev.	None	None	None	None	1.0
3	400	Gr. 3 w/o Fev.	None	None	None	None	1.0
4	400	None	None	None	None	None	0
5	500	None	Gr. 3	None	None	None	2.5
6	500	None	None	None	None	None	0
7	500	Gr. 3 w/o Fev.	Gr. 3	None	None	None	3.5
8	500	Gr. 4 w/o Fev.	None	Gr. 2	None	None	3.5
9	600	Gr. 3 w/o Fever	None	None	None	Gr. 3	2.5
10	600	Gr. 4 w/o Fever	None	None	None	None	1.5
11	600	None	Gr. 3	None	Gr. 3	None	3.0
12	600	None	Gr. 3	None	Gr. 3	None	3.0
13	700	Gr. 3 w. Fev.	None	Gr. 2	None	Gr. 3	8.5
14	700	Gr. 3 w/o Fev.	Gr. 3	None	Gr. 3	None	4.0
15	700	None	Gr. 3	None	None	Gr. 3	4.0
16	700	None	None	None	Gr. 3	None	0.5
17	600	Gr. 3 w/o Fev.	None	None	None	Gr. 3	2.5
18	600	Gr. 3 w/o Fev.	None	None	None	None	1.0
19	600	None	Gr. 3	Gr. 2	Gr. 3	None	5.0
20	600	None	Gr. 3	None	Gr. 3	None	3.0

Table 2. *(Continued)*

Patient	Dose	Myelosupp.	Dermatitis	Liver	Fatigue	Nausea	TTB
21	600	None	None	None	None	None	0
22	600	None	Gr. 3	None	None	None	2.5
23	600	Gr. 4 w/o Fev.	None	None	None	None	1.5
24	600	Gr. 4 w/o Fev.	None	None	None	None	1.5
25	700	Gr. 3 w/o Fev.	None	Gr. 2	None	None	3.0
26	700	Gr. 3 w/o Fev.	None	None	None	None	1.0
27	700	Gr. 3 w/o Fev.	None	None	None	Gr.4	3.0
28	700	Gr. 3 w/o Fev.	None	None	None	None	1.0
29	700	Gr. 3 with Fev.	None	None	None	None	5.0
30	700	Gr. 3 w/o Fev.	Gr. 3	None	None	Gr. 3	5.0
31	700	None	None	None	Gr. 3	None	0.5
32	700	None	None	None	Gr. 3	None	0.5
33	700	Gr. 4 w/o Fever	None	Gr. 2	None	Gr. 3	5.0
34	700	Gr. 3 w/o Fever	Gr. 3	None	None	None	3.5
35	700	None	None	None	None	None	0
36	700	None	None	Gr. 3	None	None	3.0

Table 3. *Posterior mean total toxicity burdens, by dose, for the illustrative trial summarized in Table 2.*

	Dose Level									
	100	200	300	400	500	600	700	800	900	1000
Prior	1.05	1.55	2.06	2.61	3.18	3.76	4.35	4.91	5.46	5.96
After Cohort 1	0.56	0.93	1.35	1.81	2.29	2.78	3.24	3.68	4.08	4.45
After Cohort 2	0.53	0.90	1.34	1.83	2.34	2.85	3.34	3.79	4.20	4.57
After Cohort 3	0.53	0.90	1.32	1.78	2.27	2.75	3.22	3.66	4.06	4.42
After Cohort 4	0.54	0.92	1.35	1.85	2.38	2.92	3.45	3.95	4.41	4.83
After Cohort 5	0.57	0.95	1.39	1.89	2.43	2.97	3.50	3.99	4.44	4.84
After Cohort 6	0.55	0.90	1.29	1.74	2.22	2.72	3.21	3.68	4.11	4.51
After Cohort 7	0.59	0.91	1.26	1.65	2.08	2.53	3.00	3.45	3.88	4.27
After Cohort 8	0.57	0.89	1.24	1.63	2.06	2.51	2.98	3.43	3.87	4.27
After Cohort 9	0.57	0.88	1.22	1.60	2.03	2.49	2.97	3.45	3.91	4.34

Table 4. *Simulation results for the sarcoma trial under nine scenarios. LS = low severity toxicities, HS = high severity toxicities, Psel = % selected, Npats = number of patients treated.*

Main Toxicities		Gemcitabine Dose, mg/m^2									
		100	200	300	400	500	600	700	800	900	1000
					$\psi^* = 3.04$ at 200 mg/m^2						
LS	Psel	1.6	**93.7**	4.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Npats	2.2	21.8	5.3	5.6	1.1	0.0	0.0	0.0	0.0	0.0
HS	Psel	4.1	**92.0**	3.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Npats	4.5	22.0	4.5	4.6	0.4	0.0	0.0	0.0	0.0	0.0
					$\psi^* = 3.04$ at 500 mg/m^2						
LS	Psel	0.0	0.0	0.0	5.6	**85.4**	9.0	0.0	0.0	0.0	0.0
	Npats	0.0	0.0	0.0	5.6	18.5	9.7	2.1	0.1	0.0	0.0
HS	Psel	0.0	0.1	0.0	5.2	**86.9**	7.8	0.0	0.0	0.0	0.0
	Npats	0.0	0.0	0.1	6.6	20.6	7.8	1.0	0.0	0.0	0.0
					$\psi^* = 3.04$ at 800 mg/m^2						
LS	Psel	0.0	0.0	0.0	0.0	0.0	0.3	10.4	**80.7**	8.3	0.2
	Npats	0.0	0.0	0.0	4.0	4.0	4.0	5.6	11.4	5.9	1.0
HS	Psel	0.0	0.0	0.0	0.0	0.0	0.1	13.3	**80.3**	6.2	0.0
	Npats	0.0	0.0	0.0	4.0	4.0	4.0	5.7	11.7	5.8	0.7

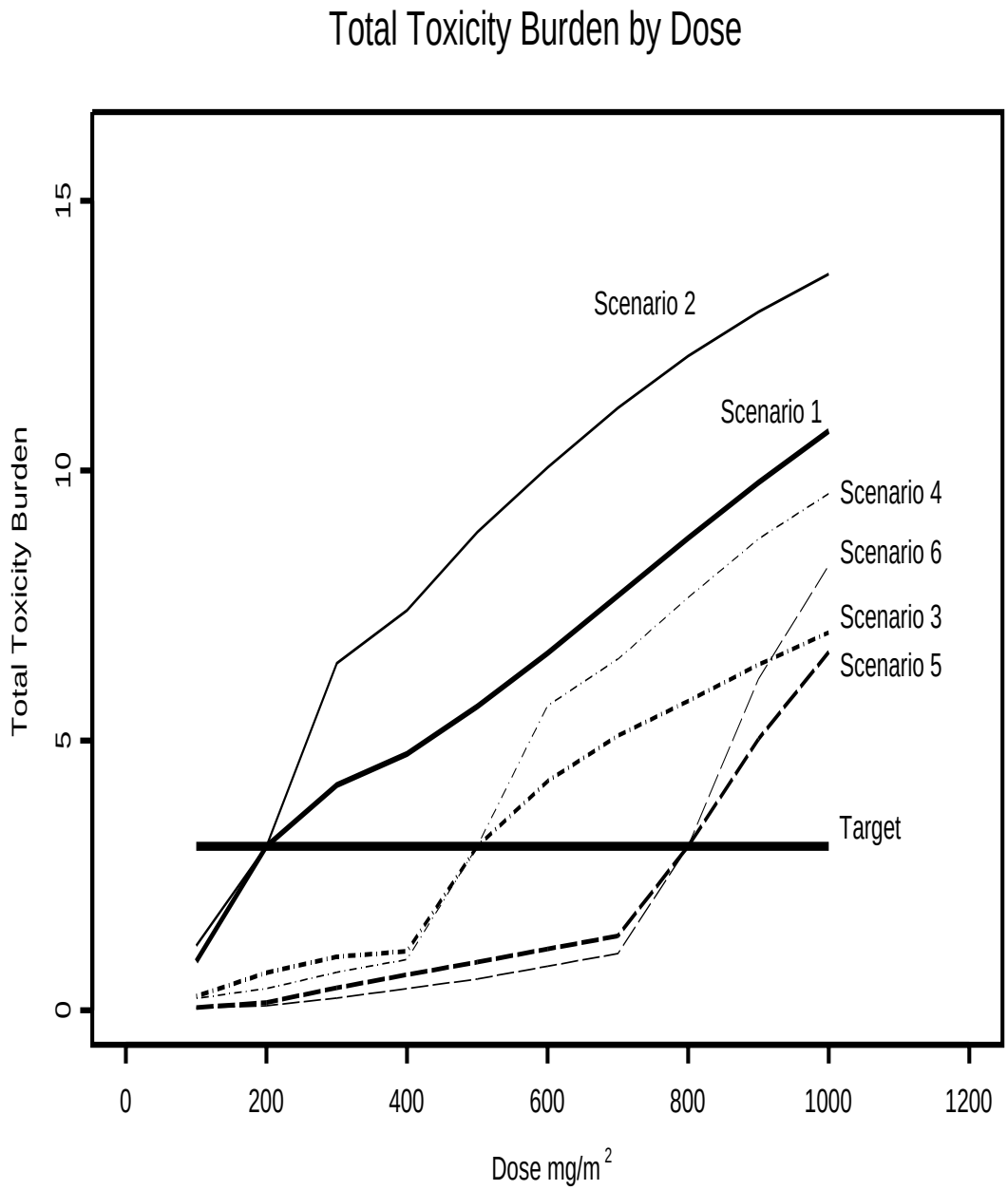

Figure 1: Total toxicity burden as a function of dose under each of the six dose-toxicity scenarios considered in the simulation study.

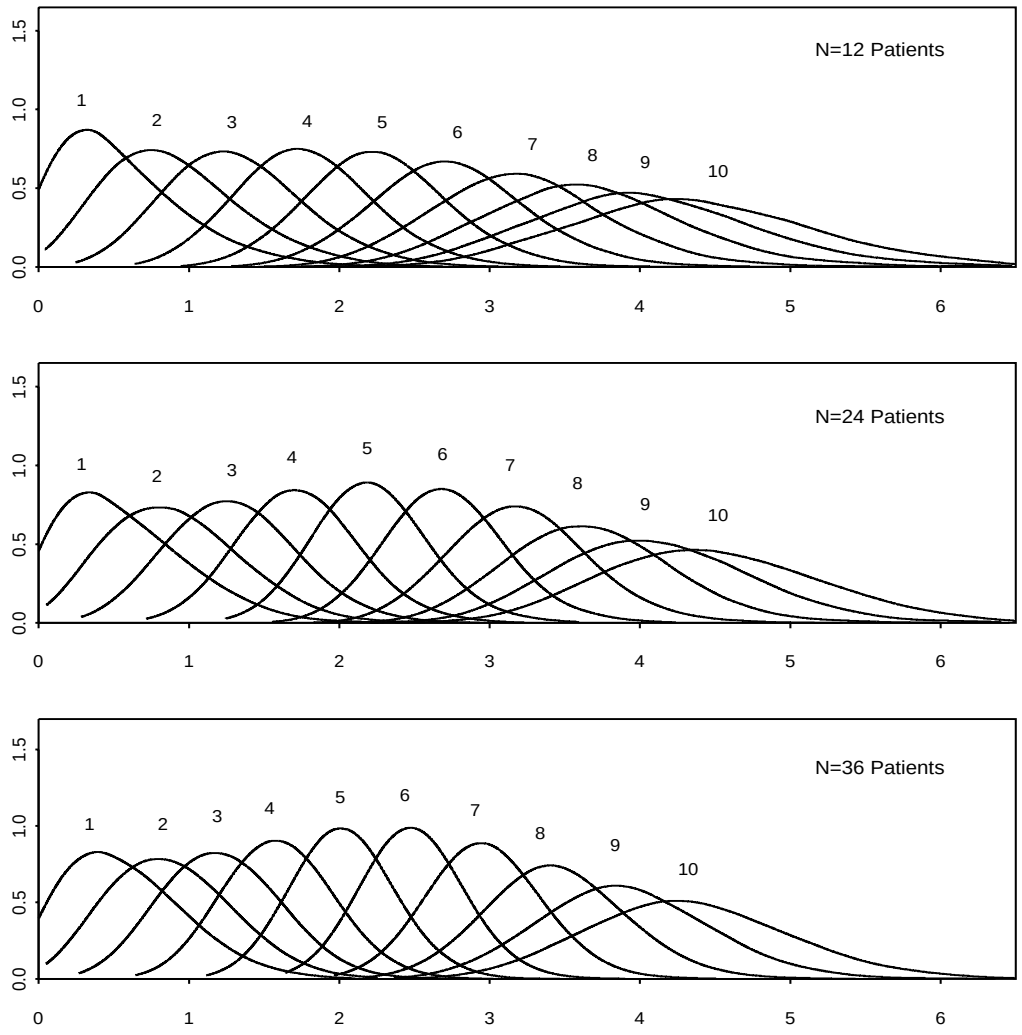

Figure 2: Posterior distribution of the total toxicity burden for each dose, after observing the data from the first 12, 24, and 36 patients in the illustrative trial summarized in Table 1.