University of Michigan School of Public Health
The University of Michigan Department of Biostatistics Working Paper Series

Year 2006 *Paper* 63

Simultaneously Optimizing Dose and Schedule of a New Cytotoxic Agent

Thomas M. Braun* Peter F. Thall†
Hoang Nguyen‡ Marcos de Lima**

*University of Michigan, tombraun@umich.edu
†University of Texas M.D. Anderson Cancer Center
‡University of Texas M.D. Anderson Cancer Center
**University of Texas M.D. Anderson Cancer Center

http://biostats.bepress.com/umichbiostat/paper63

Simultaneously Optimizing Dose and Schedule of a New Cytotoxic Agent

Thomas M. Braun, Peter F. Thall, Hoang Nguyen, and Marcos de Lima

Abstract

Traditionally, phase I clinical trial designs determine a maximum tolerated dose of an experimental cytotoxic agent based on a fixed schedule, usually one course consisting of multiple administrations, while varying the dose per administration between patients. However, in actual medical practice patients often receive several courses of treatment, and some patients may receive one or more dose reductions due to low-grade (non-dose limiting) toxicity in previous courses. As a result, the overall risk of toxicity for each patient is a function of both the schedule and the dose used at each adminstration. We propose a new paradigm for Phase I clinical trials that allows both the dose per administration and the schedule to vary, making treatment two-dimensional. We provide an outcome-adaptive Bayesian design that simultaneously optimizes both dose and schedule in terms of the overall risk of toxicity, based on time-to-toxicity outcomes. The method is illustrated with a trial of an agent hypothesized to prolong cancer remission after allogeneic bone marrow transplantation, and a simulation study in the context of this trial is presented.

Simultaneously Optimizing Dose and Schedule of a New Cytotoxic Agent

Thomas M. Braun,[1,*] **Peter F. Thall,**[2] **Hoang Nguyen**[2] **and Marcos de Lima**[3]

[1]Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI

[2]Department of Biostatistics and Applied Mathematics
The University of Texas, M.D. Anderson Cancer Center, Houston, TX

[3]Department of Blood and Marrow Transplantation
The University of Texas, M.D. Anderson Cancer Center, Houston, TX

July 27, 2006

SUMMARY. Traditionally, phase I clinical trial designs determine a maximum tolerated dose of an experimental cytotoxic agent based on a fixed schedule, usually one course consisting of multiple administrations, while varying the dose per administration between patients. However, in actual medical practice patients often receive several courses of treatment, and some patients may receive one or more dose reductions due to low-grade (non-dose limiting) toxicity in previous courses. As a result, the overall risk of toxicity for each patient is a function of both the schedule and the dose used at each adminstration. We propose a new paradigm for Phase I clinical trials that allows both the dose per administration and the schedule to vary, making treatment two-dimensional. We provide an outcome-adaptive Bayesian design that simultaneously optimizes both dose and schedule in terms

[*] *email:* tombraun@umich.edu; *phone*: 734-936-9844; *fax*: 734-763-2215

of the overall risk of toxicity, based on time-to-toxicity outcomes. The method is illustrated with a trial of an agent hypothesized to prolong cancer remission after allogeneic bone marrow transplantation, and a simulation study in the context of this trial is presented.

KEY WORDS: Adaptive design; phase I trial; bone marrow transplantation; dose escalation; CRM; maximum tolerated dose; hypomethylation; chromatin

1. Introduction

Conventional phase I clinical trials determine the maximum tolerated dose (MTD) of a new agent by characterizing patient outcome as a binary indicator of whether toxicity occurs within a short time period from the start of therapy. Generally, the MTD is the highest dose that does not present a practical limitation to therapy (Storer, 1989; Goodman et al., 1995; Babb et al., 1998). This approach has seen widespread use largely because it facilitates adaptive dose-finding methods that successively use the doses and outcomes of previous patients to select doses for new patients. A limitation of these methods is that they typically base dose-finding on a single course of therapy, whereas multiple courses typically are used in medical practice. As a result, the MTD based on a single course of treatment may prove to be overly toxic when given over multiple courses. For example, if conventional dose-finding is done with a fixed schedule consisting of one course when in fact a safe dose d^* exists with three courses and this combination has substantive anti-disease effect whereas d^* with only one course does not, then the conventional MTD of one course may lead to the erroneous conclusion in later studies that the agent is ineffective. Similarly, if conventional dose-finding is done with four courses and it turns out that the lowest dose is excessively toxic, then it may be concluded erroneously that the agent is unsafe at any dose simply because shorter schedules were not examined.

Recently, a new phase I method was proposed that determines a maximum

tolerated schedule (MTS), rather than a conventional MTD (Braun et al., 2005). The MTS is defined as the maximum number of courses that can be given without causing unacceptable cumulative toxicity. The model and method account for the patient's sequence of administrations and allows the number of courses to vary so that an optimal schedule may be determined. However, while this method allows the number of courses to vary, it requires the dose used in each administration to be fixed. Thus, if the fixed dose is ill-chosen, the MTS may be far from optimal. One may easily imagine examples similar to those given above by switching the roles of dose and schedule.

This paper is motivated by the problems, noted above, that arise in phase I trials when either the schedule is fixed and a MTD is found, or the per-administration dose is fixed and a MTS is found. We propose a new paradigm for Phase I clinical trials that simultaneously optimizes both the dose per adminstration and the overall schedule. The design examines a matrix of possible (dose, schedule) combinations. Each patient is assigned a combination using previous patients' data, with decision criteria based on the posterior under a Bayesian model using time-to-toxicity as the outcome. The goal is to determine a maximum tolerated dose and schedule (MTDS) in terms of the overall risk of toxicity. Our formulation allows both the dose and the timing of each administration to vary between patients. This accommodates settings where a patient's dose per administration is decreased if a low grade toxicity is observed, and we also allow a patient's actual doses or administration times to deviate from planned values due to logistical difficulties or human error. Consequently, although the design examines a predetermined matrix of (dose, schedule) combinations, the model allows each patient's treatment to consist of an arbitrary sequence of administration times and a corresponding sequence of doses, so that the likelihood reflects the actual data in the trial.

Section 2 describes the trial that motivated this research and that will be used for illustration. Section 3 presents notation and probability models, including

methods for eliciting and calibrating priors. Section 4 provides criteria for evaluating (dose,schedule) pairs and rules for trial conduct. Section 5 illustrates the method via simulations as applied to an allogeneic cell transplantation trial, and we conclude with a discussion in Section 6.

2. Motivating Example

In allogeneic blood or bone marrow cell transplantation (allotx) for treatment of leukemia, a patient (host) receives cells (the graft) from a donor who has been matched on a number of human leukocyte antigen sites. The graft contains T-cells and natural killer cells that coordinate a positive immune response that kills leukemia cells, called a graft-versus-leukemia (GVL) effect. However, allotx recipients who initially respond to treatment have a substantial risk of disease recurrence due to proliferation of residual leukemia cells. As a result, investigators continue to seek agents that can be given to allotx recipients after they achieve a response in order to reduce the risk of disease recurrence.

Epigenetic DNA changes are reversible modifications of the DNA-histone complex that do not require alterations in nucleotide sequences (Das and Singal, 2004). Addition of a methyl group to gene promoter areas (DNA methylation) is associated with gene silencing, and abnormal methylation patterns are commonly seen in cancer cells. Hypermethylation of promoter regions appears to suppress genes involved in leukemic cell growth. Methylation is maintained by the enzyme cytosine DNA methyltransferase, with inhibition of this enzyme leading to hypomethylation and subsequent reactivation of tumor suppressor genes. Vidaza® (5-azacitidine) inibits DNA methyltransferase by forming covalent adducts with the enzyme, activating silent genes that may lead to cell death, and also may induce phenotypic modification of the leukemic cells to facilitate immune recognition and potentiation of the donor cells' GVL effect. Vidaza® has been approved by the U.S. Food and Drug Administration for the treatment of myelodysplastic syndrome (MDS), a blood cell disease that often progresses to acute myelogenous

leukemia (AML). The recommended dose and schedule for MDS patients is 75 mg/m^2 given subcutaneously, daily for seven days, with this seven-day cycle repeated every four weeks. No data exist, however, on what a safe dose and schedule for AML patients might be.

The method described here was motivated by the desire to design a phase I trial to optimize both the schedule and the dose per administration of Vidaza® in AML. The trial currently is ongoing at M.D. Anderson Cancer Center. For the purpose of determining an optimal (dose,schedule) pair, "toxicity" is defined as any of the following adverse events (AEs): (1) severe (grade 3 or 4) toxicity of the kidney, liver, heart or lung, or neural toxicity, as defined by standard NCI grading criteria; (2) severe graft-versus-host disease; (3) systemic infection that cannot be resolved by antibiotics within 2 weeks; (4) severe hematologic toxicity, with thrombocytopenia and or neutropenia or (5) an AE of any of these types that leads to subsequent delay or termination of therapy, or a dose reduction. Each patient may receive up to four courses of therapy, and the dose may be reduced up to two times for reasons other than the toxicities listed above. Thus, a patient's treatment may consist of an initial dose and the times at which it was administered, a second, possibly lower dose, along with its administration times, and so on, up to four courses. In practice, a patient's administration times may deviate from the planned schedule due to practical difficulties in adhering to the schedule over several months of therapy, or intentional delay of a planned course by the physician to allow a patient to recover from a low grade toxicity. Additionally, a patient may receive the wrong dose due to human error. Thus, an important feature of our model is that it accommodates each patient's actual treatment sequence by accounting for the contribution of each dose and its time of administration to the patient's overall risk of toxicity.

3. Probability Model

3.1 *General Form of the Hazard and Likelihood*

Suppose one wishes to evaluate J doses, $d_1 < d_2 < \cdots < d_J$, and K nested schedules, $\boldsymbol{s}^{(1)}, \cdots, \boldsymbol{s}^{(K)}$. The k^{th} schedule is a sequence of administration times, $\boldsymbol{s}^{(k)} = (s_1, s_2, \ldots, s_{m^{(k)}})$, with $\boldsymbol{s}^{(k)}$ a subsequence of $\boldsymbol{s}^{(k+1)}$ for each $k = 1, \ldots, K-1$, and $m^{(1)} < m^{(2)} \cdots < m^{(K)}$. Here, "dose" is the amount of the agent given at each administration. For example, a patient given d_2 under schedule $\boldsymbol{s}^{(3)} = (s_1, s_2, \ldots, s_{m^{(3)}})$ receives the cumulative amount $d_2\, m^{(3)}$ of the agent in $m^{(3)}$ successive administrations, unless therapy is terminated early due to toxicity. Thus, a patient's assigned treatment is indexed by the pair (j, k), representing $(d_j, \boldsymbol{s}^{(k)})$, there are $M = JK$ such pairs under consideration, and the total amount of the agent given to the patient increases with both dose and schedule.

In the motivating study, there are three doses of interest: 8, 16 and 24 mg/m^2, and four schedules, for a total of $M = 12$ combinations. One course consists of 5 consecutive daily administrations. Ideally, the first course begins 40 days post-transplant, although this may vary since the physician may decide to delay administration due to early complications, such as infection. We thus define the time to toxicity from the time when the first course is actually begun, the patient's enrollment time. The first schedule, $s^{(1)} = (0, 1, 2, 3, 4)$, consists of one course. The second schedule includes one additional course starting 28 days after the beginning of $s^{(1)}$, so that $s^{(2)} = (0, 1, 2, 3, 4, 28, 29, 30, 31, 32) = (s^{(1)}, s^{(1)} + 28)$. The third and fourth schedules are defined similarly, with $s^{(3)} = (s^{(1)}, s^{(1)} + 28, s^{(1)} + 56)$ and $s^{(4)} = (s^{(1)}, s^{(1)} + 28, s^{(1)} + 56, s^{(1)} + 84)$. Figure 1 provides a schematic representation of the 12 (dose,schedule) combinations evaluated in the trial. We denote the maximum length of follow-up for each patient specified by the investigators by τ, which should be large enough to include toxicities arising from the longest schedule, $\boldsymbol{s}^{(K)}$. In the motivating trial, τ = 116 days, which is 28 days after the start of the fourth course.

[Figure 1 about here.]

Our model generalizes that used by Braun et al. (2005) by extending it to allow the dose per administration to vary and also using a new parameterization to facilitate computation. While other models are possible, we use this model here because it is robust (section 6.4, Table 4, below) and because our primary focus is the new algorithm for optimizing (dose,schedule). Let t^* denote a time, from the start of the trial, when one evaluates the data and either assigns a particular pair (j, k) to the next patient or terminates the trial early if no pair is acceptable. Denote by e the study time when the patient's therapy begins. We denote the time to toxicity by Y and let Y^o be the patient's observed time from e to either toxicity or last follow up at study time t^*. Thus, $Y^o = Y$ if $e + Y \leq t^*$ or $Y^o = t^* - e$ if $e + Y > t^*$. Let $\delta = I(Y^o = Y)$ indicate that the patient has toxicity by study time t^*. Let $h(u \mid \boldsymbol{\theta}, d)$ denote the hazard of toxicity associated with a single administration of dose d of the agent given u days previously, where $\boldsymbol{\theta}$ is a vector of model parameters; we define $h(u \mid \boldsymbol{\theta}, d) = 0$ for $u < 0$. Let $\boldsymbol{s} = (s_1, \cdots, s_k)$ denote the patient's sequence of administration times and $\mathbf{d}_{\boldsymbol{S}} = (d_{j(s_1)}, \cdots, d_{j(s_k)})$ the corresponding doses up to study time t^*. Thus, $j(s_\ell)$ indexes the dose given to the patient at time s_ℓ after entry, at study time $e + s_\ell$. The overall hazard of toxicity at study time t^* for a patient treated with schedule $\boldsymbol{s}$ and doses $\boldsymbol{d}_{\boldsymbol{S}}$ is $\lambda(t^* \mid \boldsymbol{\theta}, \boldsymbol{s}, \mathbf{d}_{\boldsymbol{S}}) = \sum_{\ell=1}^{k} h(t^* - e - s_\ell \mid \boldsymbol{\theta}, d_{j(s_\ell)})$. Consequently, the patient's cumulative hazard function at t^* is $\Lambda(t^* \mid \boldsymbol{\theta}, \boldsymbol{s}, \mathbf{d}_{\boldsymbol{S}}) = \sum_{\ell=1}^{k} H(t^* - e - s_\ell \mid \boldsymbol{\theta}, d_{j(s_\ell)})$, where $H(x \mid \boldsymbol{\theta}, d) = \int_0^x h(u \mid \boldsymbol{\theta}, d)du$, with survivor function $\Pr(Y > t^* \mid \boldsymbol{\theta}, \boldsymbol{s}, \mathbf{d}_{\boldsymbol{S}}) = \bar{F}(t^* \mid \boldsymbol{\theta}, \boldsymbol{s}, \mathbf{d}_{\boldsymbol{S}}) = \exp\{-\Lambda(t^* \mid \boldsymbol{\theta}, \boldsymbol{s}, \mathbf{d}_{\boldsymbol{S}})\}$ and density $f(t^* \mid \boldsymbol{\theta}, \boldsymbol{s}, \mathbf{d}_{\boldsymbol{S}}) = \lambda(t^* \mid \boldsymbol{\theta}, \boldsymbol{s}, \mathbf{d}_{\boldsymbol{S}})\bar{F}(t^* \mid \boldsymbol{\theta}, \boldsymbol{s}, \mathbf{d}_{\boldsymbol{S}})$. Let n^* denote the number of patients enrolled up to t^*. For the i^{th} patient, e_i is the entry time, the sequence of administration times up to t^* is $e_i + \boldsymbol{s}_i = (e_i + s_{i,1}, \cdots, e_i + s_{i,k_i})$, and $\mathbf{d}_{\boldsymbol{S}_i} = (d_{j(s_{i,1})}, \cdots, d_{j(s_{i,k_i})})$ is the corresponding sequence of doses, $i = 1, \cdots, n^*$. For treatment sequences $(\boldsymbol{s}_i, \mathbf{d}_{\boldsymbol{S}_i})$ and outcome data (Y_i^o, δ_i), patient i has likelihood

$L_i(\boldsymbol{\theta}|Y_i^o, \delta_i, \boldsymbol{s}_i, \mathbf{d}_{\boldsymbol{S}_i}) = \lambda(Y_i^o \mid \boldsymbol{\theta}, \boldsymbol{s}_i, \mathbf{d}_{\boldsymbol{S}_i})^{\delta_i} \bar{F}(Y_i^o \mid \boldsymbol{\theta}, \boldsymbol{s}_i, \mathbf{d}_{\boldsymbol{S}_i})$,, and the overall likelihood at t^* is $\mathcal{L}(\boldsymbol{\theta} \mid \text{data}_{n^*}) = \prod_{i=1}^{n^*} L_i(\boldsymbol{\theta}|Y_i^o, \delta_i, \boldsymbol{s}_i, \mathbf{d}_{\boldsymbol{S}_i})$.

This model accommodates each patient's *actual* sequence of administration times and doses, which often deviate from his/her planned treatment. For example, suppose a patient's planned treatment was two courses with 24 mg/m^2 at each of the 10 administrations, but the patient began the first course two days late, was reduced to 8 mg/m^2 in the second course due to a grade 2 infection, and was given 16 mg/m^2 by mistake at the tenth administration. Then the patient's actual treatment would be $\boldsymbol{s} = (2, 3, 4, 5, 6, 30, 31, 32, 33, 34)$ and $\mathbf{d}_{\boldsymbol{S}} = (24, 24, 24, 24, 24, 8, 8, 8, 8, 16)$. While this is not any of the 12 (dose, schedule) combinations being studied in the Vidaza® trial (Figure 1), the model allows this patient's data to be included in the likelihood.

It may be unclear how to score Y for some patients. The definition of toxicity in the Vidaza® trial includes grade 2 toxicities that cannot be resolved therapeutically within 2 weeks from onset or that necessitate a dose reduction. We chose to score such toxicities as occurring at the time of initial onset. For example, if a patient has a grade 2 thrombocytopenia starting at day 10 of therapy that persists beyond day 24 and requires a dose reduction, we define $Y^o = 10$ and $\delta = 1$. However, if the thrombocytopenia is resolved by day 24, then it is not scored as a toxicity, and at day 24 we define $Y^o = 24$ and $\delta = 0$, provided that no other toxicity has occurred. This approach is conservative in that toxicity is assumed to have occurred as soon as possible.

3.2 *Single Administration Hazard Function*

The probability model is determined by the particular form of h, for which we employ a reparameterized version of the triangular hazard function used by Braun *et al.* (2005). In this model, for each per-administration dose $j = 1, \cdots, J$, denote $\boldsymbol{\theta}_j = (a_j, b_j, c_j)$, with $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_J\}$ and each entry of $\boldsymbol{\theta}$ positive-valued. The

hazard of toxicity associated with a single administration of dose d_j is

$$h(u \mid \boldsymbol{\theta}_j) = \begin{cases} \frac{2a_j}{b_j+c_j}\ \frac{u}{b_j} & 0 \le u \le b_j \\ \frac{2a_j}{b_j+c_j}\ \frac{b_j+c_j-u}{c_j} & b_j < u \le b_j + c_j \\ 0 & u > b_j + c_j \ \text{ or } \ u < 0. \end{cases} \quad (1)$$

This is a triangle having base of length $b_j + c_j$ and area equal to a_j, with the height of the triangle, $2a_j/(b_j + c_j)$, occurring at $u = b_j$. The area of this triangle is the cumulative single-administration hazard $H_j = a_j$ for dose d_j. The constraint that the per-administration cumulative hazard of toxicity increases with dose says that $a_1 < a_2 < \cdots < a_J$. In some applications, however, the risk of toxicity may reach a plateau or may even decrease with a higher dose or longer schedule, such as in studies of anti-infection agents that have adverse effects, and such an ordering constraint is inappropriate.

3.3 *Establishing Priors*

To enforce the constraint $a_1 < a_2 < \ldots < a_J$ so that $F(\tau \mid \boldsymbol{\theta}, d_j, \boldsymbol{s}^{(k)})$, the probability of toxicity by τ, increases with dose, we let $a_j^* = a_j - a_{j-1}$ for $j \ge 2$, with $a_1^* = a_1$. We assume that $(a_1^*, \cdots, a_J^*)$ follows a J-variate lognormal prior with all correlations equal to zero, although *a posteriori* the a_j^*'s may be correlated. Denoting the marginals by $a_j^* \sim LN(\mu_{a_j^*}, \sigma_a^2)$, this implies that $\mathrm{E}(a_j^*) = \exp(\mu_{a_j^*} + \sigma_a^2/2)$ and $\mathrm{var}(a_j^*) = \exp(2\mu_{a_j^*} + \sigma_a^2)\{\exp(\sigma_a^2) - 1\}$. Similarly, $(b_1, \cdots, b_J)$ and $(c_1, \cdots, c_J)$ each follow J-variate lognormal priors, with marginals $b_j \sim LN(\mu_{b_j}, \sigma_b^2)$ and $c_j \sim LN(\mu_{c_j}, \sigma_c^2)$ for each j. We chose the multivariate lognormal for its generality and tractability. Denoting $\boldsymbol{\mu}_j = (\mu_{a_j^*}, \mu_{b_j}, \mu_{c_j})$ and $\boldsymbol{\sigma}^2 = (\sigma_a^2, \sigma_b^2, \sigma_c^2)$, the model has $3J$ location hyperparameters, $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \cdots, \boldsymbol{\mu}_J)$ and 3 scale hyperparameters, $\boldsymbol{\sigma}^2 = (\sigma_a^2, \sigma_b^2, \sigma_c^2)$; we denote $\widetilde{\boldsymbol{\theta}} = (\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$. In Section 5, we will illustrate how to determine $\widetilde{\boldsymbol{\theta}}$ in the context of the application.

Appropriate values for the prior mean and variance parameters may be elicited from the investigators in many ways (Gelman et al., 2004), although in general, it is easiest to elicit values on domains with which the investigator is familiar

(Tsutakawa and Lin, 1986). Thus, we elicit the *expected* values of the following three quantities for each dose $j = 1, \cdots, J$:

(1) $\xi_{j,1} = -\log\{1 - F_j(\tau \mid \boldsymbol{s}^{(1)}, \boldsymbol{\theta})\} = m^{(1)} a_j$, the transformed probability of toxicity by time τ under the shortest schedule $\boldsymbol{s}^{(1)}$

(2) $\xi_{j,2}$, the time until the maximum hazard is reached for one administration of dose j, and

(3) $\xi_{j,3}$, the time from the peak of the hazard until the hazard vanishes completely or becomes negligible for a single administration of dose j .

In practice, one elicits the mean probability $E\{F_j(\tau \mid \boldsymbol{s}^{(1)}, \boldsymbol{\theta}) \mid \widetilde{\boldsymbol{\theta}}\}$ and then derives $E\{\xi_{j,1} \mid \widetilde{\boldsymbol{\theta}}\} \approx -\log[1 - E\{F_j(\tau \mid \boldsymbol{s}^{(1)}, \boldsymbol{\theta}) \mid \widetilde{\boldsymbol{\theta}}\}]$. We will use the superscript (e) to denote elicited values, with $\xi_{j,\ell}^{(e)}$ the elicited mean of $\xi_{j,\ell}, \ell = 1, 2, 3$.

We derive additional functions of the hyperparameter vector $\widetilde{\boldsymbol{\theta}}$ by assuming that $\xi_{j,1}$ has an inverse Gamma (IG) distribution with variance $(\xi_{j,1}^{(e)})^2/(\nu_1 - 1)$, and that $\xi_{j,2}$ and $\xi_{j,3}$ have IG distributions with respective variances $(\xi_{j,2}^{(e)})^2/(\nu_2 - 1)$ and $(\xi_{j,3}^{(e)})^2/(\nu_2 - 1)$. We use the same value ν_2 in the distributions of $\xi_{j,2}$ and $\xi_{j,3}$ as they both describe time durations and differ in nature from $\xi_{j,1}$. Since $\xi_{j,1}, \xi_{j,2}$, and $\xi_{j,3}$ correspond respectively to $m^{(1)} a_j$, b_j, and c_j, we use the following method-of-moments approach to solve for $\widetilde{\boldsymbol{\theta}}$ by equating the *elicited* moments of $\xi_{j,1}, \xi_{j,2}$, and $\xi_{j,3}$ to their corresponding *theoretical* moments. Denoting $\xi_{0,1}^{(e)} = 0$, we solve the following set of equations for $\widetilde{\boldsymbol{\theta}}$:

$$\begin{aligned}
\exp(\mu_{a_j^*} + \sigma_a^2/2) &= (\xi_{j,1}^{(e)} - \xi_{j-1,1}^{(e)})/m^{(1)} && (2)\\
\exp(2\mu_{a_j^*} + \sigma_a^2)\{\exp(\sigma_a^2) - 1\} &= [(\xi_{j,1}^{(e)} - \xi_{j-1,1}^{(e)})/m^{(1)}]^2/(\nu_1 - 1) && (3)\\
\exp(\mu_{b_j} + \sigma_b^2/2) &= \xi_{j,2}^{(e)} && (4)\\
\exp(2\mu_{b_j} + \sigma_b^2)\{\exp(\sigma_b^2) - 1\} &= (\xi_{j,2}^{(e)})^2/(\nu_2 - 1) && (5)\\
\exp(\mu_{c_j} + \sigma_c^2/2) &= \xi_{j,3}^{(e)} && (6)\\
\exp(2\mu_{c_j} + \sigma_c^2)\{\exp(\sigma_c^2) - 1\} &= (\xi_{j,3}^{(e)})^2/(\nu_2 - 1). && (7)
\end{aligned}$$

Comparing Equations (2) and (3) shows that $\exp(\sigma_a^2) - 1 = (\nu_1 - 1)^{-1}$, hence $\sigma_a^2 = \log\{\nu_1/(\nu_1 - 1)\}$, which is a function solely of ν_1. Comparing Equations (4)-(5) and Equations (6)-(7), shows that $\sigma_b^2 = \sigma_c^2 = \log\{\nu_2/(\nu_2 - 1)\}$. Thus, ν_1 and ν_2 are tuning parameters that determine the informativeness of the prior. Solving for the prior location parameters gives $\mu_{a_j^*} = \log([\xi_{j,1}^{(e)} - \xi_{j-1,1}^{(e)}]/m^{(1)}) - \sigma_a^2/2$, μ_{b_j} = $\log(\xi_{j,2}^{(e)}) - \sigma_b^2/2$ and μ_{c_j} = $\log(\xi_{j,3}^{(e)}) - \sigma_c^2/2$. Thus, $\boldsymbol{\sigma}^2$ is determined by the tuning parameters ν_1 and ν_2, and given $\boldsymbol{\sigma}^2$, the elicited quantities are used to determine the prior location parameters $\boldsymbol{\mu}$.

4. Choosing (Dose, Schedule) Combinations

Because trial conduct uses decision criteria based on the most recent posterior computed when a new patient is accrued, the data must be monitored continuously, making the design computationally intensive. We compute the posterior of $\boldsymbol{\theta}$ using Markov Chain Monte Carlo methods as described in the Appendix. For each (j, k), we base decisions on $F_{jk}(\boldsymbol{\theta}) = F(\tau \mid \boldsymbol{\theta}, d_j, \boldsymbol{s}^{(k)})$, the cumulative probability of toxicity within τ days after enrollment for a patient treated with dose j and schedule k. We will say that the pair (j, k) is *acceptable* if:

$$Pr\{F_{jk}(\boldsymbol{\theta}) > F_{max} \mid \text{data}_{n^*}\} < p_u, \tag{8}$$

where F_{max} is a fixed upper bound on the probability of toxicity by τ specified by the physician and p_u is a fixed decision cut-off, typically set to 0.80 or larger. This is similar to the criterion used for defining acceptable toxicity used by Thall and Cook (2004), in the context of dose-finding based on efficacy and toxicity. We denote the set of acceptable (dose, schedule) combinations by $\mathcal{O}^*$. If $\mathcal{O}^*$ is the empty set, then all (j, k) combinations are unacceptable and the trial is terminated. If $\mathcal{O}^*$ has two or more elements, then we compute the distance measure

$$d_{jk}^* = \mid E\{F_{jk}(\theta) \mid \text{data}_{n^*}\} - \pi^o \mid, \tag{9}$$

for each $(j, k) \in \mathcal{O}^*$, where π^o is a desired target for Pr$(Y < \tau)$ specified by the physician. We assign patient $n^* + 1$ to the element of $\mathcal{O}^*$ having smallest d_{jk}^*.

To protect patient safety, we constrain $\mathcal{O}^*$ to include only (j, k) pairs with doses that are at most one dose above and/or one schedule longer than those combinations already assigned to previous patients. If (j^*, k^*) is the pair that was assigned to the previous subject, and no pair with higher dose or longer schedule has previously been tried, then the next patient may be assigned any pair (j, k) for which $j \leq j^* + 1$ and $k \leq k^* + 1$; we call this the "do not skip" rule. Thus, one may de-escalate, stay at (j^*, k^*), increase either dose or schedule by one level, or increase both dose and schedule by one level, as shown by Figure 1. These restrictions only apply to untried (j, k) pairs when escalating, and we place no restriction on de-escalation of either dose or schedule. Combining all of the above criteria and rules, our algorithm for trial conduct is as follows:

TRIAL CONDUCT

1. Treat the first patient at the lowest (dose, schedule) pair, $(j, k) = (1, 1)$.

2. For each patient after the first, based on the current posterior of $\boldsymbol{\theta}$, determine the set, $\mathcal{O}^*$, of acceptable (j, k) combinations.

3. If $\mathcal{O}^*$ is empty, then stop the trial and conclude that no (j, k) combination is acceptable.

4. If $\mathcal{O}^*$ is not empty, then assign the next patient to the element of $\mathcal{O}^*$ with smallest d^*_{jk}, i.e., the combination whose posterior mean cumulative probability of toxicity by τ is closest to π^o.

5. If the study is not terminated before N patients have been enrolled and fully evaluated with follow-up to τ, select the pair (j^*, k^*) that minimizes d^*_{jk} as the optimal (dose, schedule) combination.

For example, suppose that the first patient has been assigned to combination $(1, 1)$ and has not experienced toxicity. Due to the "do not skip" rule, there are four possible (dose,schedule) combinations for the next patient: $\mathcal{O}^*$= $\{(1, 1), (1, 2),$

$(2,1),\ (2,2)\}$. If we assign the second patient to combination $(2,1)$, and both enrolled patients have not experienced toxicity when the third patient is enrolled, then there are now six possible (dose,schedule) combinations for the third patient: $\mathcal{O}^*$= $\{(1,1),\ (1,2),\ (2,1),\ (2,2),\ (3,1),\ (3,2)\}$. This process is repeated with all successively enrolled patients. Note that if (j,k) is determined to be unsafe by Equation (8), then all pairs (j',k') with $j' \geq j$ and $k' \geq k$ must be unsafe.

5. Application

5.1 *Design and Priors*

The Vidaza® trial has three doses, $d_1 = 8$, $d_2 = 16$, and $d_3 = 24$ mg/m^2 and four schedules with $m^{(1)} = 5$, $m^{(2)} = 10$, $m^{(3)} = 15$, or $m^{(4)} = 20$ administrations. The trial will enroll a maximum of $N = 60$ patients, with each patient followed for up to $\tau = 116$ days. Using the trial conduct algorithm given above, a (dose, schedule) combination is assigned to each new patient at the time of his/her enrollment. The goal is to find a (dose,schedule) combination with mean probability of toxicity by τ closest to $\pi^o = 0.30$. The investigators believed that the single-administration hazards for the three doses have expected peaks at $\xi_{1,2}^{(e)} = 18, \xi_{2,2}^{(e)} = 14$, and $\xi_{3,2}^{(e)} = 10$ days, respectively, with expected remaining durations of $\xi_{1,3}^{(e)} = 10, \xi_{2,3}^{(e)} = 14$, and $\xi_{3,3}^{(e)} = 18$ days. They also believed that the expected probabilities of toxicity by 116 days for the three doses under the shortest schedule are $0.20, 0.25$, and 0.30, so that $\xi_{1,1}^{(e)} = -\log(1-0.20)$, $\xi_{2,1}^{(e)} = -\log(1-0.25)$, and $\xi_{3,1}^{(e)} = -\log(1-0.30)$. We derived $\widetilde{\boldsymbol{\theta}}$ using this elicited information as described in Section 3.3. The prior is characterized by $\sigma_a^2 = \sigma_b^2 = \sigma_c^2 = 1.1$, $\boldsymbol{\mu}_1 = (-3.66, 2.34, 1.75)$, $\boldsymbol{\mu}_2 = (-4.90, 2.09, 2.09)$, $\boldsymbol{\mu}_3 = (-4.83, 1.75, 2.34)$. The safety criterion parameters for applying Equation (4) were defined to be $F_{max} = 0.30$ and $p_u = 0.80$, so that a pair (j,k) is deemed acceptable if less than 80% of the posterior mass of $F_{jk}(\boldsymbol{\theta})$ is above 0.30.

5.2 *Simulation Design*

For the simulation study, we specified seven scenarios in terms of fixed true probabilities of toxicity by day $\tau = 116$, F_{jk}^{true}, for each (j, k). These scenarios are summarized in Table 1 and illustrated by Figure 2. In Scenario 1, all 12 combinations are safe, with $F_{jk}^{true} \leq 0.30$, and the pairs $(2, 4)$ and $(3, 4)$ have F_{jk}^{true} closest to 0.30. Scenarios 2 and 3 have several combinations with $F_{jk}^{true} > 0.30$, and all 12 combinations are toxic in scenario 4. In each of scenarios 5-7, combination $(2, 2)$ has $F_{22}^{true} = 0.30$, but the scenarios differ in terms of how the F_{jk}^{true} values of the other combinations vary around combination $(2, 2)$.

[Table 1 about here.]

[Figure 2 about here.]

Based on input from the principal investigator, patient inter-arrival times were simulated from an exponential distribution with mean of 2 weeks, reflecting an accrual rate of about 2 patients per month, and we simulated 10% of all toxicities to be low-grade, which are classified as dose-limiting at their onset if they fail to resolve within two weeks. Thus, there is a two-week delay in the recording of toxicities that began as low-grade toxicities.

We determined the values ν_1, ν_2 and p_u used for the actual trial by first running an extensive series of preliminary simulations using candidate values $1.1 \leq \nu_1 \leq 3$, $1.1 \leq \nu_2 \leq 14$, and $p_u \in \{0.70, 0.75, 0.80, 0.85, 0.90\}$. Initially, we examined multiple combinations of the three parameters under scenario 4, our "worst-case" scenario, and we selected a set of (ν_1, ν_2, p_u) combinations to ensure a high probability ($\geq 90\%$) of stopping early, to ensure a safe design. Once we found a subset of safe (ν_1, ν_2, p_u) combinations, we ran additional simulations using those combinations under each of the other scenarios to identify final values giving a design with good properties under all scenarios. We found that $\nu_1 = 1.5$, $\nu_2 = 1.5$ and $p_u = 0.80$ reliably chooses (j, k) pairs with F_{jk}^{true} close to the targeted $\pi^o = 0.30$

in each of the non-worst case scenarios, while forcing early termination of the study with at least 90% probability under scenario 4.

To examine the design's robustness, we generated the Y_i's from each of several different parameterizations of Weibull, exponential, and lognormal distributions under each scenario. In Tables 2 and 3, we report simulation results for an exponential distribution with scale parameter chosen so that each F_{jk}^{true} equalled the value specified in Table 1. Table 4 summarizes results under scenario 5 when the Y_i's are generated from a Weibull distribution with shape parameter 0.4 and a lognormal distribution with variance equal to that of the exponential. The scale parameters of the Weibull and lognormal distributions were chosen so that F_{jk}^{true} equalled the value specified in Table 1.

5.3 *Simulation Results*

Table 2 gives the selection frequency and the mean number of patients assigned to each (dose, schedule) pair under each of the scenarios using the proposed design, referred to as "MTDS." As a basis for comparison, we also include results for a conventional Phase I dose-finding design using the CRM (O'Quigley et al. (1990)) with the schedule fixed at schedule 4, assuming $Pr(Y_i \leq 116 \mid d_j, \text{schedule } 4) = p_j^{exp(\alpha)}$ for $j = 1, 2, 3$, with $(p_1, p_2, p_3) = (0.10, 0.30, 0.50)$, and α following a normal prior with mean 0 and variance 2. Table 3 displays summary statistics for both the MTDS and CRM designs under all seven scenarios. We consider (dose, schedule) combinations with $.20 \leq F_{jk}^{true} \leq .40$ to be acceptable choices, with boldfaced values in Table 2 corresponding to such combinations.

[Table 2 about here.]

[Table 3 about here.]

[Table 4 about here.]

In scenario 1, the MTDS design identifies one of the four acceptable combinations as optimal 87% of the time, with on average about 38 patients assigned to

one of these four combinations. The CRM assigns all 60 patients to an acceptable combination, because in this case the CRM design fortuitously only examines schedule 4. In scenario 2, one of the five acceptable combinations is identified as optimal by the MTDS design in 81% of the simulations, with an average of nearly 42 patients assigned to one of these five combinations and around 7 patients assigned to combinations with $F_{jk}^{true} > 0.40$. Because the CRM design is limited to doses with schedule 4, it can possibly find only the one acceptable combination (1,4) in scenario 2, which it identifies as the MTD 66% of the time, with 41 patients assigned to that combination. The CRM terminates early with no dose selected 16% of the time, in contrast to 0% for the MTDS design, because the other two doses are extremely toxic under schedule 4. That is, there is a probability 0.16 that a truly safe agent, when appropriately administered, will be abandoned based on a conventional Phase I trial using the CRM. The limitations of fixing schedule and only varying dose are further illustrated by Table 3. In particular, the CRM has probability 0 of selecting an acceptable dose under each of scenarios 3, 6 and 7, where all of the acceptable doses are at schedules below schedule 4.

In scenario 3, the MTDS design identifies one of the four acceptable combinations as optimal 88% of the time, with about 43 patients assigned to one of these four combinations and 13 patients assigned to combinations with $F_{jk}^{true} > 0.40$. Because all three doses are unacceptably toxic under schedule 4 in scenario 3, the CRM is unable to find an acceptable dose simply because it never examines a lower schedule. In scenario 3, the CRM exposes on average nearly 17 patients to toxic combinations with 55% of them experiencing toxicity, and nearly always terminates early with all combinations deemed unacceptable. In scenario 4, where no combination is acceptable, 29 patients on average are enrolled under the MTDS design before the study terminates, in contrast to only 14 patients under the CRM. However, in the scenario where all combinations are overly toxic the MTDS design assigns a majority of patients only to combination (1,1) and its closest neigh-

bors, with the overall incidence of toxicities similar between the MTDS and CRM designs.

Scenarios 5-7 illustrate the MTDS method for different distributions of acceptable (dose,schedule) pairs over the matrix of 12 pairs studied. Tables 2 and 3 show that the MTDS reliably identifies acceptable pairs in all of these scenarios. The conclusions reached from scenarios 1-4 are re-emphasized: because it does not allow schedule to vary, a conventional CRM dose-finding design is likely to assign a majority of patients to sub-optimal doses and is often unable to identify an optimal (dose,schedule) pair. In contrast, the MTDS design assigns more patients to acceptable combinations and often has a much greater likelihood of selecting a combination suitable for further study.

A critical issue illustrated by the simulations is that the MTDS method is superior to the CRM, or any method that searches for an optimal dose but does not allow schedule to vary. The point is simply that, if the optimal (dose,schedule) combination occurs at a schedule different from the fixed schedule assumed by a method that only varies dose, such a method will have probability 0 of finding the optimal combination, as was the case with CRM under scenarios 3, 6 and 7. Similarly, the MTDS method also is superior to the MTS method of Braun et al. (2005), which fixes dose while only varying schedule. If the fixed dose is suboptimal for all schedules, the MTS method will have probability 0 of finding the optimal combination. We ran additional simulations (not shown) under each scenario in Table 2 using an MTS design that examined all four schedules but assigned all subjects to the same dose of 24 mg/m^2 (final column of Table 1). In scenario 3, in which 24 mg/m^2 is toxic with all four schedules, the MTS design identified the slightly toxic combination $(3, 1)$ as optimal in 41% of simulations and treated an average of 29 subjects with that combination, compared to 29% of simulations and 12 subjects with the MTDS design (see Table 2). More strikingly, in scenarios 5 and 7, where the MTS design is restricted to excessively toxic com-

binations, this design terminated the trial 83% and 96% of the time, respectively, with the false negative conclusion that no optimal combination existed and thus the agent should not be studied in further clinical trials.

5.4 *Robustness*

Table 4 gives results for the MTDS design under scenario 5 with toxicity times generated from Weibull, exponential, and lognormal distributions, described earlier. The first four rows replicate the values displayed in Table 2. Table 4 indicates that the performance of the MTDS method varies very little with the time-to-event distribution, in terms of selection of acceptable (dose, schedule) pairs and numbers of patients assigned. Results for other time-to-event distributions and other scenarios, not shown, were very similar to those presented in Table 4.

[Table 5 about here.]

To study the effects of maximum sample size, we simulated the trial using the MTDS method with N = 40, 60 or 80. In scenario 1, the MTDS method identified combination $(3, 4)$ as optimal 32%, 44%, and 44% of the time with N = 40, 60, and 80, respectively. In scenario 4, where no combination is safe, the trial was terminated early 80%, 90%, and 96% of the time. In scenario 5, the four acceptable combinations were identified as optimal in 64%, 71%, and 79% of the time. Thus, although the design performs best with $N = 80$, our selected sample size of $N = 60$ provides very desirable operating characteristics and improves substantively upon $N = 40$.

6. Discussion

We have proposed a new paradigm for phase I clinical trials aiming to identify a best (dose, schedule) combination. The specific model and parameterization used here were selected for tractability and robustness, although other models certainly are possible. We also examined a version of our design that does not allow "diagonal" escalation, that is, increasing both dose and schedule simultaneously, in

order to protect patient safety when escalating. However, this restriction slowed escalation so severely that far too many patients were assigned to sub-optimal combinations and too few were assigned to optimal combinations. One also may impose the safety constraint that escalation from the current combination (j^*, k^*) cannot occur until a cohort of at least M patients have been assigned to (j^*, k^*), analogous to the usual approach in conventional Phase I designs. However, we found that $M = 1$ yielded a safe design with good operating characteristics.

ACKNOWLEDGEMENTS

Thomas Braun's research was partially supported by the University of Michigan's NIH Cancer Center Support Grant 5-P30-CA-46592. Peter Thall's research was partially supported by NCI grant RO1-CA-83932.

Appendix

We used Markov Chain Monte Carlo (MCMC) with Gibbs sampling for integrating posterior quantities over the parameter space. For each integral, we generated a series of random vectors of model parameters distributed proportionally to the posterior integrand (likelihood times prior), with each series initialized by using the mode. At the start of the trial, we initialized the posterior integrand mode to be the same as the prior mode. When a new patient enrolled, the mode was updated by random sampling around the previous mode. We worked in terms of the log of the model parameters in order to generate all random values from normal distributions. We used two levels of sampling around the previous mode to ensure a good approximation. The first level generates 10,000 normally distributed samples using a large variance for each parameter, roughly two orders of magnitude larger than that parameter's prior variance. In the rare case that this procedure failed to find a mode, we increased the variance and the number of samples and repeated. The second level takes 5000 more samples around the best mode approximation

found at the first level, using the same variance. For the Gibbs sampling, the substep of drawing from the conditional distribution uses importance sampling with a symmetric normal proposal distribution. For each parameter θ_i, $i = 1, 2, \ldots 9$, we generated $\tilde{\theta}_i \sim N(\theta_i, s_i)$, in which s_i is approximately the prior standard deviation. Denoting $\tilde{\theta} = (\theta_1, ..., \theta_{i-1}, \tilde{\theta}_i, ..., \theta_9)$, we computed $A = \min\{1, q(\tilde{\theta}|data)/ q(\theta|data)\}$, in which $q(\cdot)$ is the posterior integrand, and accepted $\tilde{\theta}_i$ as the new θ_i with probability A. MCMC convergence was monitored by comparing the Monte Carlo standard error (MCSE) to the standard deviation of the decision variables (i.e., the posterior cumulative probabilities of toxicity.) We began with 4000 samples and gradually reduced this to a minimum of 1000 samples until the MCSE was $\leq 3\%$ of the posterior standard deviation. Using the batch-means method to estimate the MCSE (with batch size 50), we observed that 1000 random samples were enough to keep the error ratio below 3%. We also used these samples to construct the posterior marginal distribution of each model parameter and confirm that each was a proper distribution with a unimodal shape.

References

BABB, J., A. ROGATKO, and S. ZACKS (1998). Cancer phase I clinical trials: efficient dose escalation with overdose control. *Statistics in Medicine* **17**, 1103–1120.

BRAUN, T. M., Z. YUAN, and P. F. THALL (2005). Determining a maximum tolerated schedule of a cytotoxic agent. *Biometrics* **61**, 335–343.

DAS, P. M. and R. SINGAL (2004). DNA methylation and cancer. *Journal of Clinical Oncology* **22**, 4632–4642.

GELMAN, A., J. B. CARLIN, H. S. STERN, and D. B. RUBIN (2004). *Bayesian Data Analysis, 2nd edition*. Chapman and Hall/CRC.

GOODMAN, S., M. ZAHURAK, and S. PIANTADOSI (1995). Some pratical improvements in the continual reassessment method for phase I studies. *Statistics in Medicine* **14**, 1149–1161.

O'QUIGLEY, J., M. PEPE, and L. FISHER (1990). Continual reassessment method: a practial design for phase I clinical trials in cancer. *Biometrics* **46**, 33–48.

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

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

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

TSUTAKAWA, R. K. and H. Y. LIN (1986). Bayesian estimation of item response curves. *Psychometrika* **51**, 251–267.

Figure 1. Schematic representation of study.

Figure 2. The fixed toxicity probabilities of each (dose, schedule) pair under each of scenario in the simulation study. The dashed horizontal line represents the target toxicity probability 0.30.

Table 1

The fixed toxicity probabilities of each (dose, schedule) pair under each of scenario in the simulation study.

		Dose (mg/m^2)		
Scenario	Schedule	8	16	24
1	4	0.22	0.26	0.30
	3	0.16	0.18	0.23
	2	0.09	0.12	0.18
	1	0.05	0.07	0.11
2	4	0.31	0.45	0.62
	3	0.18	0.32	0.54
	2	0.09	0.21	0.40
	1	0.03	0.14	0.28
3	4	0.55	0.62	0.72
	3	0.45	0.50	0.62
	2	0.30	0.32	0.50
	1	0.10	0.26	0.35
4	4	0.57	0.73	0.78
	3	0.55	0.65	0.75
	2	0.53	0.60	0.65
	1	0.50	0.54	0.58
5	4	0.30	0.48	0.70
	3	0.14	0.32	0.55
	2	0.12	0.30	0.48
	1	0.10	0.28	0.45
6	4	0.50	0.60	0.75
	3	0.30	0.50	0.60
	2	0.12	0.30	0.50
	1	0.03	0.15	0.30
7	4	0.10	0.60	0.70
	3	0.05	0.50	0.60
	2	0.03	0.30	0.55
	1	0.01	0.10	0.50

Table 2

Simulation results for the MTDS method and the CRM used within schedule 4. For each (dose,schedule) pair, column (a) gives the selection percentage and column (b) gives the mean number of patients assigned to the pair. Boldface values correspond to pairs with toxicity probability between 0.20 *and* 0.40.

			Dose (mg/m^2)					
			8		16		24	
Scenario	Method	Schedule	(a)	(b)	(a)	(b)	(a)	(b)
1	MTDS	4	**0.02**	**3.5**	**0.16**	**10.1**	**0.44**	**13.1**
		3	0.01	2.6	0.06	6.4	**0.25**	**11.4**
		2	0.00	1.6	0.00	3.5	0.05	6.5
		1	0.00	1.0	0.00	0.1	0.00	0.3
	CRM	4	**0.47**	**30.6**	**0.31**	**17.8**	**0.19**	**10.3**
2	MTDS	4	**0.17**	**8.6**	0.08	5.7	0.00	0.2
		3	0.10	6.7	**0.29**	**11.7**	0.01	1.3
		2	0.00	2.8	**0.19**	**10.6**	**0.13**	**8.7**
		1	0.00	1.1	0.00	0.5	**0.03**	**2.2**
	CRM	4	**0.66**	**41.2**	0.18	12.1	0.00	0.4
3	MTDS	4	0.01	2.9	0.00	0.5	0.00	0.0
		3	0.06	6.1	0.01	1.6	0.00	0.0
		2	**0.26**	**13.7**	**0.14**	**9.9**	0.01	1.7
		1	0.03	3.3	**0.19**	**7.3**	**0.29**	**12.4**
	CRM	4	0.03	15.6	0.00	1.1	0.00	0.0
4	MTDS	4	0.00	0.8	0.00	0.0	0.00	0.0
		3	0.00	1.7	0.00	0.2	0.00	0.0
		2	0.00	3.7	0.00	2.0	0.00	0.2
		1	0.08	11.1	0.02	5.1	0.00	3.7
	CRM	4	0.01	13.8	0.00	0.5	0.00	0.0

Table 2 (continued)

			Dose (mg/m^2)					
			8		16		24	
Scenario	Method	Schedule	(a)	(b)	(a)	(b)	(a)	(b)
5	MTDS	4	**0.19**	**8.8**	0.04	3.7	0.00	0.1
		3	0.09	6.5	**0.19**	**9.7**	0.00	0.7
		2	0.01	3.4	**0.24**	**12.3**	0.06	5.8
		1	0.00	1.4	**0.09**	**3.1**	0.08	4.6
	CRM	4	**0.72**	**43.2**	0.15	11.0	0.00	0.2
6	MTDS	4	0.09	6.9	0.00	0.9	0.00	0.0
		3	**0.32**	**12.9**	0.03	3.2	0.00	0.2
		2	0.13	9.3	**0.29**	**15.0**	0.01	3.0
		1	0.00	1.2	0.00	1.4	**0.11**	**6.1**
	CRM	4	0.08	20.6	0.00	1.7	0.00	0.0
7	MTDS	4	0.03	6.1	0.01	2.2	0.00	0.0
		3	0.00	2.7	0.09	8.3	0.00	0.4
		2	0.00	1.5	**0.54**	**20.9**	0.01	5.3
		1	0.00	1.0	0.13	3.5	0.19	8.2
	CRM	4	0.77	40.1	0.23	19.6	0.00	0.2

Table 3

Summary statistics for the MTDS and CRM designs under each scenario.

		Scenario						
	Method	1	2	3	4	5	6	7
Probability of Selecting an Acceptable Dose	MTDS	0.87	0.81	0.88	n/a	0.71	0.72	0.54
	CRM	0.97	0.66	0.00	n/a	0.72	0.00	0.00
Probability of Selecting No Dose	MTDS	0.00	0.00	0.01	0.90	0.00	0.00	0.00
	CRM	0.03	0.16	0.97	0.99	0.13	0.92	0.00
Mean Number of Patients Enrolled	MTDS	60.0	60.0	59.5	28.7	60.0	60.0	60.0
	CRM	58.7	53.6	16.7	14.3	54.5	22.4	59.9
Observed Incidence of Toxicity	MTDS	0.22	0.29	0.34	0.54	0.31	0.31	0.33
	CRM	0.25	0.34	0.55	0.58	0.34	0.51	0.27

Table 4

Simulation results for scenario 5 under different time to toxicity distributions. For each (dose,schedule) pair, column (a) gives the selection percentage, and column (b) gives the mean number of patients assigned to the pair. Boldface values correspond to pairs with cumulative toxicity probability between 0.20 *and* 0.40.

Time-to-Event Distribution	Schedule	Dose (mg/m^2) 8 (a)	8 (b)	16 (a)	16 (b)	24 (a)	24 (b)
Exponential	4	**0.19**	**8.8**	0.04	3.7	0.00	0.1
	3	0.09	6.5	**0.19**	**9.7**	0.00	0.7
	2	0.01	3.4	**0.24**	**12.3**	0.06	5.8
	1	0.00	1.4	**0.09**	**3.1**	0.08	4.6
Weibull	4	**0.25**	**9.4**	0.03	2.0	0.00	0.0
	3	0.11	8.4	**0.16**	**7.5**	0.00	0.2
	2	0.03	4.4	**0.20**	**11.7**	0.03	4.0
	1	0.00	2.2	**0.12**	**4.9**	0.08	4.9
Lognormal	4	**0.18**	**9.0**	0.07	5.5	0.00	0.1
	3	0.07	5.7	**0.17**	**9.3**	0.00	1.2
	2	0.01	2.9	**0.27**	**12.2**	0.04	6.6
	1	0.00	1.2	**0.07**	**1.9**	0.10	4.4