STATISTICS IN MEDICINE
Statist. Med. 2001; **20**:867–882

Heterogeneity in phase I clinical trials: prior elicitation and computation using the continual reassessment method

Anna T. R. Legedza[1,*,†] and Joseph G. Ibrahim[1,2]

[1]*Department of Biostatistics, Harvard School of Public Health, Boston, U.S.A.*
[2]*Dana Farber Cancer Institute*

SUMMARY

Heterogeneity in a phase I clinical trial patient population may lead to distinctly different dose–response relationships along covariate values. For a given target probability of toxicity, this implies different maximum tolerated doses (MTDs) for each distinct subpopulation. Within the framework of O'Quigley, Pepe and Fisher's (1990) continual reassessment method, we propose the notion of average and patient-specific MTDs by augmenting the dose–response model with other covariates to account for such differences. A method to elicit prior distributions on the dose and other covariate parameters are proposed, based on the predictive approach of Ibrahim and Laud (1994), Laud and Ibrahim (1995), and Ibrahim, Ryan and Chen (1998). This approach relies on prior predictions for the response vector y_0 and a quantity a_0 specifying uncertainty in y_0. Then, y_0 and a_0 are used to specify a prior for the regression coefficients in a semi-automatic fashion. The elicitation scheme for y_0 uses results from previous phase I cancer clinical trials. The average and patient-specific MTDs and an elicitation method are demonstrated in logistic regression examples. Copyright © 2001 John Wiley & Sons, Ltd.

1. INTRODUCTION

The purpose of phase I clinical trials is to estimate the maximum tolerated dose (MTD), or the highest dose at which a prespecified level of risk of toxicity is not exceeded. Storer [1] notes that phase I studies are complicated by such challenges as: '(i) the relatively long time it takes to evaluate each subject; (ii) the ethical requirement to approach the MTD conservatively; (iii) the relative subjectiveness of the response (toxic/non-toxic instead of alive/dead); (iv) the heterogeneity of the subject population, and (v) the possible difficulty in classifying response due to early drop-out for reasons unrelated to toxicity'. One can add to this list the difficulty in incorporating prior information into the human phase I clinical trial.

In order to estimate the MTD in an ethical manner, most phase I clinical trials are conducted along similar lines. The first patient (or group of patients) is administered a starting dose

* Correspondence to: Anna T. R. Legedza, Department of Biostatistics, Harvard School of Public Health, 655 Huntington Avenue, Boston, Massachusetts 02115, U.S.A.
† E-mail: anna@hsph.harvard.edu

Contract/grant sponsor: NIH; contract/grant number: CA 70101-01, CA 74015-01
Contract/grant sponsor: NIH-NCI Biostatistics/Epidemiology Training Grant; contract/grant number: 5 T32 CA09337

Received January 1999
Accepted March 2000

thought to be lower than the true MTD. A higher dose is administered to the next patient (or group). This process is repeated until a toxic response occurs. Then, the subsequent patient (or group) is administered the previous dose or a lower one. Usually 20 or so patients are recruited for a phase I trial. Once they have all responded, the MTD is estimated to be the last or penultimate dose level administered. Many variations of this method have been proposed. They all share the characteristic of experimenting at various doses in a sequential manner until too many toxic responses are observed. Their differences lie in the types of methods used to improve dose escalation, administration, and estimation. A popular Bayesian approach is the continual reassessment method (CRM) [2].

In this paper, we introduce a new methodology for phase I analysis based on CRM. Using historical data obtained from previous studies, we propose a class of prior distributions on the parameters of a new covariate-augmented dose–response model. Then, using new definitions of maximum tolerated dose, we estimate both an average MTD as well as patient-specific MTDs using CRM. These estimates can be used to determine more accurate dosages at initial and intermediate stages within phase I clinical trials. They can also motivate the design of later phase trials and, ultimately, dosage recommendations for the approved drug.

It is worthwhile to review the rudiments of CRM in the standard case to see how our innovations were motivated. At each iteration, CRM estimates the MTD, given the responses of all previous patients and prior information on the dose–response parameter. In the degenerate case of estimating the starting dose, this approach relies solely on the prior distribution. If prior information on the dose–response parameter were vague, there could be the possibility that the starting dose would be far away from the true MTD, leading experimenters to administer improper doses to the first few patients. On the other hand, if the prior were informative, it is more likely that more patients would take doses closer to the true MTD. Therefore, it is imperative to elicit as much prior information as possible on the dose–response parameter.

We utilize a class of informative prior distributions (proposed by Ibrahim *et al.* [3] and Chen *et al.* [4] in different contexts) that are appropriate for phase I clinical trials. The proposed priors are constructed from data obtained in previous studies.

The heterogeneity of certain patient populations motivates the need to elicit other types of prior information as well. Many phase I clinical trials are conducted on a homogeneous group, for example, 18–25-year-old healthy males. In this situation, the estimated MTD from a phase I trial would be valid for dosing a similar population, but would it be so for sick 18–25-year-old males? Or for older women? Another situation where this problem arises is in cancer patients. Patients in phase I cancer clinical trials are a heterogeneous lot, differing with respect to age, sex, race, disease status etc. Depending on the nature of the drug, a single MTD might not be appropriate for all these patients. For example, the recommended dosage at a given iteration of CRM might not seem appropriate for the next patient, an older female, given that all the previous patients (on whose responses the recommended dosage is based) are young males. Instead, it would be desirable to recommend a dose that takes into account her different personal characteristics. What needs to be done, then, is to incorporate patient covariate values into the dose–response model.

All these proposals will be developed in the following sections of this paper. In Section 2 we review the continual reassessment method. We introduce the standard and augmented dose–response models in Section 3 and a new class of prior distributions in Section 4. To estimate average and patient-specific MTDs, CRM is applied to the new priors and likelihood in Section 5. A novel method to elicit previous data is presented in Section 6. In Section 7,

a computational implementation of the proposals is described. We present numerical examples demonstrating our proposed methodology in Section 8, followed by a discussion in Section 9.

2. CRM FOR THE STANDARD DOSE–RESPONSE MODEL

At any particular point in time in a phase I trial, it is imperative to choose a dose that all the available evidence points to as the most appropriate estimate of the MTD x^*. In O'Quigley, *et al.*'s [2] continual reassessment method, the authors chose to work with some simple one-parameter models for the dose–response curve,

$$P(Y=1|x)=E(Y|x)=\psi(x,b)$$

where Y denotes the type of response ($1=$toxic, $0=$non-toxic), x is the dose, and b is the dose–response parameter. Suppose that there are k distinct dose levels $(x_1,\ldots,x_k)$ under investigation during the course of the trial. The goal of CRM is to find the unique value, b^*, of the parameter b for a given probability of toxicity ω such that the MTD, x^*, solves $\psi(x^*,b^*)=\omega$. Let $p(b|y_1,\ldots,y_{j-1})$ be the posterior distribution on the dose parameter b, given the current data, the observed toxic responses $(y_1,\ldots,y_{j-1})$ of the $(j-1)$ patients. In other words, this function describes all that is known about b immediately before experimenting upon the jth patient. The current estimate of the probability of toxic response at dose x_i after observing patient $(j-1)$'s response is

$$\omega_{ij}=\int \psi(x_i,b)\,p(b|y_1,\ldots,y_{j-1})\mathrm{d}b$$

Alternately, one could estimate this probability by first finding the posterior expectation of b

$$\mu(j)=\int b\,p(b|y_1,\ldots,y_{j-1})\,\mathrm{d}b$$

and then substituting

$$\omega'_{ij}=\psi(x_i,\mu(j))$$

One selects the next experimental dose level x_i from among $(x_1,\ldots,x_k)$ so that it is close to x^*, where 'closeness' is measured in terms of Euclidean distance, $\Delta(u,w)=(u-w)^2$. Thus, armed with the estimate ω_{ij} (or ω'_{ij}), the next dose x_i (from among the eligible doses $(x_1,\ldots,x_k)$) minimizes one of the three distance criteria (i) $\Delta(\omega_{ij},\omega)$, (ii) $\Delta(\omega'_{ij},\omega)$, or (iii) $\Delta(x_i,\psi^{-1}_{b=\mu(j)}(\omega))$. Having administered the chosen dose and observed the jth patient's response y_j, Bayes theorem updates the knowledge about the parameter b via the formula

$$p(b|y_1,\ldots,y_{j-1},y_j)=\frac{L(b|y_j)\,p(b|y_1,\ldots,y_{j-1})}{\int L(b|y_j)\,p(b|y_1,\ldots,y_{j-1})\,\mathrm{d}b}$$

where the likelihood of b for an observed response y_j at a dose x_i is

$$L(b|y_j)=\psi(x_i,b)^{y_j}(1-\psi(x_i,b)^{y_j})^{1-y_j}$$

Now, one can iterate anew to obtain the $(j+1)$th dose. Once all enrolled patients have responded, the MTD x^* will be the dose chosen according to one of the three criteria based on those n responses.

As described up to this point, the continual reassessment method estimates a single MTD for an assumed homogeneous group of recruited subjects. In this situation, it is not so important to obtain good probability estimates over the entire range of possible doses. Rather, it is more important to estimate only the MTD well. That is why O'Quigley *et al.* [2] chose to use a simple one parameter model $\psi(x,b)$ over a more complex one. In this paper, however, we will describe why it is desirable to account for heterogeneity among the phase I trial subjects and in the drug's intended target population. This will necessitate using more complex models because we will want to estimate a set and/or a continuum of MTDs. In order to compare the performance of such models with the original CRM, we need to add another parameter, β_0, to $\psi(.)$. Let this new two-parameter model

$$P(Y=1|x) = E(Y|x) = \psi(x,\beta_0,b)$$

be referred to as the *standard dose–response model*. It will serve to mimic the performance of the 'traditional' CRM based on $\psi(x,b)$. This supposition is tenable, based upon similar dose recommendations made from one-parameter and two-parameter model simulations in O'Quigley *et al.* [2]. In other words, since both are similar, we will use the MTD estimated from the two-parameter model in place of the MTD estimate from the one-parameter model. In reference to this MTD, we will use the terms 'standard', 'traditional', and 'two parameter' interchangeably in this paper.

To estimate the traditional MTD using the standard dose–response model, we need to amend notation. Given all n responses, let the joint posterior of the two parameters be $p(\beta_0,b|y_1,\ldots,y_n)$. Let

$$\mu_{\beta_0}(n) = \iint \beta_0\, p(\beta_0,b|y_1,\ldots,y_n)\mathrm{d}b\,\mathrm{d}\beta_0$$

$$\mu_b(n) = \iint b\, p(\beta_0,b|y_1,\ldots,y_n)\mathrm{d}\beta_0\,\mathrm{d}b$$

Suppose we use the third distance criterion to estimate the MTD from a continuum of doses, $x \in [-\infty,+\infty]$. Theoretically, the continuous set can be viewed as the limit of a very large set of discrete doses on $[-\infty,+\infty]$. Then, all we need to do is invert $\psi(.)$ such that MTD x^* solves

$$\omega = \psi(x,\beta_0 = \mu_{\beta_0}(n), b = \mu_b(n))$$

In this paper, a continuous set of candidate doses will be used to simplify computation.

3. LIKELIHOODS FOR DOSE–RESPONSE MODELS

In this section we propose a parametric form for the standard dose–response model that makes it amenable to expansion to non-standard situations. In phase I cancer clinical trials, for example, patients differ in terms of sex, age, disease status etc. We would like a model to reflect this heterogeneity and its effect on patients' responses to dosing, so we choose a parametric form that can then be expanded to incorporate covariates in what we term the *augmented dose–response model*.

3.1. Standard dose–response model

Suppose the standard dose–response model can be written as

$$P(Y = 1|x) = E(Y|x) = \psi(x, \beta_0, b) = \Psi(\eta)$$

where $\eta = \beta_0 + bx$, Y is the binary random variable of toxic response, Ψ is the inverse link function in a generalized linear model, β_0 is the intercept, b is the regression coefficient, and x is dose on the logarithmic scale. From now on, we will refer to x simply as 'dose'. Let D_n denote the data for all the patients that have completed the current phase I trial so far. Let $D_n \equiv (n, y, X)$, where n is the sample size, $y' = (y_1, y_2, \ldots, y_n)$ is the $n \times 1$ vector of the n patients' observed binary toxicity responses, and X is the $n \times 2$ matrix of their intercept and dose covariate values. Let $\eta_i = \beta_0 + bx_i$. We can write the likelihood of β_0 and b for these n patients as

$$L(\beta_0, b|D_n) = \prod_{i=1}^{n} \Psi(\eta_i)^{y_i}\{1 - \Psi(\eta_i)\}^{1-y_i}$$

3.2. Augmented dose–response model

As mentioned in the introduction, it is desirable to better estimate dosages for different patient subpopulations. To do so, we need to augment the standard dose–response model with other covariates that distinguish the subpopulations besides dose by letting

$$P(Y = 1|x, z) = E(Y|x, z) = \Psi(\eta_z)$$

where $\eta_z = \beta_0 + bx + \beta_1' z$, z is a $p \times 1$ vector of covariates, and β_1 is a $p \times 1$ vector of regression coefficients corresponding to z. Let $\eta_{zi} = \beta_0 + bx_i + \beta_1' z_i$. In this instance, we can write the likelihood of $(\beta_0,\ b,\ \beta_1)$ as

$$L(\beta_0, b, \beta_1|D_n) = \prod_{i=1}^{n} \Psi(\eta_{zi})^{y_i}\{1 - \Psi(\eta_{zi})\}^{1-y_i}$$

where $D_n \equiv (n, y, X)$. That is, X is now the $n \times (p + 2)$ matrix of patients' intercept, dose and additional covariates.

4. PRIOR DISTRIBUTIONS

Informative prior distributions have not been developed for the CRM/phase I setting before. Here we propose a class of priors similar to those of Ibrahim *et al.* [3], which in turn are based upon proposals in Ibrahim and Laud [5] and Laud and Ibrahim [6]. The priors in Ibrahim *et al.* were used in the context of score tests for carcinogenicity experiments, which is completely different from our aim here. In this paper, the priors are motivated as follows. Suppose we are given data from a previous study similar to the current phase I study. These previous data can be used to construct a predictive prior distribution on the dose and other covariate parameters in the current study's model. In following subsections, we describe novel prior distributions for the standard and augmented dose–response models.

4.1. A proposed prior for the standard dose–response model

Our prior distributions are based on the notion of the availability of prior data. It often occurs in many phase I cancer clinical trials that there exist historical data or data from similar phase I studies in which the drug was used for a different indication. Let us denote these previous data as $D_0 \equiv (n_0, y_0, X_0)$, where n_0 is the sample size of a previous study, y_0 is an $n_0 \times 1$ vector of raw responses from that study, and X_0 is an $n_0 \times 2$ matrix of intercept and dose covariates corresponding to y_0. We propose a prior distribution for (β_0, b) of the form

$$\pi(\beta_0, b|D_0, a_0) \propto \left[\prod_{i=1}^{n_0} \Psi(\eta_i)^{y_{0i}} \{1 - \Psi(\eta_i)\}^{1-y_{0i}} \right]^{a_0}$$

where $a_0 \in [0, 1]$ is a scalar prior parameter that controls how much influence the prior has on the current study. The parameter a_0 can be interpreted as a precision parameter. There is no closed form for this distribution. However, Ibrahim *et al.* [3] and Chen *et al.* [7] have shown that this joint prior distribution is proper under some very general conditions. Once the prior is specified and the jth patient has responded, the posterior distribution of the parameters is

$$p(\beta_0, b|D_0, a_0, y_1, \ldots, y_j) \propto \pi(\beta_0, b|D_0, a_0) L(\beta_0, b|D_n)$$

At this point, CRM can be applied to $p(\beta_0, b|D_0, y_1, \ldots, y_j)$, as described in Section 2.

4.2. A proposed prior for the augmented dose–response model

It often occurs in phase I studies that there exist historical data containing other covariate information. This covariate information is crucial to the development of appropriate dosing recommendations. Later in this paper, we will present two new concepts of maximum tolerated dose based on the augmented dose–response model. First, though, we need to develop a suitable prior.

For the augmented dose–response model, we again utilize the predictive prior framework for prior construction. Let the data from previous studies be denoted as $D_0 \equiv (n_0, y_0, X_0)$, where n_0 is the sample size of a previous study, y_0 is an $n_0 \times 1$ vector of raw responses from that study, and $X_0 = [1, x_0, z_0]$ is an $n_0 \times (p+2)$ matrix of intercept, dose, and other covariates corresponding to y_0. The joint prior for (β_0, b, β_1) is

$$\pi(\beta_0, b, \beta_1|D_0, a_0) \propto \left[\prod_{i=1}^{n_0} \Psi(\eta_{zi})^{y_{0i}} \{1 - \Psi(\eta_{zi})\}^{1-y_{0i}} \right]^{a_0}$$

In the next section, this generalization of the dose–response model and associated priors will be incorporated into the framework of CRM and lead to new definitions of MTD.

5. CRM FOR THE AUGMENTED DOSE–RESPONSE MODEL

In some methods, pains are taken to make conservative dose estimates at intermediate points of a phase I trial in order to avoid potential toxicity. For example, Faries [8] mentions that practitioners fear that priors may be too powerful and cause dose escalation to proceed too quickly. Although it is difficult to believe that the use of an informative prior in clinical

settings would result in anything but appropriate dose escalation, there is room for improvement. Up to now, at any particular point in a phase I trial, the next patient's dose is selected by determining what would be best for the group as a whole, not the patient in particular. Keeping a patient's best interests in mind, then, we propose a new definition of MTD – the *patient-specific MTD* – in order to provide clinicians with more information before they determine the appropriate dosage for an individual at intermediate points of a phase I trial. In order to determine the best dosage for the target population as a whole, though, we also propose the *average MTD* as a means of guiding phases II and III. Both these new types of MTD will be developed within CRM using the augmented dose–response model. The general approach to estimating these two new types of MTD is to replace the posterior distribution of (β_0, b) in Section 2 with two different ones, respectively.

5.1. Patient-specific MTD

During a phase I trial, each patient should be administered the dose level most appropriate to him, based on his distinguishing characteristics and past experience. This requirement motivated us to define an alternative dose, the patient-specific MTD x_{P}, which will explicitly take into account a patient's history via the augmented dose–response model. Furthermore, it will look at the behaviour of (β_0, b) conditional on covariate values.

The patient-specific MTD x_{P} is estimated using the joint posterior distribution of all the parameters. That is, we condition on the specific values of the covariates that interest us by plugging them into the joint posterior distribution. To establish notation, let $\theta = (\beta_0, b, \beta_1', a_0)$. Then, the posterior distribution of all the parameters is

$$p(\theta|D_j) = \frac{L(\theta|D_j)\pi(\theta|D_0)}{\int L(\theta|D_j)\,\mathrm{d}\theta}$$

where D_j denotes the data of the first j patients that have completed the phase I trial so far. That is

$$L(\theta|D_j) = \prod_{i=1}^{j} L(\theta|y_i)$$

The posterior joint distribution of the parameters of interest is

$$p(\beta_0, b, \beta_1'|D_j) = \int p(\theta|D_j)\,\mathrm{d}a_0$$

Let the posterior joint of expectation of (β_0, b, β_1') with respect to $p(\beta_0, b, \beta_1'|D_j)$ be denoted as $(\mu^{\mathrm{P}}_{\beta_0}(j), \mu^{\mathrm{P}}_{b}(j), \mu^{\mathrm{P}}_{\beta_1'}(j))$. Then, given the covariate values z for the next patient, the patient-specific MTD x_{P} is defined as the dose which solves the following expression for x:

$$\omega = \psi(x, z, \beta_0 = \mu^{\mathrm{P}}_{\beta_0}(j), \quad b = \mu^{\mathrm{P}}_{b}(j), \quad \beta_1 = \mu^{\mathrm{P}}_{\beta_1}(j))$$

This is an extension of the definition of the traditional MTD x^* (Section 2) to the case of a patient with covariate value z. The patient-specific MTD x_{P} incorporates more information and is therefore more appropriate for patient treatment.

5.2. Average MTD

The goal of a phase I trial is to estimate the maximum dose (MTD) at which a diseased population will experience few toxic side-effects, assuming the population's average toxic response. That is, the target dose will work despite other distinguishing characteristics besides the pertinent disease. This traditional MTD is a useful concept epidemiologically in that clinicians, drug companies and regulatory authorities use it to make decisions on whether to terminate or continue the clinical trial process based on their assessment of how detrimental/beneficial the recommended dose may be to the target population.

In past work, both frequentist and Bayesian, this traditional MTD was an automatic result of the marginal modelling process. However, we feel there is more to be gained by explicitly modelling potential dose–response-altering covariates using the augmented dose–response model and integrating to arrive at the usual marginal model, resulting in what we define as the average MTD x_{A}.

To estimate the average MTD x_{A}, we first need to compute the marginal posterior distribution of the parameter (β_0, b) once all patients have completed the trial. This distribution is

$$p_{\mathrm{A}}(\beta_0, b|D) = \frac{\int\int L(\beta_0, b, \beta_1|D)\pi(\beta_0, b, \beta_1|D_0, a_0)\,\mathrm{d}\beta_1}{\int\int\int L(\beta_0, b, \beta_1|D)\pi(\beta_0, b, \beta_1|D_0, a_0)\,\mathrm{d}b\,\mathrm{d}\beta_1}$$

where D refers to all data accumulated over the entire phase I trial. Let the posterior joint of expectation of (β_0, b) with respect to $p_{\mathrm{A}}(\beta_0, b|D)$ be denoted as $(\mu^{\mathrm{A}}_{\beta_0}(j), \mu^{\mathrm{A}}_{b}(j))$. The average MTD is then defined as the value x_A which solves the following expression for x:

$$\omega = \psi(x, \beta_0 = \mu^{\mathrm{A}}_{\beta_0}(j), b = \mu^{\mathrm{A}}_{b}(j))$$

6. PRIOR ELICITATION OF D_0

In this section we propose a method to elicit data from previous studies in order to construct the predictive priors developed in Section 4. The elicitation procedure consists of transforming raw data into a usable format, denoted as D_0.

In phase I trials, previous human experiments on the experimental drug are rare, save when a drug is being tested for a new indication. In either case, elicitation of D_0 is straightforward – just plug in previous study results. However, attention should be given to the tuning parameter $a_0 \in [0, 1]$. If one feels that the previous study's results are not very applicable to the current situation, then one would choose a_0 closer to 0. On the other hand, if previous data are directly pertinent, then one would choose a_0 closer to 1.

7. COMPUTATIONAL IMPLEMENTATION

The basic algorithm to compute the most appropriate MTD at each iteration of a phase I trial was outlined in Section 2. The modifications to compute our proposed average and patient-specific MTDs were introduced in Section 5. The only novelty these modifications present are the use of two alternative posterior distributions (the marginal and the joint). In this section

we show how to compute these posterior distributions. Then we show how to compute the two new MTDs using CRM.

7.1. Computing the posterior distribution of parameters

No closed forms exist for any of the posterior distributions required for the computation of the average and patient-specific MTDs within CRM. Therefore, we intend to use Markov chain Monte Carlo (MCMC) methods to approximate the posterior distributions of the parameters. In fact, we will demonstrate how using one Gibbs sampling run will give us both the necessary marginal and joint distributions.

The joint posterior distribution of the parameters does not have a closed form and is only known up to a proportionality constant:

$$p(\beta_0, b, \beta_1 | D_0, a_0, D_j) \propto$$

$$\left[\prod_{i=1}^{n_0} \Psi(\beta_0 + bx_i + \beta_1' z_i)^{y_{0i}} \{1 - \Psi(\beta_0 + bx_i + \beta_1' z_i)\}^{1-y_{0i}} \right]^{a_0}$$

$$\times \prod_{i=1}^{n} \Psi(\beta_0 + bx_i + \beta_1' z_i)^{y_i} \{1 - \Psi(\beta_0 + bx_i + \beta_1' z_i)\}^{1-y_i}$$

Since we do not know its closed form, we intend to approximate it using Gibbs sampling. For this, we need to specify the full set of conditionals, each of which is only known up to a proportionality constant. It is easy to see that β_0, b, and elements of β_1 each have the same format, except with respect to different random variables. For example, when β_0 is the random variable

$$p(\beta_0 | b, \beta_1, D_0, a_0, D_j) \propto$$

$$\left[\prod_{i=1}^{n_0} \Psi(\beta_0 + bx_i + \beta_1' z_i)^{y_{0i}} \{1 - \Psi(\beta_0 + bx_i + \beta_1' z_i)\}^{1-y_{0i}} \right]^{a_0}$$

$$\times \prod_{i=1}^{n} \Psi(\beta_0 + bx_i + \beta_1' z_i)^{y_i} \{1 - \Psi(\beta_0 + bx_i + \beta_1' z_i)\}^{1-y_i}$$

For b conditional on all the other parameters, we would use the same expression, except using b as the random variable. The same approach goes for the elements of β_1.

Once these conditionals are specified, Gibbs sampling proceeds as follows:

1. Specify starting values for β_0 and β_1.
2. Draw a random b, conditional on the other parameters' values and the data.
3. Draw a random β_0, conditional on the other parameters' values and the data.
4. Draw random elements of β_1 in turn, conditional on the other parameters' values and the data.
5. Repeat steps 2 to 4 until convergence reached.

The only remaining task is to draw the univariate random samples in steps 2 to 5. Since their conditional distributions are known only up to a proportionality constant, one can use Gilks *et al.*'s [9] adaptive rejection Metropolis sampling (ARMS) method to draw from such distributions. ARMS only requires that the user specify the logarithm of the target distribution up

to a constant. ARMS then uses secant/tangent approximations to the curve to perform piecewise exponential sampling. This algorithm is available as C code from its author, Wally Gilks [10]. We used a C program to call the ARMS code. Since all the conditional distributions are log-convex in our logistic regression context, it is only necessary to call a special case of ARMS called ARS, or adaptive rejection sampling, in order to sample from them.

Let $\{\beta_{0k}\}$, $\{b_k\}$, $\{\beta_{1k}\}$, $k = 1, \ldots, m$, denote the vectors of the respective univariate samples, where m is the total number of iterations. In the next two sections we shall demonstrate that, in practice, these Gibbs samples greatly simplify the computation of the integrals by which the patient-specific and average MTDs are defined.

7.2. Computing the patient-specific MTD

As described in Section 4.1, to determine the patient-specific MTD x_{P}, the posterior expectations $(\mu^{\mathrm{P}}_{\beta_0}, \mu^{\mathrm{P}}_{b}, \mu^{\mathrm{P}}_{\beta_1})$ must be computed first. Although their definitions involve multi-dimensional integrals, the nature of Gibbs sampling allows us to estimate them using the sums

$$\hat{\mu}^{\mathrm{P}}_{\beta_0} = \frac{\sum_{k=1}^{m} \beta_{0k}}{m}, \quad \hat{\mu}^{\mathrm{P}}_{b} = \frac{\sum_{k=1}^{m} b_{0k}}{m}, \quad \hat{\mu}^{\mathrm{P}}_{\beta_1} = \frac{\sum_{k=1}^{m} \beta_{1k}}{m}$$

Plugging these estimates into the $\psi(.)$ of Section 4.1 and solving for x numerically using the S-plus [11] function `uniroot()` yields $\hat{x}_{\mathrm{P}}$.

7.3. Computing the average MTD

To determine the average MTD x_{A}, the posterior expectations $(\mu^{\mathrm{A}}_{\beta_0}, \mu^{\mathrm{A}}_{b})$ must be computed first. Since Gibbs samples are both marginal and joint samples, we can estimate the posterior expectations using the sums from Section 6.2:

$$\hat{\mu}^{\mathrm{P}}_{\beta_0} = \hat{\mu}^{\mathrm{A}}_{\beta_0}, \quad \hat{\mu}^{\mathrm{P}}_{b} = \hat{\mu}^{\mathrm{A}}_{b}$$

Plugging these estimates into the $\psi(.)$ of Section 4.2 and solving for x numerically using the S-plus [11] function `uniroot()` yields $\hat{x}_{\mathrm{A}}$.

8. EXAMPLE

To illustrate how the augmented dose–response model improves estimation of the MTD within CRM, we simulated 200 historical data from an augmented logistic dose–response model where

$$P(Y = 1 | \text{dose}, \ \text{sex}) = \frac{\exp(2.1 + 5 \times \text{dose} + 0.7 \times \text{sex})}{1 + \exp(2.1 + 5 \times \text{dose} + 0.7 \times \text{sex})}$$

The true female- and male-specific MTDs are listed in Table I.

Four different models were run to see how estimates compare. The first two models (patient-specific (true) and standard) both include intercepts. We call the last two O'Quigley-type models because they do not have intercepts, but they mimic the first two with respect to all other parameters. For each model, we adjusted the informativeness of the prior by running all combinations of three different prior samples sizes (200, 100, and 50) with three different

Table I. Example – true doses.

Type of dose	Value
True female-specific MTD	−0.729
True male-specific MTD	−0.449

types of tuning parameters (0.9, 0.5 and 0.1). The results of all these runs are listed by model type, a_0, and prior sample size N_0 in Tables II, III, IV and V.

MTD estimates and standard errors were computed using the delta method. This method was chosen over just computing the estimated mean and standard error of the distribution of the MTD because it was more robust to occasional extreme draws. The estimates were similar in the cases where there were not any extreme draws.

Each trial was simulated with 20 patients. In all situations, the Gibbs samplers appeared to converge in a few hundred iterations. The total number of Gibbs sampling iterations varied widely by model type and values of a_0. In the models that included intercepts, the intercept and dose–response parameters Gibbs runs were both highly cross-correlated as well as autocorrelated. To account for this, we used large lags to select sample draws. In neither the true nor O'Quigley type patient-specific models were the patient-specific covariates appreciably correlated with the other parameters. Thus, we were able to use shorter lags in these situations. In all situations, however, the necessary lag size increased with decreasing a_0.

Table II contains the results of the patient-specific (true) model run. In all of the prior sample size and a_0 combinations, the true female- and male-specific MTDs were well-estimated. In addition, the average MTDs in each case were similar and fell in between the patient-specific estimates. The average MTD estimates are very similar to the standard MTD estimates in Table III. This similarity illustrates the usefulness of the patient-specific model in being able to estimate both patient and population dose estimates at once with the same number of patients, thus obviating the need for the standard model.

In both Tables II and III, note that the estimated first dose of each trial is similar to the resulting MTD estimate. However, in the case of the standard model in Table II, the first dose and all other doses were in between the patient-specific MTDs. Although our results show that the number of toxicities sustained in both types of trials were similar, it would seem that using the first model would be preferable in that it yields first doses closer to the desired targets.

Tables IV and V repeat the story for the O'Quigley-type models in which there were no intercepts. In these cases, however, the omission of the intercept led to higher variability in the estimates. For instance, the resulting patient-specific, average, and standard MTD estimates are for the most part close to the true values and the corresponding values in Table III. However, note that the standard errors are at least twice as large in the most informative cases where $a_0=0.9$ and $N_0=200$, while in less informative situations the standard errors increase even more. The prior for the $N_0=50$ is so uninformative as to render estimation numerically impossible. Note that first dose estimates are much more variable in the no-intercept models.

In Tables II to V, the last four columns represent, respectively, the total number of toxic events encountered, the number of times the dose administered was greater than the true MTD, the number of toxic events associated with doses greater than the MTD, and the increase in the probability of toxicity (over the target of 30 per cent) associated with doses greater than

Table II. Example – patient-specific (true) model results.

Model	N_0	a_0	MTD type	Estimated 1st dose	Estimated MTD (SE)	Number toxic	Number of dose > MTD	Number of toxic > MTD	Range of percentile > 30% target
Int + dose + sex	200	0.9	female	−0.740	−0.748 (0.056)	8	0	0	0%
			male	−0.474	−0.484 (0.043)				
			average	—	−0.616 (0.036)				
		0.5	female	−0.746	−0.751 (0.068)	8	1	1	0.6%
			male	−0.471	−0.490 (0.053)				
			average	—	−0.621 (0.045)				
		0.1	female	−0.792	−0.770 (0.089)	8	3	3	0.2– 3.6%
			male	−0.450	−0.526 (0.078)				
			average	—	−0.648 (0.062)				
	100	0.9	female	−0.672	−0.704 (0.062)	9	10	5	3.8– 8.6%
			male	−0.478	−0.500 (0.052)				
			average	—	−0.602 (0.042)				
		0.5	female	−0.678	−0.735 (0.079)	9	7	5	1.2– 9.4%
			male	−0.474	−0.511 (0.067)				
			average	—	−0.623 (0.054)				
		0.1	female	−0.739	−0.805 (0.084)	8	2	2	0.6– 13.1%
			male	−0.444	−0.569 (0.077)				
			average	—	−0.687 (0.059)				
	50	0.9	female	−0.738	−0.778 (0.073)	7	6	4	0.8– 3.3%
			male	−0.558	−0.553 (0.064)				
			average	—	−0.645 (0.052)				
		0.5	female	−0.755	−0.741 (0.077)	7	6	4	0.7– 4.3%
			male	−0.552	−0.539 (0.069)				
			average	—	−0.640 (0.054)				
		0.1	female	−0.810	−0.729 (0.043)	7	7	4	1.5–10.2%
			male	−0.502	−0.508 (0.043)				
			average	—	−0.618 (0.031)				

Table III. Example – standard model results.

Model	N_0	a_0	MTD type	Estimated 1st dose	Estimated MTD (SE)	Number toxic	Number of dose > MTD	Number of toxic > MTD	Range of percentile > 30% target
Int+dose	200	0.9	standard	−0.610	−0.607 (0.036)	6	10	5	13.3–14.3%
		0.5	standard	−0.613	−0.606 (0.044)	6	10	5	12.5–14.7%
		0.1	standard	−0.630	−0.674 (0.081)	8	10	4	11–17.1%
	100	0.9	standard	−0.573	−0.588 (0.041)	8	10	7	16.2–19.5%
		0.5	standard	−0.576	−0.582 (0.048)	7	10	6	16.6–19.8%
		0.1	standard	−0.508	−0.530 (0.054)	7	10	6	25.1–30.7%
	50	0.9	standard	−0.654	−0.646 (0.051)	6	10	5	6.8–11%
		0.5	standard	−0.662	−0.608 (0.051)	5	10	5	10.8–13.5%
		0.1	standard	−0.533	−0.556 (0.033)	6	11	4	20.8–25.2%

Table IV. Example – O'Quigley-type patient-specific model results.

Model	N_0	a_0	MTD type	Estimated 1st dose	Estimated MTD (SE)	Number toxic	Number of dose >MTD	Number of toxic >MTD	Range of percentile >30% target
Dose + sex	200	0.9	female	−1.244	−1.005 (0.210)	8	10	4	28.9–33.9%
			male	−0.179	−0.209 (0.129)				
			average	—	−0.607 (0.119)				
		0.5	female	−1.207	−0.953 (0.246)	5	10	4	27.8–35.8%
			male	−0.172	−0.223 (0.164)				
			average	—	−0.588 (0.145)				
		0.1	female	−2.202	−0.802 (0.427)	4	6	4	10.9–53.3%
			male	−0.068	−0.672 (0.468)				
			average	—	−0.737 (0.330)				
	100	0.9	female	−0.990	−0.902 (0.279)	9	10	6	8.1–25.3%
			male	−0.266	−0.414 (0.210)				
			average	—	−0.658 (0.181)				
		0.5	female	−1.063	−0.887 (0.364)	9	10	6	0.9–30.3%
			male	−0.245	−0.511 (0.302)				
			average	—	−0.699 (0.253)				
		0.1	female	−1.549	−0.735 (0.304)	6	10	4	5.9–48.4%
			male	−0.022	−0.141 (0.271)				
			average	—	−0.438 (0.178)				
	50	0.9	female	−1.382	−0.860 (0.357)	5	0	0	0%
			male	−0.499	−0.547 (0.310)				
			average	—	−0.704 (0.247)				
		0.5	female	−1.187	−0.856 (0.419)	6	2	2	0.3–3%
			male	−0.421	−0.569 (0.382)				
			average	—	−0.713 (0.292)				
		0.1	female	−0.589	−0.830 (0.564)	9	9	8	0.2–60.7%
			male	−0.181	−0.641 (0.634)				
			average	—	−0.736 (0.457)				

Table V. Example – O'Quigley model results.

Model	N_0	a_0	MTD type	Estimated 1st dose	Estimated MTD (SE)	Number toxic	Number of dose >MTD	Number of toxic >MTD	Range of percentile >30% target
Dose	200	0.9	standard	−0.785	−0.684 (0.149)	4	6	3	0.8–3.7%
		0.5	standard	−0.808	−0.631 (0.166)	4	7	3	3.8–8%
		0.1	standard	−0.633	−0.431 (0.153)	4	13	4	0.2–33.4%
	100	0.9	standard	−0.619	−0.595 (0.155)	6	10	5	10.1–16.4%
		0.5	standard	−0.554	−0.524 (0.156)	5	10	4	17.1–22.3%
		0.1	standard	−0.450	−0.746 (0.331)	6	3	1	1.7%

the MTD. In Tables III and V, since the estimated MTD falls between the true female- and male-specific MTDs in all but one case, it is not surprising that there are administered doses which overshot their marks by a large percentage. In the ideal case of Table II, there were overshoots, but they were small in size, indicating closeness to the target doses. In the corresponding model in Table IV with no intercept, the overshoots were larger, reflecting the difference between the true and working models. In general, the magnitude of the overshoots increased with decreasing informativeness in the prior. These results underscore the importance of modelling using patient-specific covariates and employing informative priors.

9. DISCUSSION

We presented the concept of patient-specific maximum tolerated doses by augmenting the dose–response model with other covariates. This notion can be applied in a non-informative prior context. However, it is desirable to incorporate as much prior information as possible in order to determine the MTD more accurately for both the phase I patients and those in later phases. We proposed three methods to elicit informative priors on the dose and other covariate parameters based on the notion of prior predictive responses. This is a straightforward approach to elicitation because one does not have to think of distributions on parameters, *per se*; rather, the focus is on responses and covariates. Thus, the interaction between statistician and clinician is eased. Also, a tuning parameter can be incorporated into the prior structure in order to adjust the sharpness of the prior in case the clinician feels it justified and/or for regulatory reasons. Note that we used a continuum of doses within CRM instead of a set of discrete doses to simplify computation. However, our extensions can also be applied in the discrete dose case.

In order to avoid too flat a posterior surface and numerical problems in computing, it is recommended that the prior sample size and tuning parameter be sufficiently high. As illustrated in the first example for the patient-specific models with and without intercept, all patient-specific and average MTDs were well-estimated as long as the prior was informative. Thus, it appears that even if one does not use the true model, as long as it has dose and covariate parameters, it will estimate the necessary doses if the prior is informative. Also, the first doses estimated in this context are reasonable if the prior is informative. This makes sense because one would not want to use these estimated first doses unless one were very sure of the prior information. In the cases where the prior is less informative, it appears it would

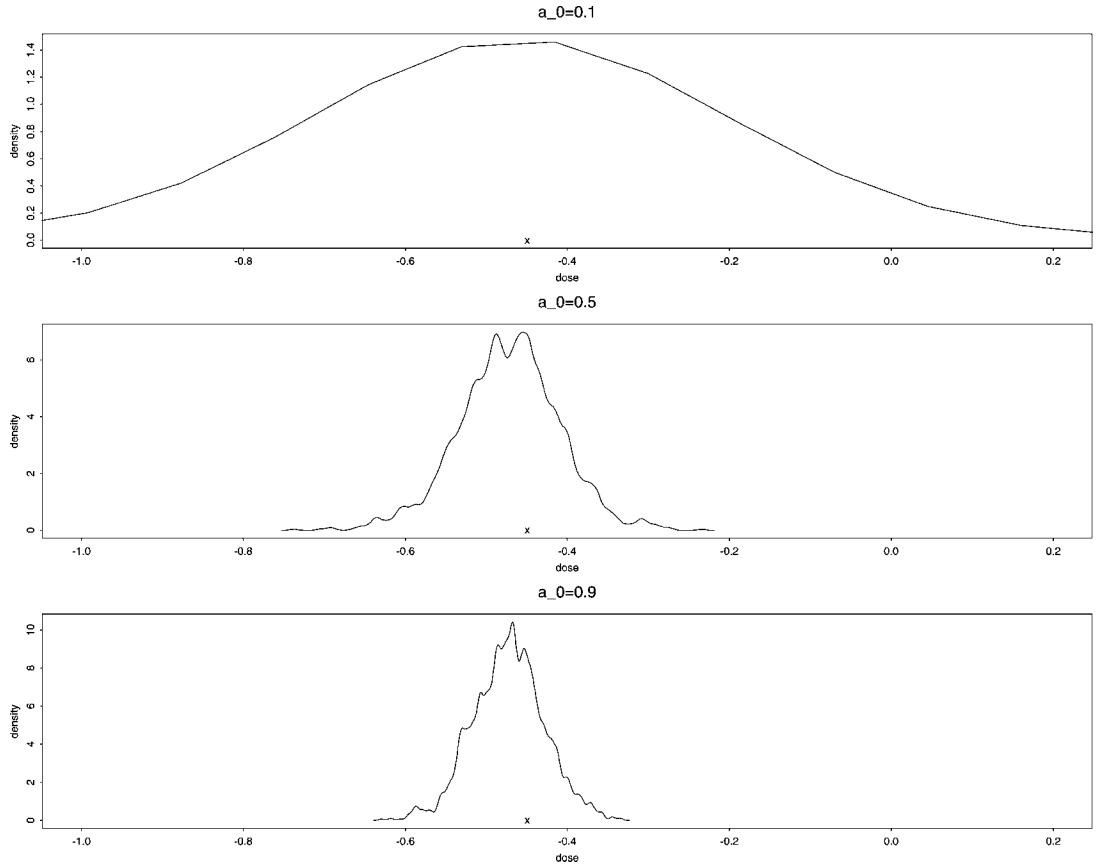

Figure 1. Distributions of first does for males under varying a_0. $\times$ indicates the true male MTD.

be better to be cautious and follow current practice by starting at lower doses. However, dose estimates can be updated with the proposed method.

It is recommended to compare estimates at different levels of the tuning parameter a_0. Initially we placed a beta prior on a_0 in order to be able to average over this parameter *a posteriori* and obtain a single MTD estimate. However, this proved both numerically and practically ill-advised. Numerically, the data do not say anything about a_0. This leads to numerical instability in some Gibbs sampling situations. From a data analytic point of view, it proved better to fix a_0 and compare estimates because, as shown in the tables, estimates tended to be more variable when one of the false models were used.

The recommendation to compare the effects of different priors is especially important when determining the starting dose. In the case of using the true model as the working model, a prior sample size of 200 and varying a_0 (Figure 1), one can see how an increasingly informative prior narrows the distribution of the estimated first dose for males toward the true dose (obtained via S-plus [11] function `density ()`). Investigators can examine Table VI for percentiles of these distributions to determine if they wish to start with lower or higher doses.

Table VI. Percentiles of first dose distribution for males under varying a_0.

a_0	5%	10%	25%	50%	75%	90%	95%
0.1	−0.732	−0.634	−0.538	−0.461	−0.381	−0.273	−0.191
0.5	−0.574	−0.550	−0.512	−0.470	−0.433	−0.397	−0.372
0.9	−0.543	−0.529	−0.502	−0.472	−0.444	−0.417	−0.400

The advantage of our Bayesian approach is that such distribution and percentile estimates are available to aid decision-making.

The use of x_P for patient treatment, however, does not preclude the calculation of the average MTD x_A during intermediate steps of the trial for overall monitoring purposes. In fact, large discrepancies between x_A and x_P may be of both clinical and regulatory value. Ultimately, perhaps the dosages on the approved drug label may be modified to recommend different dosages based on sex, age, and/or other important covariates.

ACKNOWLEDGEMENTS

Both authors gratefully acknowledge the support of NIH grants CA 70101-01 and CA 74015-01. The first author also gratefully acknowledges the support of the NIH-NCI Biostatistics/Epidemiology Training Grant in Biostatistics, number 5 T32 CA09337.

REFERENCES

1. Storer B. Design and analysis of phase I clinical trials. *Biometrics* 1989; **45**:925–937.
2. O'Quigley J, Pepe M, Fisher L. Continual reassessment method: a practical design for phase I clinical trials in cancer. *Biometrics* 1990; **46**:33–48.
3. Ibrahim JG, Ryan LM, Chen M-H. Use of historical controls to adjust for covariates in trend tests for binary data. *Journal of the American Statistical Association* 1998; **93**:1282–1293.
4. Chen M-H, Ibrahim JG, Yiannoutsos C. Prior elicitation and Bayesian computation for logistic regression models with applications to variable selection. *Journal of the Royal Statistical Society, Series B* 1998; **61**:223–242.
5. Ibrahim JG, Laud PW. A predictive approach to the analysis of designed experiments. *Journal of the American Statistical Association* 1994; **89**:309–319.
6. Laud PW, Ibrahim JG. Predictive model selection. *Journal of the Royal Statistical Society B* 1995; **57**: 981–986.
7. Chen M-H, Ibrahim JG, Shao Q-M. Power prior distributions for generalized linear models. (1997).
8. Faries D. Practical modifications of the Continual Reassessment Method for phase I cancer clinical trials. *Journal of Biopharmaceutical Statistics* 1994; **4**:147–164.
9. Gilks WR, Best NG, Tan KKC. Adaptive rejection Metropolis sampling within Gibbs sampling. *Applied Statistics* 1995; **44**:455–472.
10. Gilks WR. *C code for ARMS – Adaptive Rejection Metropolis Sampling*. 1995.
11. Mathsoft Inc. S-plus Version 3.4 Release 1 for Sun SPARC, SunOS 5.3. Mathsoft, Inc., 1996.

