

Critical aspects of the Bayesian approach to phase I cancer trials

Beat Neuenschwander^{*,†}, Michael Branson and Thomas Gsponer

Novartis Pharma AG, Lichstrasse 35, 4056 Basel, Switzerland

SUMMARY

The Bayesian approach to finding the maximum-tolerated dose in phase I cancer trials is discussed. The suggested approach relies on a realistic dose–toxicity model, allows one to include prior information, and supports clinical decision making by presenting within-trial information in a transparent way. The modeling and decision-making components are flexible enough to be extendable to more complex settings. Critical aspects are emphasized and a comparison with the continual reassessment method (CRM) is performed with data from an actual trial and a simulation study. The comparison revealed similar operating characteristics while avoiding some of the difficulties encountered in the actual trial when applying the CRM. Copyright © 2008 John Wiley & Sons, Ltd.

KEY WORDS: maximum-tolerated dose; continual reassessment method; logistic model

1. INTRODUCTION

Phase I cancer trials are aimed at finding the maximum-tolerated dose (MTD). This is traditionally achieved by exposing patients to dose levels that are adaptively selected based on cumulative toxicity data. The simplest and most widely used approach classifies safety events into two categories, dose-limiting toxicities (DLT) and non-DLTs, with the aim of finding the dose with probability of DLT closest to a targeted probability, usually 25 or 30 per cent.

Although there is agreement on the need for adaptive dose selection in phase I cancer trials, various designs have been discussed in the literature. They range from simple up-and-down schemes like the ‘3+3’ design [1] to model-based approaches, the most prominent being the continual reassessment method (CRM) by O’Quigley *et al.* [2]. The impact of both designs on clinical practice has been considerable, and they have triggered new algorithmic and model-based designs.

Up-and-down designs are algorithmic in that dose assignments follow from fixed rules that rely exclusively on observed toxicity data. The simplest up-and-down design—‘go to next lower (higher) level if an (no) event has occurred’—targeting a probability of 50 per cent was introduced

*Correspondence to: Beat Neuenschwander, Novartis Pharma AG, Lichstrasse 35, 4056 Basel, Switzerland.

†E-mail: beat.neuenschwander@novartis.com

in a non-medical context [3]. Refinements of these designs have been developed in the meantime [4, 5]. On the other hand, model-based approaches build on a dose–toxicity model that is updated with incoming trial data, from which dose recommendations for the next patient cohort are obtained. Both approaches constitute powerful tools for dose finding in phase I cancer trials.

We will focus on the model-based approach by discussing its critical aspects and comparing one- and two-parameter dose–toxicity models (for non-parametric approaches see [6–8]), which we implemented in our trials as an improvement over the rudimentary ‘3+3’ design. The ‘3+3’ design is still popular as clinicians often feel uneasy about the starting dose recommended by a modeling approach, a too aggressive dose escalation, the need of prior specifications and computational difficulties. These concerns have been taken seriously, leading to various adaptations of modeling approaches [9–13]. For example, it has been suggested to start at the lowest dose, not skip dose levels, and to use the maximum likelihood instead of the Bayesian approach. These adaptations address the inherent issues only partially and lead to *ad hoc* rules such as ‘start at the lowest dose’ or ‘do not skip a dose’, with no regard for the magnitude of the respective dose increment and the associated risk.

Our objective is to identify and clarify some of the potential shortcomings of the modeling paradigm. It will be shown that problems do not arise from the inadequacy of the modeling paradigm in general but rather from an oversimplified implementation. We will argue that these difficulties are rooted in the use of too simplistic inferential summaries and dose–toxicity models. We propose an approach that is driven by appropriate inferences (full posterior distributions) derived from a realistic dose–toxicity model, enabling transparent decision making by clinical teams.

We will discuss the modeling, inferential and decision-making aspects for standard MTD finding phase I cancer trials, but the conclusions will apply to other situations as well. Extensions of the model-based methodology for phase I designs are numerous, addressing topics such as individualized dosing [14, 15], multiple outcomes [16], time to event [17–19], longitudinal data [20], population heterogeneity and covariates [21–23], combination trials [24, 25], continuous [26] and delayed outcomes [27].

In Section 2 an in-house phase I trial will be introduced. Section 3 summarizes two widely used parametric dose–toxicity models. Dose selection procedures will be discussed in Section 4, and in Section 5 a critical comparison of the different model-based approaches will be presented for the study introduced in Section 2. The specification of a prior distribution will be presented in Section 6. A simulation study exploring operating characteristics of the designs will be presented in Section 7. Finally, computational issues will be covered in Section 8 and a discussion is provided in Section 9.

2. APPLICATION (PART I)

This study is an open-label, multicenter, non-comparative, dose-escalation cancer trial designed to characterize the safety, tolerability and pharmacokinetic profile of a drug and to determine its MTD. The pre-defined doses were 1, 2.5, 5, 10, 15, 20, 25, 30, 40, 50, 75, 100, 150, 200 and 250 mg, and it was decided to apply the modified CRM (MCRM) (see Sections 3 and 4), i.e. not to skip dose levels during dose escalations.

The first cohort of patients was treated at 1 mg. No DLTs were observed for the first four cohorts of patients, and the clinical team then decided to skip two dose levels (contradicting the planned

Table I. Posterior summaries for probabilities of DLT (CRM).

	Doses									
	1	2.5	5	10	15	20	25	30	40	50
No. of patients	3	4	5	4	—	—	2	—	—	—
No. of DLTs	0	0	0	0	—	—	2	—	—	—
<i>(A) Posterior summaries (original skeleton)</i>										
Skeleton (CRM)	0.010	0.015	0.020	0.025	0.030	0.040	0.050	0.100	0.170	0.300
Mean	0.069	0.085	0.099	0.111	0.123	0.144	0.163	0.242	0.330	0.465
Std. dev.	0.055	0.062	0.068	0.072	0.076	0.082	0.087	0.101	0.109	0.108
<i>(B) Posterior summaries (equidistant skeleton)</i>										
Skeleton (CRM)	0.063	0.125	0.188	0.250	0.313	0.375	0.438	0.500	0.563	0.625
Mean	0.024	0.054	0.090	0.130	0.176	0.226	0.281	0.341	0.405	0.475
Std. dev.	0.030	0.051	0.069	0.084	0.097	0.107	0.115	0.119	0.120	0.117

Columns in boldface highlight the recommended dose for the next cohort.

MCRM). At 25 mg, two DLTs were seen in two patients. For information regarding the first 10 doses see Table I.

The data from the fifth cohort led to discussions about the dose for the next cohort. The recommendation based on the CRM was to escalate to a higher dose. A critical re-analysis of these data and its implications will be discussed in the remaining sections.

3. ONE- AND TWO-PARAMETER MODELS

Let \mathcal{D} be the set of J available doses, $\mathcal{D} = \{d_1, \dots, d_J\}$. Further, let $\pi_\theta(d)$ be the probability of DLT at dose d , where θ denotes model parameters. The CRM is based on the one-parameter power model

$$\pi_\theta(d) = c_d^\theta, \quad \theta > 0$$

where the values of c_d are monotonically increasing, pre-specified ('skeleton') probabilities. A suggested prior for $\log(\theta)$ is a normal distribution with mean 0 and variance 1.34^2 [2]. If the prior median of θ is 1, c_d is the prior median at dose d . However, the skeleton probabilities are structural and not part of the prior.

Alternative one-parameter models are sometimes used [28, 29]. In the sequel we will only investigate the power model, mainly for reasons of simplicity and the fact that conclusions will be similar for other one-parameter models. It is important to note that one-parameter models cannot realistically describe a dose-toxicity curve and should therefore be considered as working models to identify a single chosen quantile (e.g. 30 per cent quantile) from the unknown dose-toxicity curve by means of inverse regression.

The two-parameter logistic model parameterizes the probability of a DLT using the logistic transform,

$$\text{logit}\{\pi_\theta(d)\} = \log \alpha + \beta \log(d/d^*), \quad \alpha, \beta > 0$$

where $\theta = (\log \alpha, \log \beta)$ and d^* is a reference dose allowing for the interpretation of α as the odds of a DLT at d^* . This model is more flexible and allows for a more realistic representation of the underlying dose–toxicity curve. As with the one-parameter power model, the two-parameter model is also widely used in practice. Moreover, the latter serves as a building block for more sophisticated extensions, e.g. [21, 25].

4. DOSE RECOMMENDATION

In phase I cancer trials, dose selection for the next patient cohort depends on information from patients already in the trial. In the Bayesian setting this information is the posterior distribution of the model parameters θ , which implies posterior distributions for the probabilities of a DLT, $\pi_\theta(d)$, that will then be used to choose a dose for the next patient cohort.

The standard dose recommendation rule, originally proposed for the CRM [2], is based on point estimates for the probability of a DLT at each dose, $\hat{\pi}_\theta(d)$: the next cohort will be treated at the dose with mean posterior probability of DLT closest (in absolute value or from below) to the pre-specified target probability, e.g. 30 per cent. The rule is intuitive and has the computational advantage that only the first posterior moment is needed. However, the rule is problematic in that uncertainty about $\pi_\theta(d)$ is ignored, which is of particular concern at the start of the trial. For example, suppose the target dose has a probability of DLT of 30 per cent. For simplicity, take two Beta distributions to represent posterior distributions at a specific dose, (i) Beta(0.6, 1.4) and (ii) Beta(6.6, 15.4), both have a mean of 30 per cent. In (i), with probability 0.168 the probability of DLT is greater than 60 per cent and this probability increases to 0.366 for the probability of DLT greater than 35 per cent. In (ii), these probabilities reduce to 0.002 and 0.29, respectively. Hence, decisions based on point estimates can be risky and better inferential summaries may be needed for good decision making (see also [30]).

A summary of the posterior information that has been particularly useful in our trials classifies the probability of a DLT into four categories:

Under-dosing:	$\pi_\theta(d) \in (0, 0.20]$
Targeted toxicity:	$\pi_\theta(d) \in (0.20, 0.35]$
Excessive toxicity:	$\pi_\theta(d) \in (0.35, 0.60]$
Unacceptable toxicity:	$\pi_\theta(d) \in (0.60, 1.00]$

After each patient cohort, the posterior distributions of $\pi_\theta(d)$ are summarized for each dose by the four probabilities of under-dosing, targeted, excessive and unacceptable toxicity. These probabilities will be examined for all candidate doses, and, depending on the level of conservatism the next dose will be selected. Currently in our trials the dose recommendation relies on maximizing the probability of targeted toxicity while controlling the probability of excessive or unacceptable toxicity at 25 per cent. Similar conservative escalation rules have been discussed in the literature [11].

A fully Bayesian approach would be decision analytic with a formal loss function, where the optimal decision is the one that minimizes the corresponding Bayes risk. For the simple loss

function

$$L(\theta, d) = \begin{cases} \ell_1 = 1 & \text{if } \pi_\theta(d) \in (0, 0.2] \\ \ell_2 = 0 & \text{if } \pi_\theta(d) \in (0.2, 0.35] \\ \ell_3 = 1 & \text{if } \pi_\theta(d) \in (0.35, 0.6] \\ \ell_4 = 2 & \text{if } \pi_\theta(d) \in (0.6, 1] \end{cases}$$

and the Beta distributions (i) and (ii) discussed previously, the Bayes risk ($= \ell_1 P\{\pi_\theta(d) \in (0, 0.2]\} + \ell_2 P\{\pi_\theta(d) \in (0.2, 0.35]\} + \ell_3 P\{\pi_\theta(d) \in (0.35, 0.6]\} + \ell_4 P\{\pi_\theta(d) \in (0.6, 1]\}$) is 0.999 for (i) and 0.442 for (ii). This clearly indicates the increased risk in (i) relative to (ii). Dose escalation based on this type of loss function will be illustrated in Section 5.

The Bayesian decision-analytic approach has been discussed [31–34], with loss functions comprising design optimality related to estimation efficiency and criteria for controlling the risk of overdosing. Although the decision-analytic approach is completely transparent, the specification of loss functions may be difficult and may complicate the interactions with clinical teams.

The approach we are currently using and advocating here is not decision analytic but based on a good representation (interval probabilities) of the posterior distributions of $\pi_\theta(d)$ for each dose. It is practical, pragmatic and has served us well in our trials as a basis for dose-escalation decisions. It should be emphasized that depending on the compound and/or indication under study, the DLT classification (e.g. the target interval) and the recommendations derived from the posterior distributions should be tailored to the clinical setting.

5. APPLICATION (PART II)

In this section the case study from Section 2 is revisited. Recall that the data from the fifth cohort led to discussions about the dose recommendation for the next cohort. Applying the one-parameter CRM based on the power model, the recommendation was 40 mg (or 30 mg) since the corresponding mean posterior probability of DLT was closest in absolute value (from below) to the target of 30 per cent (Table I(A), boldface highlighting the dose recommendations based on absolute value). The CRM thus recommended a dose that was counterintuitive and caused some irritation among the clinical team. Not only it was highly unlikely that the team would escalate to a higher dose but it was also eventually decided to de-escalate to 20 mg.

A closer look at the posterior distribution (Figure 1) reveals a high precision (an artifact of the special structure of the one-parameter model) for the toxicity at the highest dose level even though no patient has been treated at this dose. This makes it clear that the one-parameter model should not be used to derive inferential results such as uncertainty statements about probabilities of DLT, since the model was only designed as a working model to identify the MTD. Note that a slightly modified version of the CRM is the one-parameter logistic regression model where the intercept is fixed to 3 and the slope is the only model parameter. The same criticism holds for this model [35].

The more flexible two-parameter logistic model combined with dose recommendations based on posterior interval probabilities of Section 4 leads to more sensible recommendations. The suggestion is to de-escalate and to treat the next patient cohort at 20 mg. Table II(A) summarizes the posterior distribution and Bayes risks for examples of three different loss functions representing

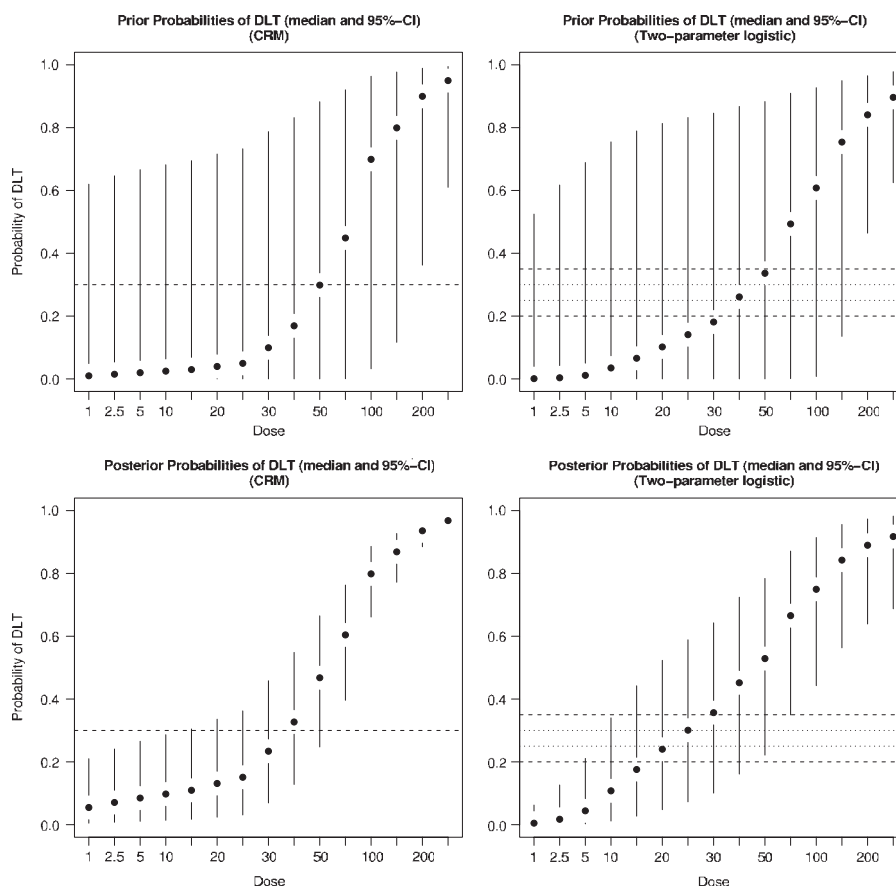


Figure 1. Prior (top panels) and posterior (bottom panels) probabilities of DLT based on CRM (left panels) and two-parameter logistic (right panels) for the data in the application. Dashed line for CRM indicates target probability of 30 per cent, dashed lines for two-parameter logistic indicate target toxicity interval.

aggressive ('1-0-1-1'), conservative ('1-0-1-2') and very conservative ('1-0-2-4') dose-escalation behavior.

For the comparison of the two models to be fair, the prior for the two-parameter logistic model was chosen such that the prior 2.5, 50 and 97.5 per cent quantiles for the probabilities of a DLT were as close as possible to the ones from the one-parameter CRM (see Section 6). For the power model a Gaussian distribution with mean zero and standard deviation 1.34 was used in the actual trial, with skeleton probabilities 0.01, 0.015, 0.020, 0.025, 0.03, 0.04, 0.05, 0.10, 0.17, 0.30, 0.45, 0.70, 0.80, 0.90, 0.95. The best matching bivariate normal prior for the two-parameter logistic model had parameters $\mu_1 = 2.15$, $\mu_2 = 0.52$, $\sigma_1 = 0.84$, $\sigma_2 = 0.80$ and $\rho = 0.20$ (using reference dose $d^* = 250$ mg). Figure 1 shows the corresponding prior and posterior quantiles for the CRM and two-parameter logistic model, for each dose level.

Another issue that may be of concern is the choice of the structural skeleton probabilities in the power model. These probabilities were indeed somewhat special in that they were chosen (based

Table II. Posterior summaries for probabilities of DLT (two-parameter logistic model): interval probabilities, mean and standard deviation and Bayes risk (see Section 4) for three loss functions.

	Doses									
	1	2.5	5	10	15	20	25	30	40	50
No. of patients	3	4	5	4	—	—	2	—	—	—
No. of DLTs	0	0	0	0	—	—	2	—	—	—
<i>(A) Posterior summaries (prior similar to CRM prior)</i>										
Under-dosing	1.000	0.996	0.970	0.809	0.581	0.377	0.234	0.140	0.050	0.017
Target	0.000	0.004	0.029	0.170	0.324	0.401	0.393	0.343	0.212	0.117
Excessive	0.000	0.000	0.001	0.021	0.094	0.216	0.352	0.464	0.574	0.544
Unacceptable	0.000	0.000	0.000	0.000	0.001	0.006	0.021	0.052	0.164	0.322
Risk 1-0-1-1	1.000	0.996	0.971	0.830	0.676	0.599	0.607	0.657	0.788	0.883
Risk 1-0-1-2	1.000	0.996	0.971	0.830	0.677	0.605	0.628	0.710	0.952	1.205
Risk 1-0-2-4	1.000	0.996	0.972	0.852	0.773	0.833	1.021	1.279	1.855	2.393
Mean	0.011	0.029	0.061	0.127	0.191	0.252	0.309	0.360	0.449	0.522
Std. dev.	0.018	0.034	0.056	0.088	0.111	0.126	0.136	0.142	0.148	0.147
<i>(B) Posterior summaries (non-informative default prior)</i>										
Under-dosing	1.000	0.998	0.968	0.740	0.476	0.287	0.173	0.110	0.051	0.027
Target	0.000	0.002	0.030	0.215	0.337	0.350	0.305	0.247	0.148	0.093
Excessive	0.000	0.000	0.001	0.044	0.179	0.319	0.413	0.450	0.432	0.357
Unacceptable	0.000	0.000	0.000	0.000	0.009	0.043	0.109	0.193	0.369	0.523
Risk 1-0-1-1	1.000	0.998	0.970	0.785	0.663	0.650	0.695	0.753	0.852	0.907
Risk 1-0-1-2	1.000	0.998	0.970	0.785	0.672	0.693	0.804	0.946	1.222	1.430
Risk 1-0-2-4	1.000	0.998	0.971	0.830	0.869	1.099	1.436	1.782	2.392	2.832
Mean	0.010	0.028	0.065	0.148	0.230	0.305	0.372	0.429	0.523	0.593
Std. dev.	0.014	0.030	0.056	0.099	0.132	0.155	0.171	0.180	0.189	0.189

Columns in boldface highlight the recommended dose for the next cohort. Upper panel: prior closest to CRM prior (in application); lower panel: default prior from Section A.1.

on pre-clinical information) to represent a relatively steep dose–toxicity curve. We checked the robustness of conclusions by using equidistant skeleton probabilities that increase linearly from 0.06 to 0.62 for the first ten dose levels. Table I-B summarizes the results based on the equidistant skeleton probabilities. For this skeleton, the CRM recommends to stay at the same dose, namely 25 mg (mean posterior probability of DLT=28.1 per cent), which is still a questionable decision. This is not to say that there is no way to revise the CRM for this example by changing the skeleton probabilities; a detailed proposal of how this can be achieved is given in [36].

In the Bayesian approach, sensitivity to the choice of prior is an issue if data are sparse. Since one prior for the two-parameter logistic model was chosen to be similar to the one used for the CRM, we also used a second prior (non-informative, default prior of Appendix A.1) for the two-parameter model.

Table II shows the results based on the original (A) and the non-informative default prior (B) with parameters $\mu_1=2.27$, $\mu_2=0.26$, $\sigma_1=1.98$, $\sigma_2=0.40$ and $\rho=-0.16$.

An interesting feature of the application is the fact that the clinical team decided to skip two dose levels after four cohorts. This was a violation of the planned MCRM, and there was no clear rationale for this decision. Although this design violation has no effect when one is following

Table III. Dose recommendations under two data scenarios for one-parameter power model and two-parameter logistic model.

	Doses										Next dose	
	1	2.5	5	10	15	20	25	30	40	50	One-parameter	Two-parameter
<i>Original data</i>												
No. of patients	3	4	5	4	—	—	2	—	—	—		
No. of DLTs	0	0	0	0	—	—	2	—	—	—	30,40*	25 [†] 15,20* 10,15 [†]
<i>Data scenarios</i>												
No. of patients	3	4	5	4	3	3	2	—	—	—		
No. of DLTs	0	0	0	0	0	0	2	—	—	—	40*	30,40 [†] 30* 25 [†]
No. of DLTs	0	0	0	0	0	1	2	—	—	—	30*	25 [†] 25* 20 [†]

*Original prior (see text).

[†]Default prior (see text).

the Bayesian approach (likelihood principle), this is not true for the CRM [36]. This leaves the question about what could have happened had the original conservative MCRM or an alternative two-stage design [10] been applied. Table III shows two possible data scenarios and corresponding dose recommendations for the one-parameter power model and the two-parameter logistic model. Irrespective of the exact dose-escalation rule, the one-parameter model tends towards higher doses. These findings confirm what was discussed for the original data, although the differences are less striking.

In summary, the introduction of a second parameter helps to better describe the dose–toxicity curve, but the question on how to derive appropriate decisions from the posterior distribution remains. If dose recommendation for the next cohort is based on point estimates, the problem of recommending a dose with a large probability of exceeding the target interval remains (see Table II mean posterior summaries). Using the uninformative (default) prior and basing the decision on maximizing the targeted toxicity interval (category 2) while controlling for excessive or unacceptable toxicity the optimal dose would be 15 mg. A formal decision-analytic approach with three different loss functions representing aggressive, conservative and very conservative dose-escalation behavior leads to dose recommendations of 20, 15 and 10 mg, respectively.

Eventually, the trial was continued with a dose of 20 mg, a total of nine patients were enrolled at that dose, and two DLTs occurred. Then the trial was stopped and 20 mg was declared as the MTD.

6. PRIOR SPECIFICATIONS

In phase I oncology trials the use of vague priors is commonplace, but this practice should not be followed blindly. Our position is to incorporate external information whenever possible if it is well substantiated, i.e. based on empirical evidence. For example, data collected on a compound already tested in a certain type of cancer might provide valuable information that can be used to define a reasonably informative prior when testing the compound in another type of cancer. Or, when entering a trial that combines two anti-cancer agents, knowledge on each single agent

will usually be available and should therefore be included as prior information in the analysis of the combination trial; for a discussion of this situation, see [25]. In the following we introduce a procedure for deriving a prior for θ .

Depending on the underlying dose–toxicity model and its parametrization θ , formulating a prior distribution can be difficult, even more so if some or all of the model parameters have no direct interpretation. One could argue that in such situations a flat improper prior is a good choice. This means that the posterior is proportional to the likelihood, which has the undesirable implication that no formal analysis is possible as long as no DLTs have been observed in the trial [10]. Another consequence of flat priors for the parameters in the two-parameter logistic model is that the corresponding priors for the probabilities of a DLT will be U-shaped. In light of these difficulties we propose

- (1) to formulate prior information on the scale of interest, i.e. the probabilities of toxicity $\pi_\theta(d)$, $d \in \mathcal{D}$;
- (2) to define the prior distribution for θ in such a way that its implied prior information on the π -scale is in good agreement with the information in (1).

Concerning (1), prior information for dose d will be defined via K quantiles

$$q_d = \{q_d(p_1), \dots, q_d(p_K)\}, \quad \Pr\{\pi_\theta(d) \leq q_d(p_k)\} = p_k, \quad k = 1, \dots, K$$

At least two quantiles will be needed for each dose (e.g. a 95 per cent credible interval). Prior information will therefore be summarized by $J \times K$ quantiles, $Q = (q_{jk})$ with $q_{jk} = q_{d_j}(p_k)$, $j = 1, \dots, J$, $k = 1, \dots, K$.

Regarding (2), prior calibration consists of finding a prior for θ that is in good agreement with the quantiles in (1). Formally, this amounts to minimizing the discrepancy between the pre-specified prior quantiles Q and the quantiles Q' arising from the prior. We will minimize the following criterion:

$$C(Q, Q') = \max_{j,k} |q_{jk} - q'_{jk}|, \quad j = 1, \dots, J, \quad k = 1, \dots, K$$

The minimization can be achieved by a stochastic optimization using a Metropolis algorithm [37, 38]. We used stochastic optimization as it worked better than more standard non-stochastic optimizers. In our applications we currently summarize prior information at a given dose using the median and 95 per cent credible intervals, denoted by $q_d = \{q_d(0.025), q_d(0.5), q_d(0.975)\}$. An alternative prior derivation based on information on the model parameters α and β is outlined in Appendix A.2.

The quantile-based approach can be used for non-informative as well as informative settings. A proposal for deriving a default non-informative prior is described in Appendix A.1. As an example, Figure 2 summarizes the prior information for the study in Section 5 by means of the median and 95 per cent credible intervals for the probabilities of a DLT at each dose (left panel). The right panel displays the same probabilities (circles) together with the fitted probabilities (dashed line) obtained by stochastic optimization.

Often in phase I oncology trials the compound has already been tested in another trial. This information can be used to set an informative bivariate normal prior, but one should be careful in deciding on the weight of historical information relative to information to be obtained in the new study. If information from a reasonably large number of historical trials is available, the between-trial variation can be estimated, and this would lead to a down-weighting of historical information

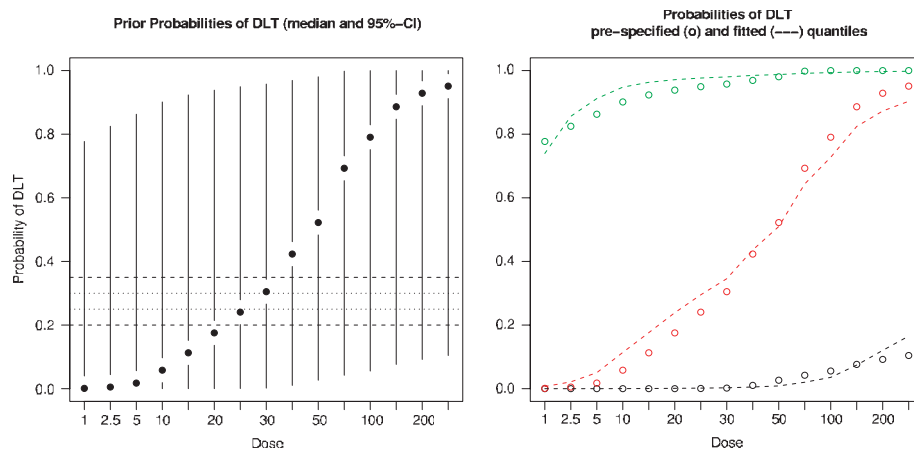


Figure 2. Non-informative default prior from Section A.1. Prior medians and 95 per cent credible intervals for each dose level (left panel) and the corresponding fitted probabilities of DLT (right panel).

under a random-effects model for treatment effects. For an overview of how to incorporate historical information see [35, 39].

7. SIMULATIONS

In order to assess the operating characteristics of the two-parameter logistic model a simulation study was performed for eight dose–toxicity scenarios (Table IV). Assuming seven dose levels, the scenarios were selected to represent a wide range of possible curves, a steep or flat curve with the true MTD at dose levels 2, 4 and 6, and two boundary scenarios with MTD at the highest or lowest dose (F7 and F1).

Five methods were compared: (1) CRM with target probability 0.27; (2) MCRM; (3) the logistic model with dose escalations and MTD definition based on the posterior mean (LRmean); (4) the logistic model with dose escalations and MTD definition based on maximizing the posterior probability of interval (0.20, 0.35] (LRcat) and (5) the logistic model with MTD definition based on maximizing the posterior probability for interval (0.20, 0.35] and 25 per cent overdose control (LRcat25). The mean-based MTD is the one closest to the target in absolute value. The methods can be classified by the underlying model, the estimate of MTD and the dose recommendation for the next cohort.

Table V summarizes the percentage of MTD recommendations, the number of patients per dose, and the average number of DLTs for the five methods using prior 1 of Table IV, based on 2000 simulated trials per scenario. For the percentage of correct recommendations of the MTD the methods perform similarly if the true MTD is at dose level 2 (scenarios S2 and F2). If the MTD is at dose level 4 (S4 and F4), the logistic model is slightly better. In these four scenarios, the conservative dose-escalation methods (MCRM and LRcat25) are very similar to the more aggressive escalation methods (CRM, LRmean and LRcat). Notable differences in performance exist if the MTD is at level 6 (S6 and F6). Here, a price has to be paid for the conservative escalations methods (MCRM and LRcat25), in that the percentages of correct MTD recommendations are smaller relative to the

Table IV. Simulation scenarios for probabilities of DLT and two prior settings with median and 95 per cent intervals for CRM and two-parameter logistic regression (LR).

Dose levels Doses	1 12.5	2 25	3 50	4 100	5 150	6 200	7 250
<i>Scenarios for probabilities of DLT</i>							
S2	0.11	0.27	0.52	0.76	0.84	0.90	0.94
F2	0.18	0.27	0.39	0.52	0.65	0.76	0.84
S4	0.01	0.04	0.11	0.27	0.52	0.76	0.90
F4	0.07	0.11	0.18	0.27	0.39	0.52	0.65
S6	0.01	0.01	0.03	0.04	0.11	0.27	0.52
F6	0.03	0.04	0.07	0.11	0.18	0.27	0.39
F7	0.01	0.03	0.04	0.07	0.11	0.18	0.27
F1	0.27	0.39	0.52	0.65	0.76	0.84	0.90
<i>Prior medians and 95 per cent intervals</i>							
First prior							
CRM*	0.05 (0.00,0.80)	0.10 (0.00,0.85)	0.15 (0.00,0.87)	0.25 (0.00,0.90)	0.40 (0.00,0.94)	0.60 (0.00,0.96)	0.90 (0.23,0.99)
LR [†]	0.01 (0.00,0.79)	0.03 (0.00,0.84)	0.09 (0.00,0.89)	0.30 (0.01,0.95)	0.53 (0.02,1.00)	0.65 (0.02,1.00)	0.73 (0.03,1.00)
Second prior							
CRM*	0.03 (0.00,0.78)	0.05 (0.00,0.81)	0.11 (0.00,0.85)	0.25 (0.00,0.90)	0.38 (0.00,0.93)	0.47 (0.00,0.95)	0.54 (0.00,0.96)
LR [‡]	0.03 (0.00,0.79)	0.05 (0.00,0.84)	0.11 (0.00,0.90)	0.25 (0.00,0.95)	0.38 (0.01,0.98)	0.47 (0.01,0.99)	0.54 (0.01,1.00)

* $\log(\alpha) \sim N(0, 1.34^2)$.[†] BVN: $(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho) = (-0.847, 0.381, 2.015, 1.207, 0)$.[‡] BVN: $(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho) = (-1.099, 0.000, 2.070, 1.000, 0)$.

less conservative methods. Conservative escalations slow down the convergence to the true MTD, which seems particularly relevant if the MTD is at a high dose. An even more striking difference can be seen when comparing the CRM with the methods based on the logistic model. For scenarios S6 and F6, the CRM is superior: the percentage of correct recommendations is somewhat larger, and the percentage of selecting a dose that is too high is considerably smaller. These results are not surprising when scrutinizing the prior. The choice of prior is indeed peculiar since *a priori* dose escalations to level 7 are fairly unlikely due to the fact that the 95 per cent prior interval for the probability of DLT at level 7 is (0.23,1.00). A truly non-informative prior should give more weight to small probabilities of a DLT. Given this particular choice of skeleton probabilities, it is not surprising that the CRM and MCRM perform very poorly if the MTD is at level 7. Of course, this does not speak against the CRM (the same could occur for the logistic model if the prior is not chosen carefully), it only illustrates the importance of a truly non-informative prior if good operating characteristics are sought for a wide range of dose–toxicity scenarios.

In order to allow for a more objective comparison, we considered a second prior with (1) the same prior medians for the power and logistic model and (2) wide confidence intervals that do not exclude dose escalations to any of the dose levels (Table IV). For prior 2, a comparison of the operating characteristics for the five methods (Table VI) reveals similar results to prior 1 for

Table V. Selection percentages of correct MTD (average number of allocated patients) per dose level, and average number of DLTs for various methods (see text), for sample sizes of 36 and 24 (prior 1).

	Scenario S2										Scenario F2									
	n=36										n=36									
	0.11	0.27	0.52	0.76	0.84	0.9	0.94	DLT			0.18	0.27	0.39	0.52	0.65	0.76	0.84	DLT		
CRM	16 (11.5)	76 (16.8)	9 (5.1)	0 (2.5)	0 (0)	0 (0)	0 (0)	10.4			24 (13.5)	53 (11)	23 (7.4)	1 (3.6)	0 (0.3)	0 (0.2)	0 (0)	10.5		
LRmean	12 (8.6)	80 (17.1)	8 (7.6)	0 (2.6)	0 (0)	0 (0)	0 (0)	11.5			20 (10.8)	50 (10.3)	28 (9.8)	3 (4.7)	0 (0.3)	0 (0)	0 (0)	11.3		
LRcat	8 (6.5)	80 (17.9)	12 (8.7)	0 (2.9)	0 (0)	0 (0)	0 (0)	12.3			14 (8.3)	47 (10.8)	34 (10.5)	4 (6.1)	0 (0.3)	0 (0)	0 (0)	11.9		
MCRM	14 (10.2)	76 (18.8)	10 (6.5)	0 (0.5)	0 (0)	0 (0)	0 (0)	10			25 (13.8)	53 (13)	22 (7.9)	1 (1.3)	0 (0)	0 (0)	0 (0)	9.8		
LRcat25	5 (13.7)	78 (17.4)	16 (4.9)	0 (0.1)	0 (0)	0 (0)	0 (0)	8.8			20 (16.1)	56 (13.1)	23 (6.5)	1 (0.3)	0 (0)	0 (0)	0 (0)	9.2		
<i>n=24</i>																				
CRM	19 (9.4)	67 (8.5)	14 (3.6)	0 (2.5)	0 (0)	0 (0)	0 (0)	7.1			27 (10.3)	44 (5.4)	26 (4.4)	3 (3.4)	0 (0.3)	0 (0.2)	0 (0)	7.1		
LRmean	16 (6.9)	70 (8.3)	15 (6.2)	0 (2.6)	0 (0)	0 (0)	0 (0)	8.2			21 (8.3)	41 (5)	32 (6.3)	6 (4.1)	0 (0.3)	0 (0)	0 (0)	7.7		
LRcat	11 (5.2)	68 (9.1)	21 (6.7)	0 (2.9)	0 (0)	0 (0)	0 (0)	8.7			17 (6.3)	35 (6.1)	38 (6.1)	9 (5.2)	0 (0.3)	0 (0)	0 (0)	8.1		
MCRM	19 (8.2)	65 (10.4)	16 (4.9)	0 (0.5)	0 (0)	0 (0)	0 (0)	6.6			29 (10.5)	43 (7.5)	26 (5)	2 (1.1)	0 (0)	0 (0)	0 (0)	6.5		
LRcat25	8 (10.1)	68 (9.3)	23 (4.5)	0 (0.1)	0 (0)	0 (0)	0 (0)	6			24 (11.5)	47 (7.1)	27 (5.1)	3 (0.3)	0 (0)	0 (0)	0 (0)	6.2		
<i>n=36</i>																				
CRM	0 (3.2)	0 (0.5)	14 (6.3)	79 (20.7)	7 (4.1)	0 (1.2)	0 (0)	9.4			0 (4.2)	3 (2.3)	29 (7.9)	51 (14)	17 (5.9)	1 (1.6)	0 (0)	8.9		
LRmean	0 (3.1)	0 (0.1)	16 (5.8)	80 (23.4)	4 (3.4)	0 (0.1)	0 (0.2)	9			0 (3.9)	2 (1.3)	25 (7.3)	58 (17.5)	13 (4.8)	2 (0.6)	0 (0.6)	9.1		
LRcat	0 (3)	0 (0.2)	6 (3)	86 (24.9)	8 (4.7)	0 (0.1)	0 (0.1)	9.8			0 (3.2)	1 (1.6)	17 (5.4)	60 (18.5)	20 (6.3)	1 (0.6)	0 (0.4)	9.4		
MCRM	0 (3.1)	0 (3.3)	12 (6.9)	78 (18.1)	9 (4.5)	0 (0.1)	0 (0)	8.2			0 (4.3)	3 (5.3)	33 (10.8)	50 (11.8)	13 (3.6)	0 (0.2)	0 (0)	7.5		
LRcat25	0 (3.2)	0 (1.6)	11 (22)	81 (8.3)	8 (0.9)	0 (0)	0 (0.1)	5.3			0 (4.5)	3 (5.9)	41 (19.9)	48 (4.6)	7 (0.8)	0 (0.1)	0 (0.2)	6.3		
<i>n=24</i>																				
CRM	0 (3.2)	0 (0.5)	20 (4.1)	69 (12.1)	11 (3)	0 (1.2)	0 (0)	6.2			0 (4.2)	6 (1.8)	29 (4.4)	42 (8.6)	21 (3.6)	2 (1.4)	0 (0)	5.8		
LRmean	0 (3.1)	0 (0.1)	18 (3.6)	75 (14.2)	7 (2.7)	0 (0.1)	0 (0.2)	5.9			0 (3.9)	4 (0.9)	25 (4.3)	53 (10.9)	16 (3)	3 (0.4)	0 (0.5)	5.9		
LRcat	0 (3)	0 (0.2)	8 (2.1)	80 (15)	12 (3.4)	0 (0.1)	0 (0.1)	6.4			0 (3.2)	2 (1.4)	17 (3.3)	55 (11.6)	22 (3.7)	2 (0.4)	0 (0.3)	6		
MCRM	0 (3.1)	0 (3.3)	15 (5.2)	70 (9.4)	15 (2.9)	0 (0.1)	0 (0)	4.9			0 (4.3)	7 (4.6)	33 (6.9)	43 (6.2)	15 (1.9)	1 (0.1)	0 (0)	4.5		
LRcat25	0 (3.1)	0 (1.3)	13 (13.9)	77 (5)	8 (0.6)	1 (0)	0 (0.1)	3.4			0 (4.4)	6 (3.8)	38 (12.4)	47 (2.7)	6 (0.4)	2 (0)	1 (0.1)	4		

Table V. *Continued.*

Scenario S6							Scenario F6						
0.01	0.01	0.03	0.04	0.11	0.27	0.52	0.03	0.04	0.07	0.11	0.18	0.27	0.39
<i>n</i> = 36							<i>n</i> = 36						
CRM	0 (3.1)	0 (0.1)	0 (0.3)	1 (4)	21 (8.4)	76 (19)	0 (3.4)	0 (0.4)	1 (1.4)	9 (6.8)	42 (10.9)	44 (11.9)	4 (1.3)
LRmean	0 (3.1)	0 (0)	0 (0.1)	1 (4.5)	26 (10.7)	62 (10.6)	0 (3.3)	0 (0.1)	1 (0.9)	11 (8.6)	29 (9.3)	36 (6.2)	24 (7.6)
LRcat	0 (3)	0 (0.1)	0 (0.1)	0 (4.1)	19 (10.5)	72 (11.9)	0 (3)	0 (0.4)	0 (0.6)	6 (7.8)	31 (10.7)	42 (6.9)	21 (6.6)
MCRM	0 (3.1)	0 (3)	0 (3.3)	1 (4.4)	30 (10.1)	66 (11.5)	0 (3.4)	0 (3.2)	1 (4.5)	14 (7.8)	51 (11.2)	31 (5.4)	2 (0.4)
LRcat25	0 (3.1)	0 (0.2)	0 (7.8)	3 (8.9)	23 (8.8)	56 (3.3)	0 (3.4)	0 (0.9)	2 (12.5)	27 (10)	31 (4.7)	22 (1.7)	17 (2.8)
<i>n</i> = 24							<i>n</i> = 24						
CRM	0 (3.1)	0 (0.1)	0 (0.3)	2 (3.8)	25 (5.5)	70 (10.4)	0 (3.4)	0 (0.4)	3 (1.2)	12 (5.4)	39 (6)	41 (6.9)	5 (0.7)
LRmean	0 (3.1)	0 (0)	0 (0.1)	2 (4.3)	30 (7.3)	51 (3.9)	0 (3.3)	0 (0.1)	1 (0.8)	14 (7)	29 (5.9)	27 (2.5)	28 (4.4)
LRcat	0 (3)	0 (0.1)	0 (0.1)	0 (4.1)	29 (7.6)	54 (4.4)	0 (3)	0 (0.4)	0 (0.6)	10 (6.7)	36 (6.7)	28 (2.7)	25 (3.9)
MCRM	0 (3.1)	0 (3)	0 (3.3)	3 (4.1)	39 (6)	54 (4.2)	0 (3.4)	0 (3.2)	4 (4.2)	20 (5.6)	48 (5.4)	26 (2)	3 (0.1)
LRcat25	0 (3.1)	0 (0.2)	0 (7.4)	8 (7.2)	35 (3.8)	34 (0)	0 (3.4)	0 (0.8)	3 (9.8)	36 (6.6)	26 (2.1)	17 (0.1)	18 (1.2)
Scenario F7							Scenario F1						
0.01	0.03	0.04	0.07	0.11	0.18	0.27	0.27	0.39	0.52	0.65	0.76	0.84	0.90
<i>n</i> = 36							<i>n</i> = 36						
CRM	0 (3.1)	0 (0.1)	0 (0.4)	2 (4.6)	16 (6.3)	64 (17)	78 (26.1)	21 (5.9)	1 (2.2)	0 (0.4)	0 (1.4)	0 (0)	0 (0)
LRmean	0 (3.1)	0 (0)	0 (0.3)	2 (5.6)	7 (5.9)	23 (5.1)	72 (22)	26 (8)	2 (4)	0 (1.9)	0 (0.1)	0 (0)	0 (0)
LRcat	0 (3)	0 (0.1)	0 (0.2)	0 (5.1)	7 (6.6)	28 (6)	63 (18.7)	32 (10.5)	5 (4.5)	0 (2.3)	0 (0.1)	0 (0)	0 (0)
MCRM	0 (3.1)	0 (3.1)	0 (3.6)	4 (5.4)	30 (9.7)	54 (9.3)	79 (26.1)	20 (7.3)	1 (2.3)	0 (0.3)	0 (0)	0 (0)	0 (0)
LRcat25	0 (3.1)	0 (0.3)	0 (9.1)	9 (9.6)	20 (5.1)	23 (2.2)	70 (27.2)	29 (6.3)	1 (2.4)	0 (0)	0 (0)	0 (0)	0 (0)
<i>n</i> = 24							<i>n</i> = 24						
CRM	0 (3.1)	0 (0.1)	0 (0.4)	3 (4.3)	19 (4.2)	60 (9.7)	73 (17)	23 (3.3)	4 (1.9)	0 (0.4)	0 (1.4)	0 (0)	0 (0)
LRmean	0 (3.1)	0 (0)	0 (0.3)	4 (5.2)	11 (4.8)	22 (2.3)	63 (14)	31 (4.6)	6 (3.5)	0 (1.9)	0 (0.1)	0 (0)	0 (0)
LRcat	0 (3)	0 (0.1)	0 (0.2)	2 (5)	14 (5.3)	25 (2.7)	55 (11.6)	34 (6.6)	10 (3.5)	0 (2.2)	0 (0.1)	0 (0)	0 (0)
MCRM	0 (3.1)	0 (3.1)	1 (3.6)	8 (4.7)	38 (5.6)	42 (3.5)	72 (16.9)	24 (4.8)	4 (2)	0 (0.3)	0 (0)	0 (0)	0 (0)
LRcat25	0 (3.1)	0 (0.3)	0 (8.1)	17 (7.4)	21 (2.8)	22 (0.1)	65 (17.4)	31 (4.2)	4 (2.4)	0 (0)	0 (0)	0 (0)	0 (0)

Table VI. Selection percentages of correct MTD (average number of allocated patients) per dose level, and average number of DLTs for various methods (see text), for sample sizes of 36 and 24 (prior 2).

	Scenario S2										Scenario F2									
	n=36										n=36									
	0.11	0.27	0.52	0.76	0.84	0.9	0.94	DLT			0.18	0.27	0.39	0.52	0.65	0.76	0.84	DLT		
CRM	19 (10.9)	69 (16)	12 (6.6)	0 (0.4)	0 (2.1)	0 (0)	0 (0)	0 (0)	11	25 (12.7)	47 (10.6)	27 (8.9)	1 (1.9)	0 (1.8)	0 (0)	0 (0.1)	10.9	25 (12.7)	47 (10.6)	27 (8.9)
LRmean	17 (9.5)	73 (14.9)	10 (8.3)	0 (1)	0 (0.1)	0 (2.3)	0 (0)	12.2	25 (10.1)	47 (9.4)	26 (10.5)	2 (3.3)	0 (0.4)	0 (2.2)	0 (0.1)	12.2	25 (10.1)	47 (9.4)	26 (10.5)	2 (3.3)
LRcat	9 (6.7)	80 (18.2)	11 (8.2)	0 (2.7)	0 (0)	0 (0)	0 (0)	12.1	15 (8.5)	49 (11.3)	33 (10.5)	3 (5.1)	0 (0.4)	0 (0.2)	0 (0)	11.8	15 (8.5)	49 (11.3)	33 (10.5)	3 (5.1)
MCRM	18 (10.7)	69 (16.9)	13 (7.9)	0 (0.5)	0 (0)	0 (0)	0 (0)	10.2	25 (13.2)	48 (11.9)	26 (9.3)	1 (1.5)	0 (0.1)	0 (0)	0 (0)	10.1	25 (13.2)	48 (11.9)	26 (9.3)	1 (1.5)
LRcat25	6 (13.3)	76 (17.2)	18 (5.1)	0 (0.3)	0 (0)	0 (0)	0 (0)	9	21 (15.8)	54 (12.7)	24 (6.7)	1 (0.8)	0 (0)	0 (0)	0 (0)	9.4	21 (15.8)	54 (12.7)	24 (6.7)	1 (0.8)
CRM	23 (8.3)	59 (8.5)	18 (4.7)	0 (0.4)	0 (2.1)	0 (0)	0 (0)	7.7	28 (9.5)	38 (5.6)	30 (5.4)	3 (1.6)	0 (1.8)	0 (0)	0 (0.1)	7.4	28 (9.5)	38 (5.6)	30 (5.4)	3 (1.6)
LRmean	24 (7)	61 (7)	15 (6.7)	0 (1)	0 (0.1)	0 (2.3)	0 (0)	9	26 (7)	38 (4.3)	30 (7.2)	6 (2.8)	0 (0.4)	0 (2.2)	0 (0.1)	8.6	26 (7)	38 (4.3)	30 (7.2)	6 (2.8)
LRcat	13 (5.4)	69 (9.4)	19 (6.4)	0 (2.7)	0 (0)	0 (0)	0 (0)	8.6	18 (6.4)	38 (6.3)	36 (6.3)	8 (4.4)	0 (0.4)	0 (0.2)	0 (0)	8	18 (6.4)	38 (6.3)	36 (6.3)	8 (4.4)
MCRM	21 (8.3)	60 (9.3)	18 (5.9)	0 (0.5)	0 (0)	0 (0)	0 (0)	6.9	28 (10)	40 (6.7)	29 (5.9)	3 (1.2)	0 (0.1)	0 (0)	0 (0)	6.7	28 (10)	40 (6.7)	29 (5.9)	3 (1.2)
LRcat25	9 (9.9)	68 (9.1)	23 (4.7)	0 (0.3)	0 (0)	0 (0)	0 (0)	6.2	24 (11.2)	44 (6.8)	29 (5.2)	3 (0.7)	0 (0)	0 (0)	0 (0)	6.4	24 (11.2)	44 (6.8)	29 (5.2)	3 (0.7)
Scenario S4																				
	n=36										n=36									
	0.01	0.04	0.11	0.27	0.52	0.76	0.90	DLT			0.07	0.11	0.18	0.27	0.39	0.52	0.65	DLT		
	0.01	0.04	0.11	0.27	0.52	0.76	0.90	DLT			0.07	0.11	0.18	0.27	0.39	0.52	0.65	DLT		
CRM	0 (3.1)	0 (0.8)	14 (7)	78 (18.1)	8 (6.5)	0 (0.1)	0 (0.3)	9.5	0 (4)	2 (2)	28 (8.5)	52 (12.4)	17 (7.6)	1 (0.7)	0 (0.8)	9.3	0 (4)	2 (2)	28 (8.5)	52 (12.4)
LRmean	0 (3.1)	0 (0.2)	16 (5.7)	80 (23.1)	4 (2.7)	0 (0.1)	0 (1.2)	9.5	0 (3.2)	2 (1)	22 (6.9)	57 (13.6)	16 (5)	2 (5.4)	0 (1)	10.6	0 (3.2)	2 (1)	22 (6.9)	57 (13.6)
LRcat	0 (3)	0 (0.2)	8 (3.7)	85 (23.6)	7 (4.1)	0 (1.3)	0 (0.1)	10.1	0 (3.2)	1 (1.6)	18 (6)	58 (16.3)	20 (6.4)	1 (2)	0 (0.4)	9.8	0 (3.2)	1 (1.6)	18 (6)	58 (16.3)
MCRM	0 (3.1)	0 (3.1)	11 (6.6)	81 (17.7)	8 (5)	0 (0.4)	0 (0)	8.6	0 (4.1)	2 (4.4)	30 (10.9)	53 (11.8)	14 (4.1)	1 (0.6)	0 (0.1)	7.9	0 (4.1)	2 (4.4)	30 (10.9)	53 (11.8)
LRcat25	0 (3.1)	0 (1)	10 (17.1)	79 (12.7)	11 (1.2)	0 (0.8)	0 (0.1)	6.7	0 (4.4)	3 (4.8)	34 (15.9)	49 (8)	13 (1.8)	1 (0.8)	0 (0.3)	7.2	0 (4.4)	3 (4.8)	34 (15.9)	49 (8)
CRM	0 (3.1)	0 (0.7)	21 (5)	67 (9.5)	12 (5.3)	0 (0.1)	0 (0.3)	6.3	0 (4)	4 (1.6)	30 (5)	44 (6.8)	20 (5.4)	2 (0.5)	0 (0.7)	6.1	0 (4)	4 (1.6)	30 (5)	44 (6.8)
LRmean	0 (3.1)	0 (0.2)	18 (3.6)	74 (13.8)	8 (2)	0 (0.1)	0 (1.2)	6.3	0 (3.2)	4 (0.6)	23 (4.2)	49 (7.2)	20 (2.7)	3 (5.1)	1 (0.9)	7.3	0 (3.2)	4 (0.6)	23 (4.2)	49 (7.2)
LRcat	0 (3)	0 (0.2)	10 (2.6)	78 (14)	12 (2.9)	0 (1.3)	0 (0.1)	6.7	0 (3.2)	3 (1.4)	19 (3.7)	50 (9.8)	24 (3.7)	3 (1.8)	0 (0.4)	6.4	0 (3.2)	3 (1.4)	19 (3.7)	50 (9.8)
MCRM	0 (3.1)	0 (3.1)	14 (5)	69 (8.9)	17 (3.4)	0 (0.4)	0 (0)	5.2	0 (4.1)	5 (4)	32 (7)	43 (6.2)	17 (2.2)	2 (0.4)	0 (0.1)	4.8	0 (4.1)	5 (4)	32 (7)	43 (6.2)
LRcat25	0 (3.1)	0 (0.9)	14 (10.6)	70 (7.8)	15 (0.7)	0 (0.8)	0 (0.1)	4.4	0 (4.3)	5 (3.2)	35 (9.8)	43 (4.8)	14 (1)	2 (0.7)	1 (0.2)	4.6	0 (4.3)	5 (3.2)	35 (9.8)	43 (4.8)

Table VI. *Continued.*

Scenario S6								Scenario F6								
n=36								n=36								
0.01	0.01	0.03	0.04	0.11	0.27	0.52		0.03	0.04	0.07	0.11	0.18	0.27	0.39		
CRM	0 (3.1)	0 (0.1)	0 (0.3)	0 (2.1)	23 (11.6)	64 (11.9)	13 (6.9)	8.3	0 (3.3)	0 (0.4)	0 (1.2)	8 (4.7)	34 (11.4)	40 (7.5)	18 (7.5)	7.8
LRmean	0 (3.1)	0 (0)	0 (0.1)	2 (5.4)	29 (9.1)	60 (10.5)	10 (7.7)	8.1	0 (3.3)	0 (0.1)	0 (0.9)	7 (7)	30 (7.8)	38 (7.1)	24 (9.9)	8.2
LRcat	0 (3)	0 (0.1)	0 (0.1)	0 (3.7)	19 (6.9)	70 (16.2)	12 (5.9)	8.3	0 (3)	0 (0.4)	0 (0.6)	4 (6.1)	29 (7.7)	43 (10.7)	24 (7.4)	8
MCRM	0 (3.1)	0 (3)	0 (3.2)	1 (4)	23 (7.4)	62 (9.9)	15 (5.3)	6.6	0 (3.3)	0 (3.1)	1 (4.2)	12 (7.2)	35 (8.4)	36 (5.9)	16 (3.9)	5.9
LRcat25	0 (3.1)	0 (0.1)	0 (4)	1 (6)	19 (8.6)	66 (9.5)	14 (4.7)	6.3	0 (3.4)	0 (0.7)	1 (6.8)	12 (8.2)	31 (6.1)	34 (5.6)	22 (5.2)	6.2
n=24								n=24								
CRM	0 (3.1)	0 (0.1)	0 (0.3)	2 (2)	30 (8.3)	51 (5.1)	18 (5.1)	5.1	0 (3.3)	0 (0.4)	2 (1)	11 (3.4)	33 (7.5)	34 (3.2)	21 (5.1)	4.8
LRmean	0 (3.1)	0 (0)	0 (0.1)	4 (5.1)	38 (5.2)	44 (4.2)	13 (6.3)	5.2	0 (3.3)	0 (0.1)	1 (0.8)	11 (5.8)	32 (4.2)	29 (3.1)	27 (6.8)	5.1
LRcat	0 (3)	0 (0.1)	0 (0.1)	1 (3.6)	24 (4.2)	58 (8.8)	18 (4.1)	5.1	0 (3)	0 (0.4)	0 (0.6)	7 (5.3)	29 (4.3)	36 (6.1)	27 (4.3)	4.9
MCRM	0 (3.1)	0 (3)	0 (3.2)	2 (3.8)	28 (4.4)	40 (3.8)	30 (2.7)	3.2	0 (3.3)	0 (3.1)	2 (4)	17 (5.3)	35 (4.3)	23 (2.4)	23 (1.5)	3.1
LRcat25	0 (3.1)	0 (0.1)	0 (3.9)	2 (5.1)	26 (4)	50 (4.9)	21 (2.9)	3.6	0 (3.4)	0 (0.6)	3 (5.7)	17 (5.9)	29 (2.8)	24 (3)	26 (2.6)	3.5
Scenario F7								Scenario F1								
n=36								n=36								
0.01	0.03	0.04	0.07	0.11	0.18	0.27		0.27	0.39	0.52	0.65	0.76	0.84	0.90		
CRM	0 (3.1)	0 (0.1)	0 (0.3)	1 (1.9)	7 (6.9)	27 (6.1)	65 (17.5)	6.8	76 (24.1)	22 (7.5)	2 (2.8)	0 (0.4)	0 (1.2)	0 (0)	0 (0)	12.1
LRmean	0 (3.1)	0 (0)	0 (0.3)	1 (4.5)	6 (3.3)	23 (5.4)	70 (19.4)	7	72 (22.2)	26 (8)	2 (3.8)	0 (1.9)	0 (0)	0 (0)	0 (0.1)	12.4
LRcat	0 (3)	0 (0.1)	0 (0.2)	0 (4.1)	6 (3.1)	26 (9.3)	68 (16.1)	6.7	66 (19)	30 (10.6)	4 (4.2)	0 (2)	0 (0.1)	0 (0)	0 (0)	12.9
MCRM	0 (3.1)	0 (3)	0 (3.5)	3 (4.8)	13 (6.1)	29 (6.2)	56 (9.3)	4.9	76 (24.8)	22 (8.1)	2 (2.9)	0 (0.2)	0 (0)	0 (0)	0 (0)	11.5
LRcat25	0 (3.1)	0 (0.2)	0 (4.7)	3 (5.6)	9 (4.4)	24 (6.1)	63 (11.9)	5.4	70 (26.8)	29 (6.3)	1 (2.6)	0 (0.2)	0 (0)	0 (0)	0 (0)	11.2
n=24								n=24								
CRM	0 (3.1)	0 (0.1)	0 (0.3)	2 (1.7)	12 (5.7)	26 (3.2)	60 (9.9)	4.1	70 (15.4)	24 (4.6)	5 (2.4)	0 (0.4)	0 (1.2)	0 (0)	0 (0)	8.4
LRmean	0 (3.1)	0 (0)	0 (0.2)	2 (4.3)	11 (2.3)	23 (2.8)	64 (11.3)	4.2	64 (14.1)	29 (4.7)	7 (3.2)	0 (1.9)	0 (0)	0 (0)	0 (0.1)	8.6
LRcat	0 (3)	0 (0.1)	0 (0.2)	1 (4)	9 (2.2)	27 (6.2)	63 (8.2)	3.9	57 (11.7)	34 (6.7)	8 (3.4)	1 (2)	0 (0.1)	0 (0)	0 (0)	9
MCRM	0 (3.1)	0 (3)	0 (3.5)	6 (4.3)	19 (4.2)	24 (3)	51 (3)	2.4	69 (16)	26 (5.2)	4 (2.5)	0 (0.2)	0 (0)	0 (0)	0 (0)	7.8
LRcat25	0 (3.1)	0 (0.2)	0 (4.4)	6 (4.9)	17 (2.5)	21 (3.8)	57 (5.2)	2.9	65 (17.1)	30 (4.2)	5 (2.5)	0 (0.2)	0 (0)	0 (0)	0 (0)	7.7

scenarios S2, F2, S4 and F4. For scenarios S6 and F6, the marked differences seen under prior 1 disappear. A head-to-head comparison of the operating characteristics between the CRM and any of the three logistic methods show that the approaches are similar, i.e. for half of the eight scenarios the CRM is superior (inferior) to the two-parameter logistic model. For the majority of scenarios the differences are minor.

In terms of the number of DLTs per trial, the conservative methods (MCRM and LRcat25) show an average of 2–3 DLTs less than the other three methods. For example, if MTD is at dose level 6 (S6 and F6), after 36 patients the average number of DLTs is approximately 8 (CRM, LRmean and LRcat) and 6 (MCRM and LRcat25), respectively (Table VI). The number of DLTs increase to 12 and 10 if the MTD is at dose level 2, respectively. The conservative methods have a smaller number of DLTs because more patients are allocated to lower dose levels.

An important consideration is when to stop the trial. A possible rule would be to stop the trial as soon as the probability of targeted toxicity exceeds a certain threshold, e.g. 50 per cent. For this threshold the simulations show that it would be unlikely to stop the trial after 36 patients, except for the case where the MTD is at a low dose (F1, S2 and F2) for which the probability of stopping is 80 to 90 per cent (results not shown). For the other scenarios these probabilities are in the range 20 to 50 per cent. For 24 patients the stopping probabilities are considerably lower. Given that phase I cancer trials are typically of relatively small size (20–40 patients), this implies that information about the true MTD will often be quite uncertain. For a more detailed discussion of stopping rules, see [29, 40].

8. COMPUTATIONS

The proposed approach based on the two-parameter logistic model relies on full posterior distributions, which can be obtained via simulation. Markov chain Monte Carlo (MCMC) for the two-parameter logistic model can easily be done using standard Bayesian software WinBUGS [41] or R [42] with MCMCpack [43]. As an alternative sampling–importance resampling can be used [37, 44]. Appendix B lists an example of WinBUGS code.

9. DISCUSSION

We explored the critical aspects of finding the MTD in phase I cancer trials from a Bayesian model-based perspective. We propose a two-parameter model for making realistic inferential statements about the probabilities of a DLT at each dose level. After each patient cohort, information is derived from the posterior distribution of the model parameters, which serves as a basis for the clinical team to decide on a sensible dose for the next patient cohort. In our current work the decision making is informal in that we provide an informative summary of the posterior distribution, but the method can be formalized to a fully decision-analytic procedure based on loss functions representing different degrees of conservatism.

The proposed method was developed after we encountered problems when applying the CRM in one of our in-house trials. Explanations for the implausible dose recommendation were identified as (1) using point estimates instead of more informative posterior summaries; (2) using the less flexible one-parameter instead of a two-parameter model and (3) a problematic choice of the structural skeleton probabilities in the one-parameter model.

It has been repeatedly shown that the CRM has good operating characteristics, and we have successfully used it in a variety of our trials. However, there are situations where the drawbacks of coupling a presumably Bayesian procedure with a mis-specified model and suboptimal dose selection rules (which ignore uncertainties) become apparent. Moreover, there is no convincing reason for avoiding a more realistic two-parameter model. Operating characteristics are similar for the two models, and differences are related to the dose-escalation behavior as well as the underlying dose–DLT curve. However, there are two important points to make. Firstly, there is a clear trade-off between finding the MTD and the willingness to be conservative. The more conservative the escalation behavior, the longer it will take to be confident to recommend the correct MTD. Secondly, small phase I MTD trials (20–40 patients) run the risk of declaring a dose as the MTD, even though this conclusion may still be highly uncertain. This is particularly true if the dose–toxicity curve is rather flat, and the MTD is at a high dose.

The proposed approach requires full posterior distributions for the probabilities of a DLT at each dose, which calls for the modern simulation-based approach to Bayesian statistics. For the two-parameter model these computations can easily be implemented using standard MCMC software or even the non-MCMC sampling–importance resampling algorithm.

The Bayesian model-based approach to phase I oncology trials is a powerful tool that relies on a good statistical model, reasonable prior specifications and decisions derived from appropriate inferential summaries. There is no disadvantage in terms of operating characteristics, within-trial recommendations are sensible, and the approach is easily extendable to more complex settings.

APPENDIX A: PRIOR SPECIFICATIONS

A.1. Quantile-based non-informative (default) prior

The approach of Section 6 works independently of the amount of prior information. In this section we suggest a non-informative default prior based on a set of quantiles for the probabilities of toxicity that are derived from minimally informative unimodal Beta distributions.

Definition A1

Given a pre-specified quantile $q(p)$ of the prior distribution, $X \sim \text{Beta}(a, b)$ is minimally informative unimodal if (i) $\Pr\{X < q(p)\} = p$, (ii) $a \geq 1$ or $b \geq 1$ (or both) and (iii) $a + b$ minimal.

For a given prior quantile $q(p)$, the parameters and quantiles of a minimally informative unimodal Beta distribution can be easily obtained. If $q(p) > p$, $\text{Beta}(a, 1)$ is minimally informative unimodal if $a = \ln(p) / \ln\{q(p)\}$. Alternatively, if $q(p) < p$, $\text{Beta}(1, b)$ is minimally informative unimodal if $b = \ln(1 - p) / \ln\{1 - q(p)\}$.

The following steps will be taken to obtain the set of prior quantiles Q for the distributions of $\pi_\theta(d)$, $d \in \mathcal{D}$:

- (1) For the lowest dose d_1 the prior probability of exceeding a certain threshold q_1 will be p_1 . The following default values will be used: $\Pr\{\pi_\theta(d_1) > 0.6\} = 0.05$, i.e. for the lowest dose the probability of unacceptable toxicity will be set to 5 per cent.
- (2) For the highest dose d_J the prior probability of falling below a certain threshold q_J will be p_J . The following default values will be used: $\Pr\{\pi_\theta(d_J) \leq 0.2\} = 0.05$, i.e. for the highest dose the probability of under-dosing will be set to 5 per cent.

- (3) Assuming a minimally informative unimodal Beta distribution in (1) and (2) leads to prior medians for the probabilities of toxicity $\pi_\theta(d_1)$ and $\pi_\theta(d_J)$, μ_1 and μ_J say.
- (4) Prior medians μ_1, \dots, μ_J are assumed to be linear in log-dose on the logit scale.
- (5) For each dose d_j prior 95 per cent intervals will be derived using minimally informative unimodal Beta distributions with prior medians equal to μ_j .

For the two-parameter logistic model the above constructed quantiles Q are then compared with the quantiles Q' coming from a bivariate normal distribution. The minimization of $C(Q, Q')$ leads to the optimal parameter for the prior distribution $\eta = (\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$.

A.2. Priors based on information about α and β

An alternative to the quantile-based approach consists of incorporating information about α and β directly. The parameterization of the two-parameter model allows for the interpretation of the parameters and their prior setting, as follows:

- (i) $\log(\alpha)$ is the log-odds of a DLT at reference dose d^* (see model definition). The normal distribution of $\log(\alpha)$ would comprise prior information for this dose. For example, if one wanted to be uninformative about the log-odds of a DLT at the reference dose d^* , one could use two quantiles (e.g. the 2.5 and 97.5 per cent quantile) of a minimally informative unimodal Beta distribution as in Section A.1 and derive the corresponding normal distribution matching these quantiles on the logit scale. If one sets the reference dose d^* to the *a priori* anticipated MTD, the mean of $\log(\alpha)$ would follow from the targeted probability (e.g. 0.30), and only one additional quantile would be needed to obtain the prior standard deviation.
- (ii) For two doses d_i and d_j , the parameter β is essentially equal to the log-odds ratio of a DLT, i.e.

$$\beta = \frac{\text{logit}\{\pi(d_j)\} - \text{logit}\{\pi(d_i)\}}{\log(d_j/d_i)}$$

For example, the parameters of the normal distribution for $\log(\beta)$ can be obtained by specifying two quantiles for the change in the odds of a DLT if the dose is doubled. For an uninformative prior on β one should allow for flat as well as steep dose-toxicity curves, i.e. for small and large changes between doses d_i and d_j in the odds of a DLT.

APPENDIX B: WinBUGS CODE

```
model{
  # prior covariance matrix
  cova[1,1] <- Prior[3]*Prior[3]
  cova[2,2] <- Prior[4]*Prior[4]
  cova[1,2] <- Prior[3]*Prior[4]*Prior[5]
  cova[2,1] <- cova[1,2]
  prec[1:2,1:2] <- inverse(cova[,])
  log.alphabeta[1:2] ~ dnmnorm(Prior[1:2],prec[1:2,1:2])
  alphabeta[1] <- exp(log.alphabeta[1])
  alphabeta[2] <- exp(log.alphabeta[2])
  # sampling model
  for (j in 1:Ncohorts){
```

```

logit(Pr.Tox1[j])<- log.alphabeta[1]+alphabeta[2]*log(DosesAdm[j]/DoseRef)
Ntox[j] ~ dbin(Pr.Tox1[j],Npat[j])
}
# for each dose: probabilities of toxicity, category probabilities, risks
for (i in 1:Ndoses) {
  lin[i]<-log.alphabeta[1]+alphabeta[2]*log(Doses[i]/DoseRef)
  logit(Pr.Tox[i]) <- lin[i]
  Pr.Cat[i,1] <- step(Pint[1]-Pr.Tox[i])
  Pr.Cat[i,2] <- step(Pint[2]-Pr.Tox[i])-step(Pint[1]-Pr.Tox[i])
  Pr.Cat[i,3] <- step(Pint[3]-Pr.Tox[i])-step(Pint[2]-Pr.Tox[i])
  Pr.Cat[i,4] <- step(1-Pr.Tox[i])-step(Pint[3]-Pr.Tox[i])
  Risk1[i] <- inprod(Pr.Cat[i,],LossFunction1[1:4])
  Risk2[i] <- inprod(Pr.Cat[i,],LossFunction2[1:4])
  Risk3[i] <- inprod(Pr.Cat[i,],LossFunction3[1:4])
}
} # data
list( Ncohorts=5, DoseRef=250,
DosesAdm=c(1,2.5,5,10,25),Npat=c(3,4,5,4,2),
Ntox=c(0,0,0,0,2), Ndoses=10,Doses=c(1,2.5,5,10,15,20,25,30,40,50),
Prior=c(2.15,0.52,0.84,0.78,0.20),Pint=c(0.2,0.35,0.6),
LossFunction1=c(1,0,1,1),LossFunction2=c(1,0,1,2),LossFunction3=c(1,0,2,4))

```

ACKNOWLEDGEMENTS

The authors thank Amy Racine, Jeff Eisele, Stuart Bailey and Mauro Gasparini for their helpful comments and suggestions. The authors would also like to thank the referees for their careful reading and suggestions.

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. Dixon W, Mood A. A method for obtaining and analyzing sensitivity data. *Journal of the American Statistical Association* 1948; **43**:109–126.
4. Rosenberger WF, Haines LM. Competing designs for phase I clinical trials: a review. *Statistics in Medicine* 2002; **21**:2757–2770.
5. Ivanova A, Montazer-Haghighi A, Mohanty SG, Durham SD. Improved up-and-down designs for phase I trials. *Statistics in Medicine* 2003; **22**:69–82.
6. Gasparini M, Eisele J. A curve-free method for phase I clinical trials. *Biometrics* 2000; **57**:609–615.
7. O'Quigley J. Curve-free model-based continual reassessment method designs. *Biometrics* 2002; **58**:245–249.
8. Leung DH, Wang YG. Isotonic designs for phase I trials. *Controlled Clinical Trials* 2001; **22**:126–138.
9. Faries D. The continual reassessment method in cancer phase I clinical trials: a simulation study. *Journal of Biopharmaceutical Statistics* 1994; **4**(2):147–164.
10. O'Quigley J, Shen L. Continual reassessment method: a likelihood approach. *Biometrics* 1996; **52**:673–684.
11. Babb JS, Rogatko A, Zacks S. Cancer phase I clinical trials: efficient dose escalation with overdose control. *Statistics in Medicine* 1998; **17**:1103–1120.
12. Zhou Y, Whitehead J. Practical implementation of Bayesian dose-escalation procedures. *Drug Information Journal* 2003; **37**:45–59.
13. Zhou Y, Lucini M. Gaining acceptability for the Bayesian decision-theoretic approach in dose-escalation studies. *Pharmaceutical Statistics* 2005; **4**:161–171.
14. Babb JS, Rogatko A. Patient specific dosing in a cancer phase I clinical trial. *Statistics in Medicine* 2001; **20**:2079–2090.
15. Rogatko A, Babb JS, Tighiouart M, Khuri FR, Hudes G. New paradigm in dose-finding trials: patient-specific dosing and beyond phase I. *Clinical Cancer Research* 2005; **11**:15.

16. Loke Y-C, Tan S-B, Cai Y, Machin D. A Bayesian dose finding design for dual endpoint phase I trials. *Statistics in Medicine* 2006; **25**:3–22.
17. Cheung YK, Chappell R. Sequential designs for phase I clinical trials with late-onset toxicities. *Biometrics* 2000; **56**:1177–1182.
18. Braun TM, Zheng Y, Thall PF. Determining a maximum-tolerated schedule of a cytotoxic agent. *Biometrics* 2005; **61**:335–343.
19. Braun TM. Generalizing the TITE-CRM to adapt for early- and late-onset toxicities. *Statistics in Medicine* 2006; **25**(12):2071–2083.
20. Legedza ATR, Ibrahim JG. Longitudinal design for phase I clinical trials using the continual reassessment method. *Controlled Clinical Trials* 2000; **21**:574–588.
21. Piantadosi S, Liu G. Improved designs for dose escalation studies using pharmacokinetic measurements. *Statistics in Medicine* 1996; **15**:1605–1618.
22. Legedza ATR, Ibrahim JG. Heterogeneity in phase I clinical trials: prior elicitation and computation using the continual reassessment method. *Statistics in Medicine* 2001; **20**:867–882.
23. Whitehead J. Heterogeneity in phase I clinical trials: prior elicitation and computation using the continual reassessment method (letter to the editor). *Statistics in Medicine* 2002; **21**:1172.
24. Kramar A, Lebecq A, Candalh E. Continual reassessment methods in phase I trials of the combination of two drugs in oncology. *Statistics in Medicine* 1999; **18**:1849–1864.
25. Thall P, Millikan R, Mueller P, Lee S-J. Dose-finding with two agents in phase I oncology trials. *Biometrics* 2003; **59**:487–496.
26. Whitehead J, Zhou Y, Mander A, Ritchie S, Sabin A, Wright A. An evaluation of Bayesian designs for dose-escalation studies in healthy volunteers. *Statistics in Medicine* 2006; **25**(3):433–445.
27. Thall PF, Lee JJ, Tseng C-H, Estey EH. Accrual strategies for phase I trials with delayed patient outcome. *Statistics in Medicine* 1999; **18**:1155–1169.
28. Chevret S. The continual reassessment method in cancer phase I clinical trials: a simulation study. *Statistics in Medicine* 1993; **12**:1093–1108.
29. Zohar S, Chevret S. The continual reassessment method: comparison of Bayesian stopping rules for dose ranging studies. *Statistics in Medicine* 2001; **20**:2827–2843.
30. Ishizuka N, Ohashi Y. The continual reassessment method and its applications: a Bayesian methodology for phase I cancer clinical trials. *Statistics in Medicine* 2001; **20**:2661–2681.
31. Whitehead J, Brunier H. Bayesian decision procedures for dose determining experiments. *Statistics in Medicine* 1995; **14**:885–893.
32. Patterson S, Francis S, Ireson M, Webber D, Whitehead J. A novel Bayesian decision procedure for early-phase dose-finding studies. *Journal of Biopharmaceutical Statistics* 1999; **9**(4):583–597.
33. Leung DH, Wang YG. An extension of the continual reassessment method using decision theory. *Statistics in Medicine* 2002; **21**:51–63.
34. Haines LM, Perevozskaya I, Rosenberger WF. Bayesian optimal designs for phase I clinical trials. *Biometrics* 2003; **59**:591–600.
35. Spiegelhalter DJ, Abrams KR, Myles JP. *Bayesian Approaches to Clinical Trials and Health-care Evaluation*. Wiley: U.K., 2004.
36. O'Quigley J. Retrospective analysis of sequential dose-finding designs. *Biometrics* 2005; **61**:749–756.
37. Robert CP, Casella G. *Monte Carlo Statistical Methods*. Springer: New York, 2004.
38. Neuenschwander B, Gsponer T, Branson M. A Bayesian approach to phase I cancer trials. *Technical Report 2006-05-11*, Novartis Pharma AG, Basel, Switzerland, 2006.
39. Neuenschwander B, Capkun-Niggli G, Branson M, Spiegelhalter D. *Using Historical Controls in Clinical Trials*, in preparation.
40. Heyd JM, Carlin BP. Adaptive design improvements in the continual reassessment method for phase I studies. *Statistics in Medicine* 1999; **18**:1307–1321.
41. Spiegelhalter D, Thomas A, Best N. *WinBUGS 1.4.1*. MRC Biostatistics Unit: Cambridge, U.K., 2005.
42. R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria, 2004. ISBN 3-900051-07-0.
43. Martin AD, Quinn KM. *MCMCpack: Markov Chain Monte Carlo (MCMC) Package*. R Package Version 0.6-4, 2005.
44. Smith AFM, Gelfand AE. Bayesian statistics without tears: a sampling–resampling perspective. *The American Statistician* 1992; **46**(2):44–88.