A Bayesian dose-finding design for drug combination clinical trials based on the logistic model

Marie-Karelle Riviere, Ying Yuan, Frédéric Dubois, Sarah Zohar

A Bayesian dose-finding design for drug combination clinical trials based on the logistic model

Marie-Karelle Riviere[(a,b)], Ying Yuan[(c)], Frédéric Dubois[(b)] and Sarah Zohar[(a)]

(a) INSERM, UMRS 1138, Equipe 22, Centre de Recherche des Cordeliers, Université Paris 5, Université Paris 6, Paris, France

(b) IRIS (Institut de Recherches Internationales Servier), Suresnes, France

(c) Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, U.S.A.

Abstract

In early phase dose-finding cancer studies, the objective is to determine the maximum tolerated dose, defined as the highest dose with an acceptable dose-limiting toxicity rate. Finding this dose for drug-combination trials is complicated due to drug-drug interactions, and many trial designs have been proposed to address this issue. These designs rely on complicated statistical models that typically are not familiar to clinicians, and are rarely used in practice. The aim of this paper is to propose a Bayesian dose-finding design for drug combination trials based on standard logistic regression. Under the proposed design, we continuously update the posterior estimates of the model parameters to make the decisions of dose assignment and early stopping. Simulation studies show that the proposed design is competitive and outperforms some existing designs. We also extend our design to handle delayed toxicities.
Keywords: Bayesian inference; Dose-finding; Drug combination; Oncology; Phase I trial.

1 Introduction

For oncologists, the objective of phase I dose-finding studies is to determine the maximum tolerated dose (MTD), defined as the highest dose with a relatively acceptable dose-limiting toxicity (DLT) [1, 2]. DLT is usually defined as a toxicity of grade 3 or higher according to the U.S. National Cancer Institute toxicity criteria [3]. In practice, patients included in phase I clinical cancer trials have already been heavily pre-treated and in many cases no alternative therapeutic options are available to them. For cytotoxic anti-cancer drugs, a dose-toxicity effect is assumed whereby the higher the dose, the greater the risk of DLT and the greater the efficacy.

Dose finding for drug combination trials is more difficult than that for conventional single-agent trials due to complicated drug-drug interactions. Moreover, when combining several agents, the order of the toxicity probabilities is not fully known. Should investigators wish to gradually increase the acceptable level of toxicity during the trial, the appropriate order in which the doses for the various drugs in the combination should be increased would be of great interest. For instance, when combining two cytotoxic agents, it remains difficult to decide how to escalate or de-escalate the dose combination, even when a partial ordering is known [4, 5].

Recently, many phase I dose-finding designs have been proposed for drug combination trials. Thall et al. proposed a Bayesian dose-finding method based on a six-parameter model [6]. Wang and Ivanova developed a 3-parameter model-based method in which the parameters are estimated using Bayesian inference [7]. Mandrekar et al. proposed an approach incorporating the toxicity and efficacy of each agent into the identification of an optimal dosing region for the combination by using a continuation ratio model to separate each agent's toxicity and efficacy curves [8, 9]. Yin and Yuan developed a Bayesian adaptive design based on latent 2x2 tables in which the combination's toxicity probabilities in the two-dimensional space are estimated using a Gumbel-type model [10]. Yin and Yuan extended their method by changing to a copula-type model to simulate the effect of two or more drugs in combination [11]. Bailey et al. introduced a second agent as a covariate in a logistic model [12]. Wages et al. considered an approach based on the continual reassessment method and taking into account different orderings with partial order between combinations. In this case, the MTD is estimated for the order associated with the highest model-selection criterion [5]. Most of these existing designs rely on complicated statistical models that typically are not familiar to clinicians, which hinder their acceptance and application in practice. In addition, the performance of these designs seems comparable and there is no consensus which design should be used [13]. As a result, many of the dose-finding clinical trials conducted to evaluate drug combinations still use the conventional "3+3" approach, which was developed for single agents and was shown to be inefficient in terms of dose identification [14, 15, 16, 17, 18].

In this paper, we propose an Bayesian dose-finding design for drug combination trials based on standard logistic regression. Under the Bayesian paradigm, data monitoring, early stopping and dose assignment occur continuously throughout the trial by updating the posterior estimates of the model parameters. Simulation studies show that in general the proposed design provide better performance than some existing designs. To facilitate the use of the proposed design by clinicians, R package will be developed for implementing the new design.

This manuscript is laid out as follows: in Section 2, we propose the simple statistical method for modeling the toxicity probabilities of the drug combination under evaluation. Moreover, we present the likelihood function and the prior specifications for the unknown parameters. In this section, we also describe our allocation and dose-finding method, as well as propose stopping rules.

We conduct extensive simulation studies to examine the operating characteristics of our design in Sections 3 and 4, and conclude with a discussion in Section 5.

2 Methods

2.1 Statistical method for combination evaluation (LOGISTIC)

Dose-combination model

Let there be a two-drug combination used in a phase I dose-finding clinical trial for which the dose-toxicity relationship is monotonic and increases with the dose levels. Let (j, k) denote the dose level of a combination in which j refers to Agent 1 $(j = 1, \ldots, J)$, and k refers to Agent 2 $(k = 1, \ldots, K)$. Y_i is a Bernoulli random variable, denoting the toxicity that is equal to 1 if DLT occurs in patient i and 0 otherwise $(i = 1, \ldots, N)$. We assume that $n_{j,k}$ patients are allocated at combination (j, k) and that a total of $t_{j,k}$ DLTs are observed for that combination. We define $\pi_{j,k}$ as the toxicity probabilities of combination (j, k), θ as the target probability of toxicity, and $p_1, \ldots, p_J$ and $q_1, \ldots, q_K$ as the respective prior toxicity probabilities of Agent 1 and Agent 2 taken alone. For simplicity purposes, we refer to the selected combination as the MTD in order to maintain the same designation as in single-agent trials. In this manuscript, we focus on finding one MTD at the end of the trial. We note that in some cases, it is of interest to find multiple MTDs that can be further tested in phase II trials, see Yuan and Yin [19] and Ivanova and Wang [20] for related designs.

Let u_j and v_k denote the "effective" or standardized doses ascribed to the jth level of Agent 1 and k level of Agent 2, respectively. We model the drug combination-toxicity relationship using a 4-parameter logistic model, as follows:

$$\text{logit}(\pi_{j,k}) = \beta_0 + \beta_1 u_j + \beta_2 v_k + \beta_3 u_j v_k, \tag{1}$$

where β_0, β_1, β_2 and β_3 are unknown parameters that represent the toxicity effect of Agent 1 (β_1), that of Agent 2 (β_2) and that of the interaction between the two agents (β_3). These parameters are defined such that $\beta_1 > 0$ and $\beta_2 > 0$, ensuring that the toxicity probability is increasing with the increasing dose level of each agent alone, $\forall k, \beta_1 + \beta_3 v_k > 0$ and $\forall j, \beta_2 + \beta_3 u_j > 0$, ensuring that the toxicity probability is increasing with the increasing dose levels of both agents together, and intercept $-\infty < \beta_0 < \infty$. The standardized dose of two agents are defined as $u_j = \log\left(\frac{p_j}{1-p_j}\right)$ and $v_k = \log\left(\frac{q_k}{1-q_k}\right)$, where p_j and q_k are the prior estimates of the toxicity probabilities of the jth dose level of Agent 1 and the kth dose level of Agent 2, respectively, when they are administered individually *as a single agent*. Before two agents can be combined, each of them typically have been thoroughly investigated individually. Therefore, there is often rich prior information on p_j's and q_k's and their values can be readily elicited from physicians. Using the prior information to define standardized dose has been widely used in dose-finding trial designs, and the most well known example perhaps is the skeleton of the continuous reassessment method (CRM) [21] with a logistic model. Research has shown that this approach improves the estimation stability and trial performance [2, 22]. Our definition of u_j and v_k can be viewed as an extension of the skeleton of the CRM (with a logistic model) to drug-combination trials.

Likelihood and posterior inference

Under the proposed model, the likelihood is simply a product of the Bernoulli density, given by

$$L(\beta_0, \beta_1, \beta_2, \beta_3 | \text{data}) \propto \prod_{j=1}^{J} \prod_{k=1}^{K} \pi_{j,k}^{t_{j,k}} (1 - \pi_{j,k})^{n_{j,k} - t_{j,k}}.$$

We assume the prior distributions of the model parameters are independent. For β_0 and β_3, we assign a vague normal prior $N(0, 10)$ centered at 0 to indicate that *a priori* we do not favor either positive or negative values for these parameters and let the observed data speak for themselves through posteriors. For β_1 and β_2, we assume an informative prior distribution $\text{Exp}(1)$ centered at 1, as those parameters should not be too far from 1. Then, the joint posterior distribution of parameters $\beta_0, \beta_1, \beta_2$ and β_3 is given by

$$f(\beta_0, \beta_1, \beta_2, \beta_3 | \text{data}) \propto L(\beta_0, \beta_1, \beta_2, \beta_3 | \text{data}) f(\beta_0) f(\beta_1) f(\beta_2) f(\beta_3). \tag{2}$$

We sample this posterior distribution using Gibbs sampler, which sequentially draws each of the parameters from their full conditional distributions (see Appendix). These full conditional distributions do not have closed forms, and we use the Adaptive Rejection Metropolis Sampling (ARMS) method [23] to sample them. In order to impose the constrain that $\forall k, \beta_1 + \beta_3 v_k > 0$ and $\forall j, \beta_2 + \beta_3 u_j > 0$, at each iteration of Gibbs sampling, if sampled β_1, β_2 and β_3 fail to satisfy the constraint, we re-sample β_1, β_2 and β_3. Let $(\beta_0^{(\ell)}, \beta_1^{(\ell)}, \beta_2^{(\ell)}, \beta_3^{(\ell)})_{\ell=1,\dots,L}$ denote the L posterior samples obtained from Gibbs sampler, the posterior toxicity probabilities can be estimated using Monte Carlo by:

$$\tilde{\pi}_{j,k} = \frac{1}{L} \sum_{\ell=1}^{L} \frac{\exp(\beta_0^{(\ell)} + \beta_1^{(\ell)} u_j + \beta_2^{(\ell)} v_k + \beta_3^{(\ell)} u_j v_k)}{1 + \exp(\beta_0^{(\ell)} + \beta_1^{(\ell)} u_j + \beta_2^{(\ell)} v_k + \beta_3^{(\ell)} u_j v_k)}.$$

Dose finding algorithm and determination of the MTD

During the trial conduct, we use the dose-finding algorithm proposed by Yin and Yuan [11, 10] to determine dose escalation and deescaltion, and propose a different criterion for MTD selection at the end of the trial. In our design, similar to Yin and Yuan [11], we restrict dose escalation and de-escalation one level at a time (i.e, we do not allow dose escalate or de-escalate along the diagonal) based on the practical consideration that physicians are conservative and typically do not allow two agent to escalate at the same time for patient safety. Nevertheless, we note that Sweeting and Mander [24] showed that diagonal escalation strategy may be more efficient in reaching the target toxicity level quicker with fewer patients treated at sub-optimal doses and have a higher percentage of correct selection at the end of the trial. We take this strategy in our start-up phase described later.

Let c_e be the probability threshold for dose escalation and c_d the probability threshold for dose de-escalation. We require $c_e + c_d > 1$ to avoid that the decisions of dose escalation and deescalation occur at the same time. Our dose-finding algorithm can be described as follows:

- If the current combination is (j, k) and $P(\pi_{j,k} < \theta | \text{data}) > c_e$,

4

we escalate the combination dose level to an adjacent combination dose level $\{(j+1,k),$ $(j,k+1), (j+1,k-1), (j-1,k+1)\}$ that has a toxicity probability that is higher than the current value and closest to θ. If the current combination is the highest of the combination space, (J, K), we retain the same combination dose level for the next cohort.

- If the current dose combination is (j, k) and $P(\pi_{j,k} > \theta|\text{data}) > c_d$,

  we de-escalate the combination dose level to an adjacent combination dose level $\{(j-1,k),$ $(j,k-1), (j+1,k-1), (j-1,k+1)\}$ that has a toxicity probability that is lower than the current value and closest to θ. If the current combination is the lowest of the combination space, $(1, 1)$, we retain the same combination dose level for the next cohort.

- If the current combination is (j, k) and $P(\pi_{j,k} < \theta|\text{data}) \leq c_e$ and $P(\pi_{j,k} > \theta|\text{data}) \leq c_d$,

  we treat the next cohort of patients at the current combination dose level.

Once we reach the maximum sample size, we select the MTD as the combination associated with the highest posterior probability, $P(\pi_{j,k} \in [\theta - \delta; \theta + \delta])$, and have been used to treat at least one cohort of patients. If, for example, the target toxicity θ is 0.3, and the length around the targeted interval is $\delta = 0.1$, which gives a targeted interval defined as $[0.2; 0.4]$, then, as illustrated in Figure 5, at the end of the trial we can obtain the posterior densities of the toxicity probability for each combination. Shown in light gray in the figure is the area under the curve (AUC) for a toxicity probability lower than 0.2, which is equal to the probability of under-dosing. Shown in medium gray is the AUC for a toxicity probability between 0.2 and 0.4, which is the probability of being in the targeted interval. Shown in dark gray is the AUC for a toxicity probability greater than 0.4, which is equal to the probability of overdosing. For each combination, given the probabilities of being in the targeted interval, the combination already administered to at least one cohort at the end of the trial and corresponding to the highest probability is selected as the MTD. For example, dose combination $(2, 3)$, as illustrated in Figure 2.

The probability thresholds c_e and c_d are critical for the performance of the design as they control the dose escalation and deescalation. The values of c_e and c_d should be carefully calibrated through simulation to ensure good operating characteristics of the design. In practice, this can be done as follows: first define a set of representative dose-toxicity scenarios that may be encountered in the trial, and then conduct simulation under different values of c_e and c_d to evaluate the performance of the design. This is a trial-and-error process and may involve repeatedly tuning the values of c_e and c_d based on the simulation results. The goal is to find the values of c_e and c_d that yield good overall performance across different scenarios (e.g., the percentage of correct selection of the MTD, the number of patients exposed to over-toxic combinations or under-toxic combinations). Such a calibration based approach has been widely used in clinical trial designs [25, 6, 11, 10].

Due to the limited availability of information at the beginning of the trial, the posterior estimates of the toxicity probabilities may not be reliable. Therefore, as suggested by other authors [10, 16, 26, 27, 20], we implement an algorithm-based start-up phase in order to gather enough information to estimate the $\pi_{j,k}$. Our start-up phase shares the spirit of accelerated titration design [28] and can be described as follows: treat the first cohort of patients at the lowest dose combination $(1, 1)$; if no toxicity is observed, escalate the dose along the diagonal and treat the second cohort of patients at $(2, 2)$; we continue this dose escalation along the diagonal until the observation of a toxicity or at least one agent is at its maximum. If one of the agents is already at its maximum dose level and

still no toxicity is observed, we increase the dose of the other agent until both agents reach their maximum dose level. At any time of the start-phase, once the first toxicity event is observed, the start-up phase is completed and the design switches to the model-based dose-finding algorithm as described above.

Stopping rules

Our aim is to propose a design that, should the investigators wish, will stop the trial when all combinations are estimated to be unacceptable in terms of toxicity. If the current combination is $(1, 1)$, at least two cohorts have been included and $P\left(\pi_{j,k} < \theta\right) \geq 0.975$, then the trial is stopped.

2.2 Some existing designs for combination studies

BCOPULA and BGUMBEL methods

Yin and Yuan proposed two Bayesian methods that use copula regression for combinations. The first method [11] uses a Clayton-copula regression in order to express the joint probability of combination (j, k) with the marginal true probabilities of toxicity (p_j^α, q_k^β) :

$$\pi_{j,k} = 1 - \left((1 - p_j^\alpha)^{-\gamma} + (1 - q_k^\beta)^{-\gamma} - 1\right)^{-\frac{1}{\gamma}}.$$

The second design [10] uses a Gumbel model:

$$\pi_{j,k} = 1 - (1 - p_j^\alpha)(1 - q_k^\beta)\left[1 + p_j^\alpha q_k^\beta \frac{e^\gamma - 1}{e^\gamma + 1}\right],$$

where $\gamma, \alpha, \beta > 0$ are unknown parameters. The parameter γ characterizes the drug interactive effect, and α and β the uncertainty of the initial guesses. The combination allocation algorithm is the same as that presented in Section 2.1. The final MTD is the combination with a toxicity probability closest to the target among the combinations already administered in the trial.

I2D method

Wang and Ivanova proposed the I2D method [7], which is a two-dimensional model-based design, defined as follows [7]:

$$\pi_{j,k} = 1 - (1 - p_j)^\alpha (1 - q_k)^\beta \exp\left(-\gamma \log(1 - p_j) \log(1 - q_k)\right),$$

where $\alpha > 0$, $\beta > 0$, the interaction term $\gamma > 0$, and p_j, q_k are the working models for Agent 1 and Agent 2, respectively.

The combination that will be allocated to the next cohort is the combination closest to the target that belongs to $\{(j + 1, k), (j, k + 1), (j - 1, k + 1), (j - 1, k), (j, k - 1), (j + 1, k - 1), (j, k)\}$. The final MTD is defined as the combination with a toxicity probability closest to the target among the combinations already administered in the trial.

LOGODDS

Gasparini, Bailey and Neuenschwander [29] have proposed an alternative model, $\pi_{j,k}$, to the copula proposed by Yin and Yuan [11] containing an interaction parameter γ which explicitly quantifies departure from no interaction on the log-odds scale:

$$\gamma = \text{logodds}(\pi_{j,k}) - \text{logodds}(\pi_{j,k}^{\perp})$$

where $\pi_{j,k}^{\perp}$ is the no-interaction model defined as $\pi_{j,k}^{\perp} = p_j + q_k - p_j q_k$. By backsolving this equation, an explicit expression for the probability of toxicity is obtained. A normal prior centered in 0 and with variance of 100 was chosen for the interaction parameter. The rest of the dose allocation process, estimation and MTD determination was the same as our proposed design in order to compare our method involving a simple interaction logistic model with other logistic models.

TMML

Thall et al. [6] proposed a six-parameter model defined as follows:

$$\pi_{j,k} = \frac{\alpha_1 x_1^{\beta_1} + \alpha_2 x_2^{\beta_2} + \alpha_3 (x_1^{\beta_1} x_2^{\beta_2})^{\beta_3}}{1 + \alpha_1 x_1^{\beta_1} + \alpha_2 x_2^{\beta_2} + \alpha_3 (x_1^{\beta_1} x_2^{\beta_2})^{\beta_3}}$$

where $\alpha_1, \beta_1, \alpha_2, \beta_2, \alpha_3, \beta_3$ are unknown positive parameters. Following Thall [6], a prior gamma distribution with parameters $(0.3, 0.3)$ was chosen for α_1, α_2 and α_3, and a gamma distribution with parameters $(0.003, 0.03)$ was chosen for β_1, β_2 and β_3. The rest of the dose allocation process, estimation and MTD determination was the same as our proposed design in order to compare our method involving a simple interaction logistic model with other logistic models.

One-dimensional CRM

In practice, one-dimensional CRM sometimes is used to conduct dose-combination trials [13]. Under this method, we first preselect a subset of combinations, for which the toxicity probability order is known, and then apply the standard CRM to find the MTD. The drawback of such an approach is that we only investigate a subset of the whole two-dimensional dose space and may miss the true MTD. Following [13], we chose the subset of combinations as the combinations located at anti-diagonal path: $D_{1,1} \longrightarrow D_{1,2} \longrightarrow D_{2,2} \longrightarrow D_{3,2} \longrightarrow D_{4,2} \longrightarrow D_{5,2} \longrightarrow D_{5,3}$.

3 Simulations

We simulated 2000 independent replications of phase I trials that evaluate two agents in drug combinations, with five dose levels for Agent 1 and three for Agent 2, giving 15 possible combinations. We simulated 14 scenarios to represent the possible true underlying combination toxicities (Table 1). We studied several locations of the MTD in the combination space, as well as a number of correct combinations (or MTDs) that could shift from 0 to 3. We fixed the toxicity target at 0.3 and used an overall sample size of 60. In this paper, we compare the performance of our method with those of five other designs, including the BCOPULA [11] and BGUMBEL [10] methods proposed

by Yin and Yuan, and the I2D method proposed by Wang and Ivanova [7], and with two other logistic models, LOGODDS proposed by Gasparini et al. [29] and TMML proposed by Thall [6].We also compared the performance of this multidimensional designs with a one-dimensional CRM, CRM anti-diag, as described previously. The design parameters of all the designs (e.g., working model) have been calibrated via simulation before used for the comparison.

To ensure comparability, we chose a cohort size of 3 for all methods and did not use stopping rules when comparing all the methods. The start-up phase was implemented in all simulations. Due to ethical concerns, we started each trial at the lowest dose combination $(1, 1)$. We selected these features in order to employ typical trial set-ups in our simulation study. For Agent 1, we specified the marginal initial guesses of toxicities, p_j, as $(0.12, 0.2, 0.3, 0.4, 0.5)$, and for Agent 2, we specified q_k as $(0.2, 0.3, 0.4)$. The same working models were used for BCOPULA, BGUMBEL, LOGODDS and I2D. We set the length around the targeted interval, δ, at 0.1. Based on a sensitivity analysis (data not shown), we set the probability thresholds c_e and c_d at 0.85 and 0.45. We recorded 5000 posterior samples of the model parameters after 2000 burn-in iterations to make inference. After comparing the proposed design to the other designs, we also investigated the performance of our design with the stopping rule that we introduced above. For the one-dimensional CRM, we used the "dfcrm" R package with restrictions to avoid skipping doses. The working model was generated using the "getprior" function with an indifference interval $\delta = 0.05$, a initial guessed MTD at dose level 4 for a trial with 7 doses, and the same settings as for multidimensional designs were used.

4   Results

For each scenario, we present the correct MTD selection rate, or percentages of correct selection (PCS), in Table 2. In general, the proposed method for combination evaluation (LOGISTIC) performed better than the other existing model-based designs. Indeed, except in scenarios 4 and 5, in which the PCS were already very high (86.7% and 80.4% respectively), those in other scenarios were either the best or not less than 2% of those of the other designs. Moreover, the PCS were greater than 55% in all scenarios and higher than 60% in 12 scenarios out of 14.

Scenarios 1, 3 and 9 included three possible MTDs that were located on one diagonal of the combination space; combinations $(2, 3)$, $(3, 2)$, $(4, 1)$ in scenario 1, combinations $(3, 3)$, $(4, 2)$, $(5, 1)$ in scenario 3 and combinations $(1, 3)$, $(2, 2)$, $(3, 1)$ in scenario 9 (Table 1). For these scenarios, all model-based designs gave high PCS, whereas LOGISTIC seemed to perform better than the other methods. In addition, for the LOGISTIC, the distribution of allocated dose levels around the MTDs (Table 3) seemed to be spread unevenly between possible MTDs.

In scenarios 2 and 10, in which two possible MTDs were located on the same diagonal at the lower end or at the higher end of the combination space, high PCS were observed for the LOGISTIC ($> 75\%$). When the correct combination was in the lower $(1, 1)$ extremity of the combination space, as in scenario 4, the TMML model and then I2D method performed best (92.1% and 89.7% respectively), but the PCS for the LOGISTIC was close to this value (86.7%). By contrast, when the correct combination was in the higher $(5, 3)$ extremity of the combination space, as in scenario 5, the LOGODDS model gave the highest PCS with 90.0%.

In scenarios 6, 7, 11, 12, 13 and 14, when the true MTDs were not located on the same diagonal but randomly located in the combination space, the proposed method performed best with the PCS

that was consistently higher than 56%. In contrast, the PCS of other methods can be as low as 30%. For example, under scenario 6, the PCS of BCOPULA, BGUMBEL, I2D, and LOGODDS were all lower than 47%; and under scenario 12, the PCS of TMML was only 47.9%.

Finally, when the true MTD was unique and located in the middle of the combination space, as in scenario 8, the PCS decreased. The TMML gave the best PCS (69.7%), nevertheless the LOGISTIC was the second most efficient design and gave good results with PCS of 56.9% for that scenario (Table 2). The PCS of the LOGODDS was merely 12.1%.

Under our simulation setting, the one-dimensional CRM (based on an anti-diagonal subset of combinations) yielded higher PCS than the multidimensional designs. However, this result should not be generalized. The reason that the one-dimensional CRM performed very well is simply because the pre-selected subset of combinations happened to include the true MTD. As the CRM focused on only a subset of doses, it had more patients per dose to find the target doses than the multidimensional designs, given the same total number of patients. It is easy to see that if the true MTD(s) is (are) not contained in the pre-selected subset of ordered combinations, then the one-dimensional CRM will perform badly and can never select the true MTD.

The mean number of DLTs observed throughout the trial in each scenario when using the LOGISTIC was similar to that observed when using the BCOPULA, BGUMBEL, and I2D methods (Table 2). Over all the scenarios, the mean number of DLTs observed during the trial was 15.4, which is lower than 18, the expected number of DLTs for 60 patients with a 0.30 toxicity probability.

According to Table 2, the LOGISTIC tended to treat as many patients at the MTD combination as the other methods. The lowest percentage of patients treated at the MTD combination was 24.3% in scenario 8, and the highest was 78.0% in scenario 4. In general, the true MTD combination was assigned to at least $1/3$ of the patients, and the mean overall percentage of patients treated at the MTD combination was 42.6%.

After employing the stopping rule for all scenarios when using the LOGISTIC, we added a scenario in which all the dose combinations were unacceptably toxic (Table 4). The addition of the stopping rule for unacceptable toxicity resulted in PCS that were similar to those presented above (Table 5), except in scenario 4 in which the first dose combination, $(1, 1)$, was the MTD. For this scenario, the PCS decreased by 16.9%, although it was nevertheless high (69.8%). In practice, the trial would not be completely stopped under these circumstances, but lower dose combinations would be added to the trial. In the additional scenario 15, where all dose combinations were too toxic, the trial was stopped in 83.7% of the cases. Therefore, the stopping rule seemed efficient.

4.1 Sensitivity analysis

We conducted a sensitivity analysis in order to study the performance of our design using different prior distributions and parameters values. We used normal distributions centered at 0 for β_0, β_3 with high or low variance (from 10 to 50), and exponential or gamma distributions centered at 1 for β_1, β_2 with high or low variance (from 1 to 10). According to Table 5, we can see that the PCS for all scenarios were very similar under these different prior distributions.

4.2 Time-to-event outcome

In practice, a longer follow-up time may be required to assess the toxicity outcome. Therefore, the toxicity outcomes of some patients already treated in the trial may be unobserved (or censored) when a new patient is enrolled in the trial and is ready for dose assignment. Waiting to assess toxicity outcomes for each cohort before including a new one in the trial can greatly increase the duration of the trial. To overcome this issue, we extended our method by modeling toxicity as a time-to-event outcome. Following the TITE-CRM [30], we considered a weighted dose-toxicity relationship $\hat{w}F(D, \beta_0, \beta_1, \beta_2, \beta_3)$ where $\hat{w}$ is monotone and increasing with patient follow-up time such that $0 \leq \hat{w} \leq 1$, and the toxicity probability model $F(D, \beta_0, \beta_1, \beta_2, \beta_3) = \pi_{j,k}$ is the same as that in (1).

Let T be a (maximum) fixed time window during which patients are followed, and $y_{i,N}$, $C_{i,N}$, and $\hat{w}_{i,N}$, respectively, be the indicator of a DLT for patient i prior to the entry of the $(N+1)^{\text{th}}$ patient, the follow-up time of patient i prior to the entry of the $(N+1)^{\text{th}}$ patient, and the weight assigned to patient i prior to the entry of the $(N+1)^{\text{th}}$ patient. Let X_i denote the time to toxicity of the ith patient, and (j_i, k_i) the combination received by patient i. Before combination assignment, the likelihood is defined as

$$L(\beta_0, \beta_1, \beta_2, \beta_3 | \text{data}) = \prod_{i=1}^{N} (\hat{w}_{i,N} \pi_{j_i,k_i})^{y_{i,N}} (1 - \hat{w}_{i,N} \pi_{j_i,k_i})^{1-y_{i,N}}.$$

Following Cheung and Thall [31], we chose the weights $\hat{w}_{i,N}$ as

$$\hat{w}_{i,N} = \frac{\# \{m/X_m \leq C_{i,N}, C_{m,N} \geq T\} + \hat{w}_{i,N}^0}{\# \{m/X_m \leq T, C_{m,N} \geq T\} + 1},$$

where m refers to patient, $\# \{m/X_m \leq T, C_{m,N} \geq T\}$ is the number of completely followed patients on who a toxicity was observed, and $\hat{w}_{i,N}^0$ is the linear weight for patient i defined as $\hat{w}_{i,N}^0 = C_{i,N}/T$, that is the proportion of time patient i was followed compared to the full follow-up time.

We simulated the time-to-toxicity outcomes using an exponential distribution such that the toxicity probabilities at the end of follow-up matched those given in Table 1. Using the same working model, we chose a full follow-up time of 3 weeks, and simulated patient accrual using a Poisson process with parameter 1, meaning that an average of one patient arrived every week.

Table 6 shows the results of the extended LOGISTIC for all 15 scenarios. We observe that the performance of the design decreases only slightly by 2%, and the PCS for all scenarios are still very high. This demonstrates that the LOGISTIC can be used in trials when toxicity cannot be assessed quickly.

5 Discussion

We have proposed a statistical method for clinical trial designs that evaluate various dose combinations for two agents. This method seems to be capable of identifying the MTD with a high percentage in all scenarios. Indeed, this method works for a wide variety of drug combinations. Our aim was not to find the correct order of toxicity within the drug combination space but to identify the right dose combination to be evaluated further in terms of efficacy. We compared our

method with five other model-based designs for combination drug trials. All the designs seem to be efficient when the MTDs are located on the same diagonal in the combination space. One benefit of our method compared with the other proposed designs is that it is also efficient when the MTDs are not necessarily located on the same diagonal.

We have designed our method to enable the selection of only one MTD at the end of the trial. There are various reasons why investigators may wish to select one MTD per level of Agent 2. For this purpose, our method can be modified by selecting, for each level k of Agent 2, the combination (j_k, k) such that $j_k = \underset{1 \leq j \leq J}{\operatorname{argmax}} P(\pi_{j,k} \in [\theta - \delta; \theta + \delta])$. In the case where it cannot be assumed that one MTD exists per level of Agent 2, a minimum probability of being in the targeted toxicity interval or a maximum probability of overdosing can be added in the determination of the MTD so as not to recommend a combination in the case of an MTD not existing for this dose level of Agent 2. Our design considers only toxicity responses, which can be improved upon by taking into account efficacy when determining the combination to be used in subsequent phases.

When combining several agents, designs developed for single-agent dose-finding trials cannot be applied to combination studies. For instance, the standard algorithm-based method for phase I dose-finding clinical trials in oncology is the so-called "3+3" design, which is referred to as "memory-less" since allocation to the next dose level for an incoming group of 3 patients depends only upon what has happened to the total of 3 to 6 patients previously treated at the current dose level [14, 15, 16, 17, 18]. This method was not designed for cases in which the full toxicity ordering is unknown. Indeed, the full ordering between toxicity probabilities of each combination is not fully known. That is why some investigators fix the dose level of one agent when they set up a combination study. They can thereby bring the problem back to one dimension and use single-agent designs. Another approach studied in this paper is to pre-select a subset of combinations whose toxicity ordering is known, and then apply a single-agent design to this reduced number of combinations. This approach performs well if the target dose combinations happen to be included in the subset. However, in practice, due to potential interaction between combined drugs, it can never be guaranteed that the target doses are always pre-selected into the subset of combinations to be inverstigated. If target doses are not in the subset, the trial would completely fail to find the target doses.

The advantage of our method is that it uses a logistic model that can be understood by a broad panel of readers and the performance seems to outperform existing designs. We also added stopping rules to the design in order to accommodate practical issues. The method can be efficient with different cohort sizes. Nevertheless, estimating parameter values and the different probabilities continues to require some advanced computational skills. To overcome this challenge, we have developed executable files that can be used to (1) determine the next combination and the MTD and estimate toxicity probabilities from data in actual trials, and (2) perform simulations before setting up a combination dose-finding trial. These files are freely available upon request.

Acknowledgements

References

[1] Zhou Y. Choice of designs and doses for early phase trials. *Fundam Clin Pharmacol* Jun 2004; **18**(3):373–378.

[2] Chevret S. *Statistical Methods for dose-Finding Experiments*. Statistics in Practice, John Wiley and Sons Ltd.: Chichester, 2006.

[3] *Common Terminology Criteria for Adverse Events (CTCAE) and Common Toxicity Criteria (CTC)*. Rockville, Maryland, USA 2009.

[4] Conaway MR, Dunbar S, Peddada SD. Designs for single- or multiple-agent phase I trials. *Biometrics* Sep 2004; **60**:661–669.

[5] Wages NA, Conaway MR, O'Quigley J. Continual Reassessment Method for Partial Ordering. *Biometrics* 2011; .

[6] Thall PF, Millikan RE, Mueller P, Lee SJ. Dose-finding with two agents in Phase I oncology trials. *Biometrics* Sep 2003; **59**(3):487–496.

[7] Wang K, Ivanova A. Two-dimensional dose finding in discrete dose space. *Biometrics* 2005; **61**:217–222.

[8] Mandrekar SJ, Cui Y, Sargent DJ. An adaptive phase I design for identifying a biologically optimal dose for dual agent drug combinations. *Stat Med* May 2007; **26**:2317–2330.

[9] Mandrekar SJ, Qin R, Sargent DJ. Model-based phase I designs incorporating toxicity and efficacy for single and dual agent drug combinations: methods and challenges. *Stat Med* May 2010; **29**:1077–1083.

[10] Yin G, Yuan Y. A latent contingency table approach to dose finding for combinations of two agents. *Biometrics* 2009; **65**:866–875.

[11] Yin G, Yuan Y. Bayesian dose finding in oncology for drug combinations by copula regression. *JRSS* 2009; **58**:211–224.

[12] Bailey S, Neuenschwander B, Laird G, Branson M. A Bayesian case study in oncology Phase I combination dose-finding using logistic regression with covariates. *J Biopharm Stat* 2009; **19**:469–484.

[13] Riviere MK, Dubois F, Zohar S. Competing designs for drug combination in phase I dose-finding clinical trials. *Stat Med* Jan 2014; .

[14] Faries D. Practical modifications of the continual reassessment method for phase I cancer clinical trials. *J Biopharm Stat* 1994; **4**:147–64.

[15] Goodman SN, Zahurak ML, Piantadosi S. Some practical improvements in the continual reassessment method for phase I studies. *Stat Med* Jun 1995; **14**(11):1149–1161.

[16] Korn EL, Midthune D, Chen TT, Rubinstein LV, Christian MC, Simon RM. A comparison of two phase I trial designs. *Stat Med* 1994; **13**:1799–806.

[17] O'Quigley J, Zohar S. Experimental designs for phase I and phase I/II dose-finding studies. *Br. J. Cancer* Mar 2006; **94**(5):609–613.

[18] Iasonos A, Wilton AS, Riedel ER, Seshan VE, Spriggs DR. A comprehensive comparison of the continual reassessment method to the standard 3+3 dose escalation scheme in Phase I dose-finding studies. *Clin Trials* 2008; **5**(5):465–477.

[19] Yuan Y, Yin G. Sequential continual reassessment method for two-dimensional dose finding. *Stat Med* Nov 2008; **27**(27):5664–5678.

[20] Ivanova A, Wang K. A non-parametric approach to the design and analysis of two-dimensional dose-finding trials. *Stat Med* 2004; **23**:1861–1870.

[21] O'Quigley J, Pepe M, Fisher L. Continual reassessment method: a practical design for phase 1 clinical trials in cancer. *Biometrics* 1990; **46**:33–48.

[22] Zohar S, Resche-Rigon M, Chevret S. Using the continual reassessment method to estimate the minimum effective dose in phase II dose-finding studies: a case study. *Clin Trials* 2013; **10**(3):414–421.

[23] Gilks WR, Best NG, Tan KKC. Adaptive Rejection Metropolis Sampling within Gibbs Sampling. *JRSS* 1995; **44**:455–472.

[24] Sweeting MJ, Mander AP. Escalation strategies for combination therapy Phase I trials. *Pharm Stat* Mar 2012; .

[25] Thall PF, Simon R. Practical Bayesian guidelines for phase IIB clinical trials. *Biometrics* Jun 1994; **50**(2):337–349.

[26] Storer BE. Design and analysis of phase I clinical trials. *Biometrics* 1989; **45**:925–37.

[27] Ivanova A, Montazer-Haghighi A, Mohanty SG, Durham SD. Improved up-and-down designs for phase I trials. *Stat Med* Jan 2003; **22**(1):69–82.

[28] Simon R, Freidlin B, Rubinstein L, Arbuck SG, Collins J, Christian MC. Accelerated titration designs for phase I clinical trials in oncology. *J. Natl. Cancer Inst.* Aug 1997; **89**(15):1138–1147.

[29] Gasparini M, Bailey S, Neuenschwander B. Bayesian dose finding in oncology for drug combinations by copula regression. *JRSS C* 2010; **59**:543–546.

[30] Cheung YK, Chappell R. Sequential designs for phase I clinical trials with late-onset toxicities. *Biometrics* Dec 2000; **56**(4):1177–1182.

[31] Cheung YK, Thall PF. Monitoring the rates of composite events with censored data in phase II clinical trials. *Biometrics* Mar 2002; **58**(1):89–97.

Appendix

Gibbs sampling

Gibbs sampling was used to generate a sequence of samples from the joint probability distribution of $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)$. The Gibbs sampler relies on the availability of all complete conditional distributions. We start from an arbitrary point $\beta^{(0)} = (\beta_0^{(0)}, \beta_1^{(0)}, \beta_2^{(0)}, \beta_3^{(0)})$ in the distribution, we sampled in turn in each of the full conditional distribution in updating them as the process goes on.

---

(1) Initialize $\beta^{(0)} = (\beta_0^{(0)}, \beta_1^{(0)}, \beta_2^{(0)}, \beta_3^{(0)})$

For m from 1 to $M + N$

(2) Sample

- $\beta_0^{(m)}$ from its full conditional distribution $f(\beta_0|\beta_1^{(m-1)}, \beta_2^{(m-1)}, \beta_3^{(m-1)}, \text{data})$
- $\beta_1^{(m)}$ from its full conditional distribution $f(\beta_1|\beta_0^{(m-1)}, \beta_2^{(m-1)}, \beta_3^{(m-1)}, \text{data})$
- $\beta_2^{(m)}$ from its full conditional distribution $f(\beta_2|\beta_0^{(m-1)}, \beta_1^{(m-1)}, \beta_3^{(m-1)}, \text{data})$
- $\beta_3^{(m)}$ from its full conditional distribution $f(\beta_3|\beta_0^{(m-1)}, \beta_1^{(m-1)}, \beta_2^{(m-1)}, \text{data})$

$\Rightarrow$ return $(\beta^{(M+1)}, \ldots, \beta^{(M+N)})$

We chose $M = 2000$ number of "burn-in" iterations to be discarded before convergence. $\beta^{(M)} = (\beta_0^{(M)}, \beta_1^{(M)}, \beta_2^{(M)}, \beta_3^{(M)})$ converges in distribution to the posterior joint distribution $f(\beta_0, \beta_1, \beta_2, \beta_3|\text{data})$. The following $N = 5000$ values retained are then considered as a sample from this distribution.

Adaptive Rejection Metropolis Sampling within Gibbs sampling

Adaptive Rejection Sampling (ARS) can not be used to sample from non log-concave distributions. When this is the case, Gilks, Best and Tan [23] propose to replace the rejection sampling in favor of the Hastings-Metropolis algorithm to update one parameter at a time. But to avoid high probabilities of rejection, they adapted the proposal density to the shape of the full conditional density using ARS. They added to ARS a single Hastings-Metropolis step thus creating ARMS within Gibbs sampling. The detailed method can be found in Gilks [23].

Table 1: Toxicity scenarios for the two-agent combinations.

	Agent 1									
Agent 2	1	2	3	4	5	1	2	3	4	5
	Scenario 1					Scenario 2				
3	0.15	**0.30**	0.45	0.50	0.60	0.45	0.55	0.60	0.70	0.80
2	0.10	0.15	**0.30**	0.45	0.55	**0.30**	0.45	0.50	0.60	0.75
1	0.05	0.10	0.15	**0.30**	0.45	0.15	**0.30**	0.45	0.50	0.60
	Scenario 3					Scenario 4				
3	0.10	0.15	**0.30**	0.45	0.55	0.50	0.60	0.70	0.80	0.90
2	0.07	0.10	0.15	**0.30**	0.45	0.45	0.55	0.65	0.75	0.85
1	0.02	0.07	0.10	0.15	**0.30**	**0.30**	0.45	0.60	0.70	0.80
	Scenario 5					Scenario 6				
3	0.07	0.09	0.12	0.15	**0.30**	0.15	**0.30**	0.45	0.50	0.60
2	0.03	0.05	0.10	0.13	0.15	0.09	0.12	0.15	**0.30**	0.45
1	0.01	0.02	0.08	0.10	0.11	0.05	0.08	0.10	0.13	0.15
	Scenario 7					Scenario 8				
3	**0.30**	0.50	0.60	0.65	0.75	0.08	0.15	0.45	0.60	0.80
2	0.15	**0.30**	0.45	0.52	0.60	0.05	0.12	**0.30**	0.55	0.70
1	0.07	0.10	0.12	0.15	**0.30**	0.02	0.10	0.15	0.50	0.60
	Scenario 9					Scenario 10				
3	**0.30**	0.37	0.42	0.47	0.52	0.08	0.10	0.15	**0.30**	0.50
2	0.15	**0.30**	0.37	0.43	0.48	0.04	0.07	0.12	0.16	**0.30**
1	0.10	0.12	**0.30**	0.40	0.45	0.01	0.03	0.06	0.08	0.10
	Scenario 11					Scenario 12				
3	0.50	0.60	0.70	0.80	0.90	**0.30**	0.42	0.52	0.62	0.70
2	0.10	**0.30**	0.50	0.70	0.80	0.10	0.20	**0.30**	0.40	0.50
1	0.06	0.10	0.15	**0.30**	0.50	0.05	0.12	0.20	**0.30**	0.40
	Scenario 13					Scenario 14				
3	0.42	0.52	0.62	0.70	0.80	**0.30**	0.42	0.52	0.70	0.80
2	0.20	**0.30**	0.40	0.50	0.67	0.10	0.20	**0.30**	0.50	0.67
1	0.12	0.20	**0.30**	0.40	0.60	0.04	0.06	0.08	0.20	**0.30**

The true MTD combinations are shown in boldface.

Table 2: Percentage of correct selection in comparison of 6 dose-finding designs. Percentage of patients allocated to a true MTD during the trial and mean number of observed DLTs throughout the trial.

scenario	sc1	sc2	sc3	sc4	sc5	sc6	sc7	sc8	sc9	sc10	sc11	sc12	sc13	sc14
Percentages of correct selection (PCS)														
CRM anti-diag	73.7	74.8	71.9	84.9	80.0	71.4	73.2	84.3	58.4	74.8	82.8	61.0	59.4	73.2
BCOPULA	66.2	71.8	71.7	84.1	78.1	30.7	49.6	43.5	66.5	60.9	42.7	**57.3**	44.8	42.3
BGUMBEL	67.1	72.5	68.4	87.5	77.9	33.6	48.0	49.5	64.7	66.2	51.2	**56.3**	48.7	42.9
I2D	68.0	73.7	66.9	89.7	83.7	37.2	41.9	50.4	63.8	66.7	47.1	45.3	41.6	47.4
LOGODDS	62.8	72.7	63.7	87.5	**90.0**	46.7	**73.2**	12.1	57.3	68.4	61.1	**56.5**	47.0	54.5
TMML	70.1	70.5	65.8	**92.1**	83.9	49.6	65.6	**69.7**	61.7	66.0	**75.9**	47.9	**60.1**	56.0
LOGISTIC	**75.4**	**80.5**	**74.9**	86.7	80.4	**63.7**	**71.2**	56.9	**69.6**	**75.1**	**77.8**	**56.7**	**60.0**	**61.0**
Percentage of patients allocated to a true MTD during the trial														
CRM anti-diag	49.2	55.4	45.2	74.5	45.5	43.5	52.6	58.4	44.1	42.6	59.8	42.1	43.6	49.3
BCOPULA	40.0	50.1	40.3	84.1	27.8	16.6	38.3	23.6	46.2	32.2	34.3	37.5	26.7	33.8
BGUMBEL	40.8	52.8	39.5	81.6	30.5	20.0	34.1	26.0	46.2	34.3	34.2	37.3	31.5	32.0
I2D	44.1	55.6	38.9	79.8	34.6	23.0	32.0	24.0	46.2	30.8	33.5	32.2	29.9	32.4
LOGODDS	40.2	45.3	40.3	82.0	62.5	29.2	47.8	9.8	38.2	43.2	42.8	25.2	31.3	24.8
TMML	34.5	40.3	33.7	89.8	42.0	22.2	33.5	24.8	36.5	30.0	35.3	25.0	27.0	26.7
LOGISTIC	44.0	50.5	46.0	78.0	40.0	33.0	43.5	24.3	47.3	38.0	44.0	33.0	40.7	34.3
Mean number of observed DLTs throughout the trial														
CRM anti-diag	16.5	18.4	15.3	20.3	11.5	14.8	17.6	17.1	17.0	14.1	18.0	16.2	17.0	16.5
BCOPULA	14.2	16.1	12.7	19.5	9.2	12.8	14.6	14.3	15.0	11.5	15.5	14.5	15.2	14.5
BGUMBEL	14.6	16.6	13.2	19.7	9.5	13.5	15.6	14.7	15.4	11.9	16.1	15.0	15.5	15.2
I2D	15.3	17.6	14.1	19.9	10.1	14.3	16.1	15.3	16.2	12.9	16.6	15.6	16.6	15.8
LOGODDS	16.8	17.5	16.4	20.0	13.7	16.6	17.0	16.3	17.1	16.0	16.7	16.8	16.8	16.8
TMML	13.8	15.7	13.0	19.2	11.4	13.1	14.2	13.9	14.5	12.4	14.7	14.1	14.8	13.9
LOGISTIC	15.2	17.7	14.1	20.4	11.4	14.3	15.5	15.3	15.9	12.7	16.2	15.7	16.4	15.4

In the top third of the table, the best PCS and those associated with 2% less are shown in boldface.

Table 3: Combination selection percentages of the proposed method for combination evaluation.

Agent 2	Agent 1									
	1	2	3	4	5	1	2	3	4	5
	Scenario 1					Scenario 2				
3	6.2	**27.5**	7.1	0.1	0.0	0.2	0.1	0.0	0.0	0.0
2	0.1	4.6	**41.7**	2.9	0.1	35.4	11.2	0.1	0.0	0.0
1	0.0	0.1	3.2	**6.2**	0.1	5.8	**45.1**	2.1	0.0	0.0
	Scenario 3					Scenario 4				
3	0.5	6.2	**43.1**	7.5	0.1	0.0	0.1	0.0	0.0	0.0
2	0.0	0.2	4.5	**29.1**	2.0	4.2	0.1	0.0	0.0	0.0
1	0.0	0.0	0.1	4.0	**2.7**	86.7	9.0	0.0	0.0	0.0
	Scenario 5					Scenario 6				
3	0.0	0.1	1.5	8.5	**80.4**	3.0	**44.0**	14.1	1.4	0.1
2	0.0	0.0	0.1	1.0	7.8	0.0	1.2	11.1	**19.7**	2.4
1	0.0	0.0	0.0	0.1	0.5	0.0	0.0	0.2	2.1	0.5
	Scenario 7					Scenario 8				
3	**15.8**	2.4	0.1	0.0	0.0	0.9	15.5	17.0	0.0	0.0
2	4.2	**54.4**	15.6	1.1	0.0	0.0	0.9	**56.9**	0.5	0.0
1	0.0	0.5	3.4	1.6	**1.0**	0.0	0.0	2.6	5.7	0.0
	Scenario 9					Scenario 10				
3	**8.1**	5.7	1.8	0.2	0.0	0.1	0.4	7.1	**43.4**	8.2
2	6.4	**48.2**	11.7	1.5	0.0	0.0	0.0	0.3	8.1	**31.7**
1	0.1	1.8	**13.3**	1.4	0.0	0.0	0.0	0.0	0.1	0.8
	Scenario 11					Scenario 12				
3	6.7	0.2	0.0	0.0	0.0	**22.7**	12.5	0.5	0.0	0.0
2	2.4	**75.3**	6.0	0.1	0.0	0.4	21.3	**31.4**	2.1	0.0
1	0.0	0.2	6.6	**2.5**	0.1	0.0	0.2	6.2	**2.6**	0.0
	Scenario 13					Scenario 14				
3	7.0	0.8	0.0	0.0	0.0	**24.5**	12.7	0.6	0.0	0.0
2	13.7	**48.0**	5.3	0.0	0.0	0.1	18.1	**35.6**	3.3	0.1
1	0.4	11.9	**12.0**	0.9	0.0	0.0	0.0	0.7	3.5	**0.9**

The percentages for the true MTD combinations are shown in boldface.

Table 4: Additional toxicity scenario for the two-agent combinations.

Agent 2	Agent 1				
	1	2	3	4	5
	Scenario 15				
3	0.60	0.65	0.75	0.85	0.90
2	0.55	0.60	0.70	0.75	0.80
1	0.45	0.50	0.55	0.60	0.65

Table 5: Percentage of correct selection (PCS) and percentage of stopping for toxicity without MTD recommendation. Results of the sensitivity analysis with different prior distributions, using for β_0, β_3 normal distribution with high or low variability and for β_1, β_2 exponential or gamma priors with high or low variability.

scenario	sc1	sc2	sc3	sc4	sc5	sc6	sc7	sc8	sc9	sc10	sc11	sc12	sc13	sc14	sc15
$\beta_0, \beta_3 \sim N(0,10)$, $\beta_1, \beta_2 \sim \mathbb{E}(1)$															
PCS	76.7	79.5	74.9	69.8	80.4	64.6	70.6	57.6	68.3	75.1	77.8	56.7	60.7	61.0	-
stop	0.0	0.7	0.0	16.8	0.0	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.4	0.0	83.7
$\beta_0, \beta_3 \sim N(0,50)$, $\beta_1, \beta_2 \sim \Gamma(0.1, 0.1)$															
PCS	72.3	81.8	71.0	68.9	74.8	63.4	70.4	57.9	70.4	72.2	75.1	57.0	58.0	60.9	-
stop	0.0	0.9	0.0	17.8	0.0	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.4	0.0	83.5
$\beta_0 \sim N(0,50)$, $\beta_1, \beta_2 \sim \Gamma(0.5, 0.5)$, $\beta_3 \sim N(0,10)$															
PCS	75.0	82.1	70.4	70.0	79.0	62.0	70.0	60.3	68.6	72.8	79.4	55.0	57.5	61.6	-
stop	0.0	0.9	0.0	17.6	0.0	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.5	0.0	84.1

Table 6: Percentage of correct selection (PCS) and percentage of stopping for toxicity without MTD recommendation when toxicity is considered as a time-to-event outcome.

scenario	sc1	sc2	sc3	sc4	sc5	sc6	sc7	sc8	sc9	sc10	sc11	sc12	sc13	sc14	sc15
TITE-LOGISTIC															
PCS	74.0	77.5	74.1	68.0	78.2	63.3	66.5	57.6	64.7	74.6	76.2	56.5	60.4	62.7	-
stop	0.1	0.8	0.0	16.6	0.0	0.1	0.1	0.0	0.2	0.0	0.1	0.1	0.2	0.0	82.4

Figure 1: Posterior densities of the toxicity probability for each combination. The dashed line represents the toxicity target.
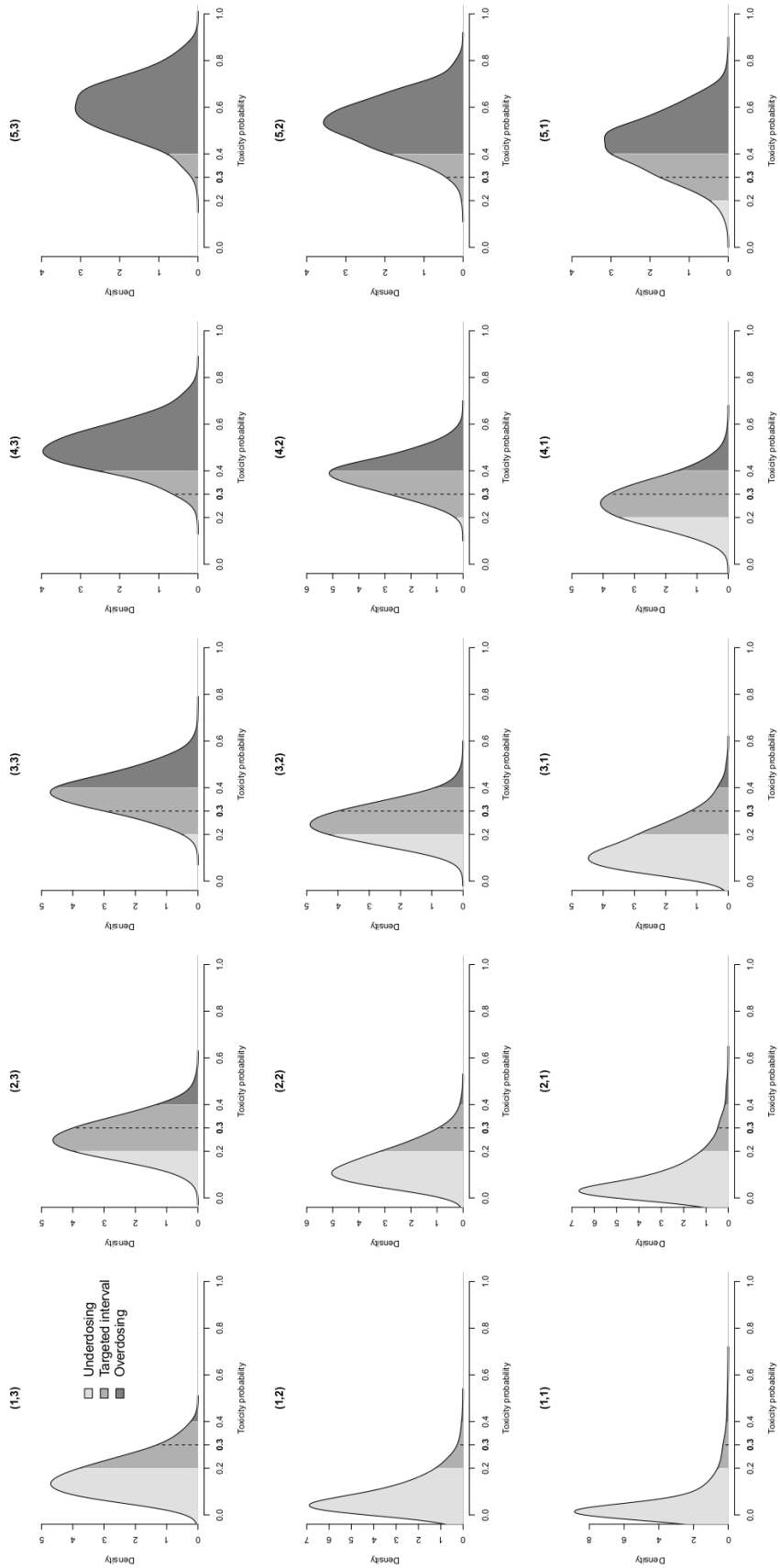
Figure 2: Probability of achieving the targeted interval, $P\left(\pi_{j,k} \in [\theta - \delta; \theta + \delta]\right)$ at the end of the trial for each combination. The dashed vertical line corresponds to the highest probability, which determines the combination selected as the MTD ($(2, 3)$ in this example).