A comparison of Phase I dose finding designs in clinical trials with monotonicity assumption violation

Rachid Abbas*[1], Caroline Rossoni[1], Thomas Jaki[2,3], Xavier Paoletti[4], and Pavel Mozgunov[2]

[1]*ONCOSTAT Team CESP INSERM U1018, Univ. Paris-Saclay and Biostatistics and Epidemiology department, Gustave Roussy Cancer Center, Villejuif, France.*

[2]*Department of Mathematics and Statistics, Lancaster University, UK.*

[3]*MRC Biostatistics Unit, University of Cambridge, Cambridge, UK.*

[4] *Université Versailles St Quentin & INSERM U900 STAMPM, Institut Curie, Paris, France.*

[*]*corresponding author: Dr ABBAS Rachid, 114 rue Edouard Vaillant, Gustave Roussy Cancer Center, 94805 Villejuif Cedex, France.*

rachid.abbas@gustaveroussy.fr (phone number: 0033 142115398)

Short title: Phase I with monotonicity assumption violation

4853 words

Abstract

Background/Aims: In oncology, new combined treatments make it difficult to order dose levels according to monotonically increasing toxicity. New flexible dose-finding designs that take into account uncertainty in dose levels ordering were compared to classical designs through simulations in the setting of the monotonicity assumption violation. We give recommendations for the choice of dose-finding design.

Methods: Motivated by a clinical trial for patients with high-risk neuroblastoma, we considered designs that require a monotonicity assumption, the Bayesian Continual Reassessment Method, the modified Toxicity Probability Interval, the Bayesian Optimal Interval design, and designs that relax monotonicity assumption, the Bayesian Partial Ordering Continual Reassessment Method and the No Monotonicity Assumption design. We considered 15 scenarios including monotonic and non-monotonic dose-toxicity relationships among 6 dose levels.

Results: The No Monotonicity Assumption and Partial Ordering Continual Reassessment Method designs were robust to the violation of the monotonicity assumption. Under non-monotonic scenarios, the No Monotonicity Assumption design selected the correct dose level more often than alternative methods on average. Under the majority of monotonic scenarios, the Partial Ordering Continual Reassessment Method selected the correct dose level more often than the No Monotonicity Assumption design. Other designs were impacted by the violation of the monotonicity assumption with a proportion of correct selections below 20% in most scenarios. Under monotonic scenarios, the highest proportions of correct selections were achieved using the Continual Reassessment Method and the Bayesian Optimal Interval design (between 52.8% to 73.1%). The costs of relaxing the monotonicity assumption by the No Monotonicity Assumption design and Partial Ordering Continual Reassessment Method were decreases in the proportions of correct selections under monotonic scenarios ranging from 5.3% to 20.7% and from 1.4% to 16.1%, respectively compared to the best performing design and were higher proportions of patients allocated to toxic dose levels during the trial.

Conclusions: Innovative oncology treatments may no longer follow monotonic dose levels ordering which makes standard phase I methods fail. In such a setting, appropriate designs, as the No Monotonicity Assumption or Partial Ordering Continual Reassessment Method designs, should be used to safely determine recommended for phase II dose.

Keywords: Dose Escalation; Monotonicity assumption; Oncology; Partial Ordering; Phase I.

Introduction

The main objective of phase I clinical trials is to identify the maximum tolerated dose (MTD). The MTD is defined as the highest dose that can be administrated, targeting an acceptable rate of toxicity defined as dose-limiting toxicities (DLT). A number of study designs has been proposed to identify the MTD in single agent trials: algorithm-based, e.g. the 3+3 method[1], model-based e.g. the Continual Reassessment Method[2], and model-assisted designs, e.g. the Toxicity Probability Interval method[3] or the Bayesian Optimal Interval designs[4]. The common feature of these methods is the assumption of a monotonic relationship between toxicity and dose that holds in single-agent trials.[5] However, it becomes more common in oncology to combine targeted treatments with established cytotoxic chemotherapy regimens.[6] In case of a synergistic interaction between two treatments, non-monotonic shapes, namely a plateau or a bi-modal relationship, can be expected.[7]

An example is the SIOPEN's clinical trial in high-risk neuroblastoma which motivated this work. Neuroblastoma is the most frequent individual type of solid tumour in children.[8] Pre-clinical data suggests that the use of immunotherapy (the dinutuximab beta targeting the disialoganglioside GD2) in combination with a conventional chemotherapy can improve the induction treatment.[9;10;11] The SIOPEN clinical trial will assess the toxicity of dinutuximab combined with the induction chemotherapy under different

schedules. The combination of dinutuximab with this chemotherapy regimen varies depending on the dose of dinutuximab and also on the schedule of administration (see Figure 1). In the following, a *dose level* refers to a combination of immunotherapy dose and extent of co-administration with chemotherapy. The starting dose level corresponds to a sequential administration of chemotherapy and dinutuximab with no overlapping administration: standard chemotherapy regimen drugs (cisplatin and etoposide) are given from day 1 to 4 and dinutuximab is started on day 5. The dose escalation is designed to reach the highest dose of immunotherapy and a concomitant administration of immunotherapy and chemotherapy, dose level 5 and 6.

The main challenge of this trial is that not all dose levels can be ordered according to a monotonically increasing toxicity. Specifically, it is unknown whether dose levels 3 and 4 are more or less toxic than dose levels 4 and 5, respectively. Thus, the monotonicity assumption upon which standard single-agent dose-finding designs are based is violated, and corresponding methods cannot be efficiently applied. Furthermore, the combination space in this example, while partially ordered, is not a matrix of combination comprised of discrete levels of two agents that is the framework, for which many existing combination methods are developed for.[12] As a result, these methods cannot be applied to conduct dose-finding in the motivating example. Others approaches considering the related schedule-finding problem have been published. Braun et al.[13] proposed a design to find the best

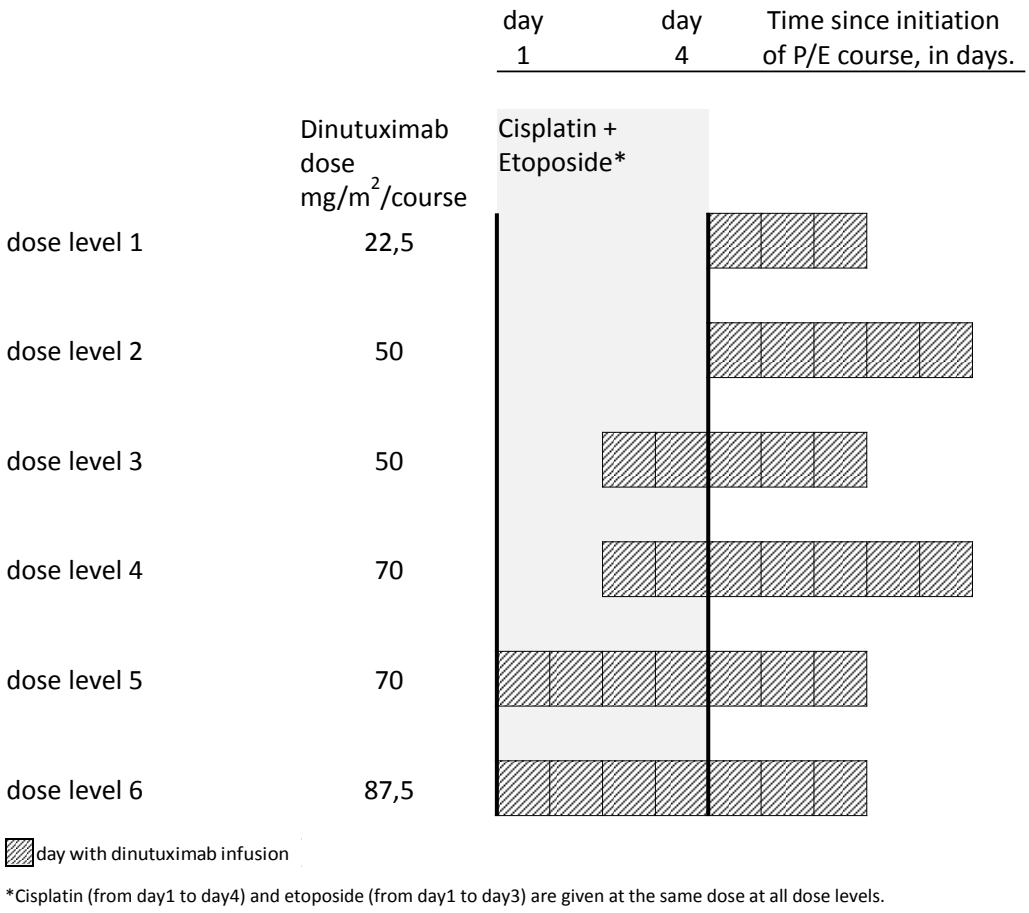

Figure 1: Dose levels for the combination trial of dinutuximab and chemotherapy course (P/E).

'Maximum Tolerated Schedule'. In this design, the main concern is delayed toxicity occurring after repeated treatment administrations. Thus, the primary endpoint is the time to toxicity. This is of interest for applications where toxicity is expected after the first administration. In our setting, toxicity is expected in the first cycle. Thall et al.[14] proposed an extension of the method by Braun et al.[13] to jointly optimize dose and schedule in a two step procedure. Guo and Yuan[15] also proposed a design based on a Bayesian dynamic model that jointly optimizes dose and schedule for toxicity and efficacy. These dose-schedule designs assume that "the duration of therapy

increases" and do not straightforwardly allow for the incorporation of the uncertainty in the toxicity ordering between "dose levels". For example, in their approach, of Guo and Yuan assume that the mean toxicity probability can be written as the function of difference in two doses. In our example, however, it remains unclear what the distance between two doses would be as the "dose-levels" are "combination of two treatments given at different doses". Furthermore, in the motivating setting, one needs to deal with the effect arising from the therapies given sequentially or with overlapping (for various number of days) that complicates the problem further. Therefore, we did not include these methods in our comparison. Instead, two recently proposed dose-finding designs, the model-based Partial Ordering Continual Reassessment Method by Wages et al.[16], and the model-free "No Monotonicity Assumption" design by Mozgunov and Jaki[17], are flexible enough to be applied to the motivating trial directly.

The aim of our study is to investigate the performance of several Phase I dose-finding designs in the setting of the monotonicity assumption violation. Specifically, we aim to investigate the consequences of applying designs based on the monotonicity to such settings, and to study how well the alternative designs overcome the uncertainty about the ordering. While several comparison of dose-finding designs have been proposed in the literature [18;19;20;21], they primarily concern settings with a known monotonicity ordering, or specific combination settings (see e.g. Riviere et al.[12]); the comparison studies in settings similar to the motivating trial are still sparse to date.

This article is structured as follows. In the methods section, we introduce the designs that relax monotonicity assumption and their statistical principles. In the results section, we specify the simulation settings, the elicitation of optimal parameters for each design and provide operating characteristics in the scenarios inspired by the motivating example. We conclude with practical recommendations and discussion in the discussion section.

Methods

Consider a clinical trial in which M dose levels $x_1, \ldots, x_M$ are studied and N patients are available in the trial. Let Y be a binary random variable, where $y_i = 1$ denotes the observation of a DLT for patient i. The DLT probability given x_m is denoted by $p_m = \mathbb{P}(Y = 1|x_m)$, $m = 1, \ldots, M$. The goal of the trial is to find the MTD x_T such that $\mathbb{P}(Y = 1|x = x_T) = \gamma$ where γ is the targeted rate of toxicity.

Below, we recall dose finding designs that relax the monotonicity assumption, the Partial Ordering Continual Reassessment Method and the No Monotonicity Assumption design. We refer the reader to O'Quigley et al.[2] for the details of the Continual Reassessment Method, to Ji et al.[3] for the details of the Toxicity Probability Interval method, and to Liu and Yuan[4] for the details of the Bayesian Optimal Interval design.

Partial ordering continual reassessment method

Consider a trial, in which at least two dose levels can not be ordered. Conaway et al.[22] proposed a framework of partial order restrictions in order to provide efficient toxicity estimations. Wages et al.[16] extended this approach using a working parametric model, such as proposed in O'Quigley et al.[2]. Let us assume that S monotonic orderings are considered as clinically plausible. For ordering s, DLT probabilities are assumed to have a parametric form $p_m = \psi_s(\alpha_m, \theta)$ where α_m is the standardized level for dose m and θ is the working model's unknown

parameter. Denote the prior distribution of θ by f_0 and let $r_i \in \{x_1, \ldots, x_M\}$ be the dose level recommended for the patient i. Then, as toxicity data for n patients is available, the likelihood under ordering s can be computed as

$$\mathcal{L}_{n,s}(\theta) = \prod_{i=1}^{n} \phi_s(r_i, y_i, \theta)$$

where ϕ_s is

$$\phi_s(r_i, y_i, \theta) = \psi_s(r_i, \theta)^{y_i}(1 - \psi_s(r_i, \theta))^{1-y_i}. \tag{1}$$

The posterior distribution for parameter θ under ordering s can be computed using Bayes's theorem as

$$f_{n,s}(\theta) = \frac{f_0(\theta) \prod_{i=1}^{n} \phi_s(r_i, y_i, \theta)}{\int_{\mathbb{R}} f_0(u) \prod_{i=1}^{n} \phi_s(r_i, y_i, u) \mathrm{d}u}. \tag{2}$$

In total, S dose-toxicity models are fitted. Given the prior distribution of orderings $q_s = \{q_1, \ldots, q_S\}$, the posterior probability of ordering s after n patients takes the form

$$\pi_{n,s} = \frac{q_s \int_R \mathcal{L}_{s,n}(u|y_1, \ldots, y_n) f_0(u) \mathrm{d}u}{\sum_{s=1}^{S} q_s \int_R \mathcal{L}_{s,n}(u|y_1, \ldots, y_n) f_0(u) \mathrm{d}u}. \tag{3}$$

The next group of patients is allocated based on ordering s^* corresponding to the maximum of $\pi_{n,s}$, $s = 1, \ldots, S$. For ordering s^*, let $\hat{\theta}_{s^*,n}$ be the corresponding posterior mean. Using the "plug-in" estimate for the DLT probability $\hat{p}_{m,n} = \psi_s^*(\alpha_{s^*}, \hat{\theta}_{s^*})$, we assign the next group of patients to the

dose level r_{n+1} minimizing for all m

$$r_{n+1} = \min_{m}(|\hat{p}_{m,n} - \gamma|) \tag{4}$$

The design proceeds until the maximum number of patients N have been treated or if the following safety constraint

$$\mathbb{P}\left(\psi(\alpha_1, \theta) > \gamma\right) > \xi \tag{5}$$

is violated, where ξ is a probability controlling overdosing, and the left-hand side is found with respect to the posterior density.

The Partial Ordering Continual Reassessment Method design is a model-based design and relies on a particular working model and pre-specified set of S ordering. One can benefit from using model-free designs in a setting with unexpected dose-toxicity shapes or in case of a large number of possible orderings.

No Monotonicity Assumption design

As an alternative to model-based designs, Mozgunov and Jaki[17] proposed an information-theoretic approach to dose-finding trials that does not require any parametric or monotonicity assumption. We will refer to this design as to a "No Monotonicity Assumption" design. This approach models DLT probabilities for

all dose levels as independent Beta random variables and uses the criterion

$$\delta(p_m) = \frac{(p_m - \gamma)^2}{p_m(1 - p_m)} \tag{6}$$

for the dose-escalation. The No Monotonicity Assumption design does not imply a monotonicity assumption in the decision rules. It was found that due to special properties of (6), the correct MTD selection can be achieved without monotonicity assumption. We extend the estimator used by Mozgunov and Jaki[17;23] to decrease the contribution of the prior information faster and proportionally to n_m. The estimate of p_m, after observing t_m DLT for n_m patients assigned to the dose level x_m takes the form

$$\hat{p}_m = \frac{t_m + \frac{\nu_m}{n_m^\lambda}}{n_m + \frac{\beta_m}{n_m^\lambda}}$$

where λ is the down-weighting parameter and $\beta_m, \nu_m > 0$ are sets of parameters of the prior Beta distribution $\mathcal{B}(\nu_m + 1, \beta_m - \nu_m + 1)$ of probability p_m.

The first allocation is based on prior information ν_m, β_m, only. Assume that n_m patients were assigned to each dose levels and t_m DLT were observed, respectively. The next cohort of patients will be assigned to dose level x_m^* for which the criterion (6) is minimized. The design proceeds until the total number of patient N is reached or a safety constraint is violated. The No Monotonicity Assumption design uses a time-varying safety constraint[17;23].

The dose level x_m is *unsafe* if after n_m patients accrued at this dose level

$$\int_{\gamma}^{1} f_{n_m}(p)\mathrm{d}p \geq \xi_{n_m} = max(1 - kn_m, \xi_{final}) \tag{7}$$

where ξ_{n_m} is the overdosing probability, f_{n_m} is the Beta posterior, k the rate of the safety constraint strictness and ξ_{final} the final level of confidence. The increasing overdosing probability implies that the safety constraint becomes stricter as the trial goes. Similarly, the No Monotonicity Assumption design includes a futility constraint to avoid patient allocation to dose level below the MTD. Dose level x_m is futile if after n_m patients

$$\int_{\gamma'}^{1} f_{n_m}(p)\mathrm{d}p \leq \zeta \tag{8}$$

where γ' is the futility bound and ζ the controlling probability. Parameters of both safety and futility constraint should be calibrated over scenarios of interest using information about partial ordering. The No Monotonicity Assumption design incorporates the information about monotonicity between any pair of dose levels by restrictions known as *the coherent escalation/de-escalation principles*[24] formulated as follows:

- *Coherent escalation*: if at least one DLT was observed given a current dose level for a previous cohort, more toxic dose levels (with respect to a partial monotonic ordering) cannot be selected for next patient.
- *Coherent de-escalation*: if no DLT outcomes were observed for a previous cohort given the current dose level then less toxic dose cannot be selected for the next cohort.

At the end of the trial, the dose level minimizing the criterion (6) is designated as the MTD.

Results

Trial setting

To compare the operating characteristics of the dose-finding designs, we consider a simulation study in the setting of (partially) unknown ordering of dose levels. The SIOPEN's clinical trial for patients with high risk neuroblastoma which motivated this work, was a first step to improve the induction treatment[9;10;11], i.e. the first part of the treatment aiming to reduce the tumor burden in order to facilitate surgery and subsequent treatments. Different chemotherapy regimens have been evaluated with increasing intensities of conventional chemotherapies over the last four decades[25;26]. The SIOPEN's clinical trial will assess the toxicity of a combined treatment. We considered six dose levels combining immunotherapy and chemotherapy using different schedules with $n = 30$ patients. Based on past experience, the expected accrual rate is 4 patients by month and the toxicity is evaluated after 3 weeks, so we considered the inclusion of 3 patients cohorts in this trial. The goal was to find the MTD corresponding to the dose level with the toxicity probability closest to the target value of $\gamma = 0.30$. We considered a trial with one de-escalation dose level (d_1), a starting dose level (d_2) and four escalation dose levels (d_3 – d_6). The main challenge was that the administration of the drugs in combination (either simultaneously or sequentially) could have a

significant impact on the toxicity. A clinician could not put in a monotonic order dose levels d_3, d_4 and d_4, d_5. This resulted in two partial orderings (partial in the sense that we can only order 5 out of 6 dose levels):

$$d_1 < d_2 < d_3 < d_5 < d_6 \tag{9}$$

and

$$d_1 < d_2 < d_4 < d_5 < d_6 \tag{10}$$

These two orderings were studied by various scenarios given in Figure 2. Scenarios 1 and 2 corresponded to cases of excessive toxicity with the MTD

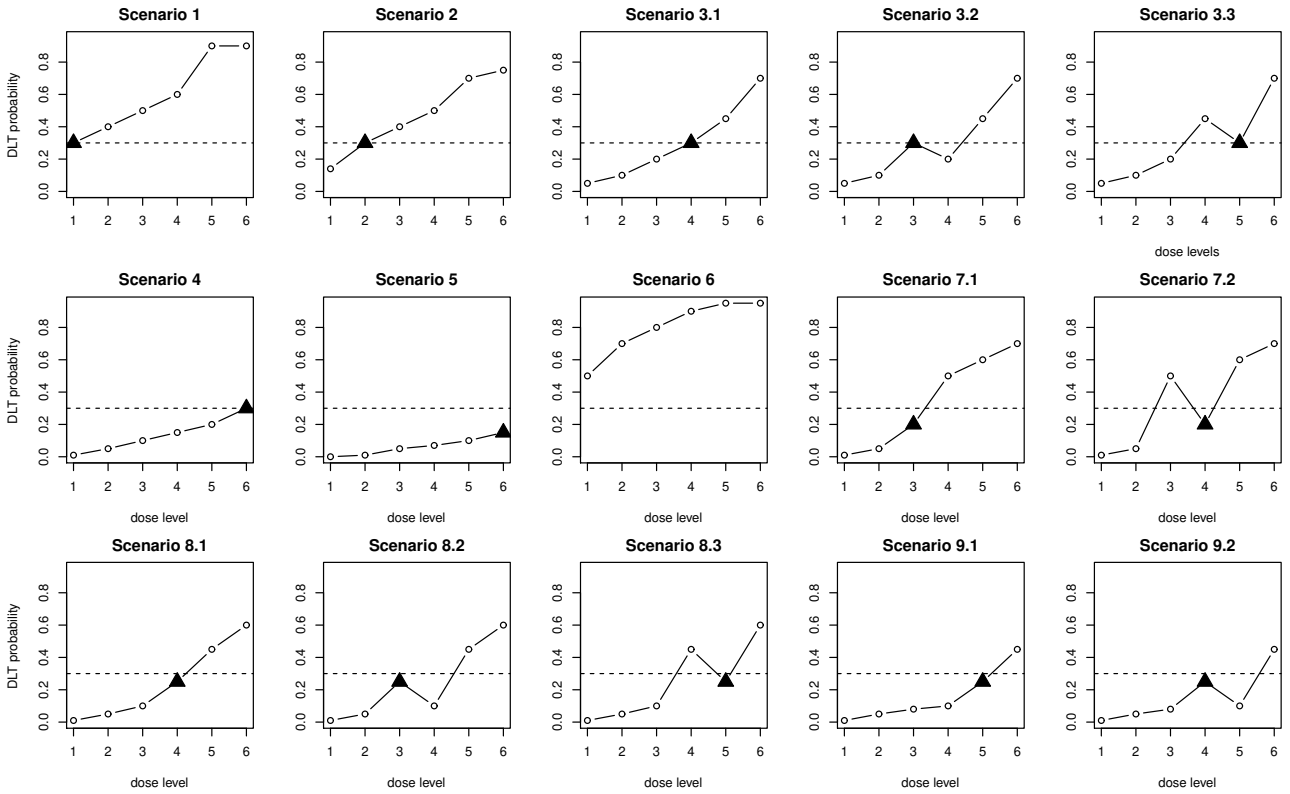

Figure 2: Considered dose-toxicity shapes. The MTD is marked by a triangle and the dashed horizontal line corresponds to the target DLT rate $\gamma = 0.3$.

being d_1 and d_2, respectively. For instance, scenario 1 accounted for the case in which an unexpected mechanism of toxicity occurred and the starting dose

level toxicity was higher than expected by a clinician. These scenarios would reveal the ability of the designs to avoid allocation to too toxic dose levels and recommend the MTD at the beginning of the dose-escalation. For scenarios 3, all possible permutation of the orderings due to unknown order of d_3, d_4 and d_4, d_5 were considered. The scenario 3.1 was a monotonic scenario with the MTD being d_4. Scenarios 3.2 and 3.3 accounted for cases where orders of d_3, d_4 and d_4, d_5 were inverted, with the MTD being d_3 and d_5 respectively. The non-monotonic order resulted in a non-trivial shape of the dose-toxicity relation. Scenarios 4 and 5 corresponded to flat dose-toxicity relations and were included to investigate the ability of the designs to escalate quickly and avoid patient's allocation to sub-therapeutic doses. Scenario 5 corresponded to the case of under dosing at all dose levels. Scenario 6, in which all doses were toxic, evaluated the capacity of the safety constraints to avoid an unethical recommendation to a toxic dose and an early stopping of study. Scenarios 7-9 corresponded to cases where none of the dose levels matches exactly the target toxicity probability of 0.30. These scenarios were chosen to check the designs trend to select overdosing or underdosing levels. Note that the permutation of dose levels d_3, d_4 and d_4, d_5 that did not change the MTD location are not included in those scenarios.

In the simulation study, we focused on (i) the proportion of dose level selection, (ii) the average number of patients assigned to each dose level, (iii) the proportion of early stopping due to excessive toxicity, (iv) the probability of selecting toxic dose levels as the MTD, (v) the percentage of patients

assigned to toxic dose levels (above the MTD). Before the analysis, each of the four designs were calibrated according to the stated trial setting.

Designs specification

For methods, which require parameters calibration, the "operational prior" (the prior working uniformly well in many different scenarios) were specified as follows.

For the Continual Reassessment Method and the Partial Ordering Continual Reassessment Method simulations, the working model for dose-toxicity was the one-parameter power model. For both methods, the following standardized values were used:

$$(0.20, 0.30, 0.40, 0.50, 0.59, 0.67)$$

These values were obtained using using the `getprior` function from `dfcrm` package[27] using a half width of 0.05, a target toxicity of $\gamma = 0.30$, and the prior MTD being the dose level 2. Sensitivity analyses showed that a half width of 0.05 provided better operating characteristics than other values (see Supplemental Materials for more details, considered values were 0.04, 0.06, 0.08 and 0.10). The least informative normal prior according to the algorithm by Cheung[28], $\theta \sim \mathcal{N}(0, 0.75^2)$ was used for both Continual Reassessment Method and Partial Ordering Continual Reassessment Method. Overdosing probability $\xi = 0.80$ was used for the safety constraint (5). In addition, the Partial Ordering Continual Reassessment Method required orderings to be specified. Using partial orderings 9 and 10, there were three feasible orderings

in this setting

$$d_1 < d_2 < d_3 < d_4 < d_5 < d_6$$

$$d_1 < d_2 < d_4 < d_3 < d_5 < d_6$$

$$d_1 < d_2 < d_3 < d_5 < d_4 < d_6$$

These three orderings were included in the Partial Ordering Continual Reassessment Method design and a clinician considered that all of them were equally possible a-priori $q_1 = q_2 = q_3 = 1/3$.

The Toxicity Probability Interval method requires specification of an equivalence interval according to a physician's advice. Following the motivating example the parameters $\epsilon_1 = \epsilon_2 = 0.10$ were chosen. For others parameters, the dose level 2 was used as the prior MTD and the cutoff probability for excessive toxicity was $\xi = 0.90$

The Bayesian Optimal Interval design required specification of an optimization interval, we used default values (0.18 - 0.42) as recommended by Liu and Yuan[4]. For sake of comparability, the stopping rule based on the cumulative number of patients treated at the same dose was not used. The safety parameters of the Bayesian Optimal Interval design proposed in the original work were used: the elimination dose threshold, $p_E = 0.95$, and the early stopping rule, $\delta = 0.05$. The use of the early stopping rule, as explained by Zhou et al.[29], was motivated by the fact that investigators emphasized the importance of conducting the motivating trial as safely as possible. The

operating characteristics of the Bayesian Optimal Interval design without using the early stopping rule can be found in Supplementary Materials.

The No Monotonicity Assumption design requires vectors $\nu = [\nu_1, \ldots, \nu_m]^{\mathrm{T}}$ and $\beta = [\beta_1, \ldots, \beta_m]^{\mathrm{T}}$ to be specified prior to the trial. For the prior information strength, $\beta = 1$ was chosen to emphasize a limited amount of information available about dose levels. Then, ν represents prior mode probabilities for all treatment levels (as standardized values above). These values were calibrated as described by Mozgunov and Jaki[17] over the considered set of scenarios. The operational prior used in the simulations was

$$\nu = [0.20, 0.23, 0.26, 0.29, 0.32, 0.35]^{\mathrm{T}}$$

This prior reflected the clinician elicitation that dose levels were in monotonic order. Note, that this ordering could change as the trial progresses in contrast to other methods. The prior probabilities were chosen to be close to each other to ensure that all dose levels could be tested if data suggested so. For the parameter of down-weighting $\lambda = 0.25$ was chosen. The coherent escalation/de-escalation principles were applied with respect to the partial orderings (9) and (10). Parameters of safety (7) and futility (8) constraints were tuned over all scenarios and $k = 0.005$, $\xi_{final} = 0.9$, $\gamma' = 0.25$, $\zeta = 0.3$ were fixed for this simulation study.

The optimal non-parametric benchmark design[30;31] was included to get an

optimal design reference.

For each scenario, 10 000 simulations were run using R[32]. The R code underlying the simulations is available on GitHub at (https://github.com/dose-finding/comparison-non-mono).

Operating characteristics

Operating characteristics for all considered designs are given in Tables 1-4. They present the proportion of each dose selections with proportion of correct selections in bold, the average number of patients assigned to each dose level and several safety metrics: the proportion of early stopping due to safety (Stop), the probability of selecting toxic doses as the MTD (SelTox) and the average proportion of patients assigned to toxic doses above the MTD (%Tox). Note that for scenarios where no dose levels met the target toxicity probability, given in Tables 3 - 4, the dose level with true toxicity rate just below the target was considered correct.

We start from considering the proportion of correct selections in scenarios where the monotonicity assumption holds. The benefits of using a parametric model were shown in monotonic scenarios as the Continual Reassessment Method had the highest proportion of correct selections in scenario 2 (53.7%), scenario 3.1 (52%); the Toxicity Probability Interval method had the highest proportion of correct selections in scenario 1 (65.4%) and the Bayesian Optimal Interval design had the highest proportion of correct selections in scenario 4 (55.8%). The Bayesian Optimal Interval design also showed the

highest proportion of correct selections in scenarios where no dose level exactly matches the target toxicity probability: scenario 7.1, 8.1 and 9.1 with 73.1%, 61.1% and 60.8% respectively. The designs relaxing the monotonicity assumption, No Monotonicity Assumption and Partial Ordering Continual Reassessment Method, can lead to lower proportion of correct selections in the monotonic scenarios. For instance, in scenario 3.1, No Monotonicity Assumption and Partial Ordering Continual Reassessment Method had 40.0% and 40.7% proportion of correct selections, nearly a 13% decrease compared to Continual Reassessment Method. The differences were larger in scenarios where no dose level exactly matches the target toxicity level, for example, scenario 9.1 where the No Monotonicity Assumption design had a lower proportion of correct selections by 21% compared to the Bayesian Optimal Interval design.

Considering the proportion of correct selections in scenarios where the monotonicity assumption was violated, the No Monotonicity Assumption design showed the highest proportion of correct selections in most scenarios with, for instance, 65% and 53.9% in scenarios 7.2 and 8.3. The Partial Ordering Continual Reassessment Method showed the second highest proportion of correct selections, after the No Monotonicity Assumption design, in all non-monotonic scenarios with an exception in scenario 3.2, in which Partial Ordering Continual Reassessment Method had the highest proportion of correct selections of 44.7% against 38.5% for the No Monotonicity Assumption design. The violation of the monotonicity

assumption impacted the proportion of correct selections of the Bayesian Optimal Interval design, Toxicity Probability Interval method and Continual Reassessment Method design with proportions below 20% in the majority of non-monotonic scenarios. In scenario 3.2, the Continual Reassessment Method and Bayesian Optimal Interval design had 33.5% and 27.0% proportion of correct selections, respectively, but they were still below the No Monotonicity Assumption design and the Partial Ordering Continual Reassessment Method, which had respectively 38.5% and 44.7%.

Investigating safety, the Continual Reassessment Method, the Partial Ordering Continual Reassessment Method, the Bayesian Optimal Interval design and the Toxicity Probability Interval method were able to terminate a trial earlier for safety requirements in the highly toxic scenario 6 in nearly 90% of trials. The No Monotonicity Assumption design allowed early terminations in 86.7% of simulated trials in scenario 6. The proportion of early stopping for safety in scenario 1 exceeded 20% for the Bayesian Optimal Interval design, the Continual Reassessment Method and the Partial Ordering Continual Reassessment Method designs. In all other scenarios, all designs had a probability of stopping close to 0. Considering the probability of selecting a toxic dose in monotonic scenarios, the Toxicity Probability Interval method had the lowest probability in all scenarios.

In monotonic scenarios, the Partial Ordering Continual Reassessment Method had a higher probability to select toxic dose in scenarios 8.1 and 9.1 while

performing similarly to No Monotonicity Assumption design under the rest of monotonic scenarios. Considering the probability of selecting a toxic dose in non-monotonic scenarios, the Toxicity Probability Interval method was still the safest design. The Continual Reassessment Method showed the highest probability of selecting a toxic dose in all non-monotonic scenarios with, for instance, 52.5% in scenario 7.2. Comparing designs relaxing the monotonicity assumption, the No Monotonicity Assumption design had the same or lower probability of selecting a toxic dose than the Partial Ordering Continual Reassessment Method under twelve scenarios, with a reduction of up to 6.5% in scenario 6. Under scenarios 3.2, 6 and 7.1, the No Monotonicity Assumption design resulted in higher proportions of toxic dose recommendations than the Partial Ordering Continual Reassessment Method with differences between 1.4% to 5.7%.

Concerning the percentage of patients assigned to toxic dose levels under monotonic scenarios, the designs relaxing monotonicity, Partial Ordering Continual Reassessment Method and No Monotonicity Assumption design, resulted in more patients receiving toxic doses compared to the rest of the designs, especially in scenario 2 with 35.6% and 41.5% of patients assigned to toxic dose levels, respectively compared to 26% for the Toxicity Probability Interval method corresponding to the least proportion of patients allocated to toxic doses. Under non-monotonic scenarios however, all designs showed comparable percentage of assignment to toxic doses except for the conservative Toxicity Probability Interval method. Comparing the designs

relaxing monotonicity, the No Monotonicity Assumption design assigned more patients to toxic doses than the Partial Ordering Continual Reassessment Method under scenarios 3.2 and 8.2 where the dose level after the MTD drop below the target DLT rate. The largest difference was in scenario 3.2 where the Partial Ordering Continual Reassessment Method assigned 15.4% versus 25.3% patients on toxic doses for the No Monotonicity Assumption design. Both designs had comparable assignment to toxic doses in other non-monotonic scenarios.

Discussion and recommendations

In this work, we have considered the setting, in which neither single agent nor the majority of combination dose-finding methods can be directly applied. This motivated the consideration of more flexible designs relaxing the monotonicity assumption, namely, Partial Ordering Continual Reassessment Method and No Monotonicity Assumption design that are unique in their ability to handle partially ordered combinations that do not necessarily form a grid of combinations. Both model-based and model-assisted designs based on monotonicity assumption failed to identify the MTD consistently when this assumption was violated. It was found that the ordering can have a large impact on a phase I study and the monotonicity assumption should be carefully justified during the planning of a phase I trial with a combination of treatments. The designs relaxing the monotonicity assumption led to robust

performances in all permutations of orderings. On average across all considered scenarios, the No Monotonicity Assumption design selected less often toxic doses than the Partial Ordering Continual Reassessment Method but the Partial Ordering Continual Reassessment Method allocated on average less patients to toxic doses.

The motivating trial assesses toxicity of a combination of a cytotoxic chemotherapy with an immunotherapy. The assumption that higher dose leads to higher efficacy and toxicity holds, but for dose levels 3, 4 and 5. Yet, one may have interest in assessing both toxicity and efficacy as in Wages and Tait[33]. In the setting of our motivating trial the proposed endpoint for efficacy would have been response to treatment. This requires tumour response evaluation after nearly 85 days of treatment, which was not feasible, as efficacy outcomes couldn't be obtained soon enough to make adaptive decision for a new cohort given an accrual rate of 4 patients per month. The problem of late onset outcome in early phase clinical trials was discussed by Cheung and Chappell[34], and later by Braun[35] and Liu and Ning[36]. An application can be found in Ick et al.[37].

If a monotonic order cannot be specified, two alternatives, the Partial Ordering Continual Reassessment Method and the No Monotonicity Assumption design, should be considered. Regardless which design is chosen, we advocate that the first (and essential) step is to restrict the number of orderings in each particular setting. Wages et al.[38] showed that the

performance of the Partial Ordering Continual Reassessment Method decreases as the number of orderings increases. Also, it has been demonstrated by Mozgunov and Jaki[17] that No Monotonicity Assumption design accuracy could also decrease if the number of potential dose-toxicity relations for calibration is large.

The choice of the orderings should be made after an extensive discussion with experts whatever the design. The first point to discuss is the validity of the monotonicity assumption. Then we recommend that the choice of the design should be guided by a comprehensive simulation study taking into account the preference for each scenario. For instance, if an investigator believes that a scenario with the MTD being the last dose is the most probable in a non-monotonic relationship, the Partial Ordering Continual Reassessment Method should be considered as a primary candidate. On the other hand, if a scenario with the MTD located in the middle of the dose range about which the monotonicity assumption can be violated, the No Monotonicity Assumption design should be prioritized. The No Monotonicity Assumption design should also be prioritized if interaction that can lead to large difference in toxicity can occur as in non monotonic scenarios 7.2 and 9.2. Depending on the scenario, small changes in the set of parameters used for the Continual Reassessment Method and the Partial Ordering Continual Reassessment Method can impact the operating characteristics of the design. As investigators usually do not have data prior the trial to guess the most likely scenario, simulations should be conducted with great care with respect to the

choice of the initials parameters and the scenarios. Importantly, as stated by Mozgunov and Jaki[23] and shown above, relaxing of monotonicity assumption might result in higher number of DLTs for both No Monotonicity Assumption design and Partial Ordering Continual Reassessment Method. It can be explained by the fact that they require to investigate more dose levels before selecting the correct ordering. This cost of considering different orderings should be taking into account when planning a study. The parameters of safety constraints should be extensively discussed with the experts to avoid the exposure of patients to excessively toxic dose levels.

References

[1] Storer BE. Design and analysis of phase i clinical trials. *Biometrics* 1989; 45(3): 925–937.

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

[3] Ji Y, Liu P, Li Y et al. A modified toxicity probability interval method for dose-finding trials. *Clinical Trials* 2010; 7(6): 653–663.

[4] Liu S and Yuan Y. Bayesian optimal interval designs for phase i clinical trials. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 2015; 64(3): 507–523.

[5] Skipper HE. Implications of biochemical, cytokinetic, pharmacologic, and

toxicologic, relationships in the design of optimal therapeutic schedules. *Cancer Chemother Rep* 1970; 54: 431–450.

[6] Pinto NR, Applebaum MA, Volchenboum SL et al. Advances in risk classification and treatment strategies for neuroblastoma. *Journal of clinical oncology* 2015; 33(27): 3008.

[7] Reynolds AR. Potential relevance of bell-shaped and u-shaped dose-responses for the therapeutic targeting of angiogenesis in cancer. *Dose-response* 2010; 8(3): dose–response.

[8] Steliarova-Foucher E, Colombet M, Ries LA et al. International incidence of childhood cancer, 2001–10: a population-based registry study. *The Lancet Oncology* 2017; 18(6): 719–731.

[9] Mujoo K, Cheresh DA, Yang HM et al. Disialoganglioside gd2 on human neuroblastoma cells: target antigen for monoclonal antibody-mediated cytolysis and suppression of tumor growth. *Cancer Research* 1987; 47(4): 1098–1104.

[10] Mueller BM, Romerdahl C, Gillies S et al. Enhancement of antibody-dependent cytotoxicity with a chimeric anti-gd2 antibody. *The Journal of Immunology* 1990; 144(4): 1382–1386.

[11] Zeng Y, Fest S, Kunert R et al. Anti-neuroblastoma effect of ch14. 18 antibody produced in cho cells is mediated by nk-cells in mice. *Molecular immunology* 2005; 42(11): 1311–1319.

[12] Riviere MK, Yuan Y, Dubois F et al. A bayesian dose finding design for clinical trials combining a cytotoxic agent with a molecularly targeted agent. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 2015; 64(1): 215–229.

[13] Braun TM, Yuan Z and Thall PF. Determining a maximum-tolerated schedule of a cytotoxic agent. *Biometrics* 2005; 61(2): 335–343. DOI: 10.1111/j.1541-0420.2005.00312.x.

[14] Thall PF, Nguyen HQ, Braun TM et al. Using joint utilities of the times to response and toxicity to adaptively optimize schedule–dose regimes. *Biometrics* 2013; 69(3): 673–682.

[15] Guo B, Li Y and Yuan Y. A dose–schedule finding design for phase i–ii clinical trials. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 2016; 65(2): 259–272.

[16] Wages NA, Conaway MR and O'Quigley J. Continual reassessment method for partial ordering. *Biometrics* 2011; 67(4): 1555–1563.

[17] Mozgunov P and Jaki T. An information-theoretic approach for selecting arms in clinical trials. *Preprint arXiv:170802426 Department of Mathematics and Statistics, Lancaster University, Lancaster* 2017; .

[18] Iasonos A, Wilton AS, Riedel ER et al. A comprehensive comparison of the continual reassessment method to the standard 3+ 3 dose escalation scheme in phase i dose-finding studies. *Clinical Trials* 2008; 5(5): 465–477.

[19] Onar-Thomas A and Xiong Z. A simulation-based comparison of the traditional method, rolling-6 design and a frequentist version of the continual reassessment method with special attention to trial duration in pediatric phase i oncology trials. *Contemporary Clinical Trials* 2010; 31(3): 259 – 270. DOI:https://doi.org/10.1016/j.cct.2010.03.006. URL `http://www.sciencedirect.com/science/article/pii/S1551714410000364`.

[20] Doussau A, Asselain B, Deley ML et al. Dose-finding designs in pediatric phase i clinical trials: Comparison by simulations in a realistic timeline framework. *Contemporary Clinical Trials* 2012; 33(4): 657 – 665. DOI:https://doi.org/10.1016/j.cct.2011.11.015. URL `http://www.sciencedirect.com/science/article/pii/S1551714411002886`.

[21] Ji Y and Wang SJ. Modified toxicity probability interval design: a safer and more reliable method than the 3+ 3 design for practical phase i trials. *Journal of Clinical Oncology* 2013; 31(14): 1785–1791.

[22] Conaway MR, Dunbar S and Peddada SD. Designs for single-or multiple-agent phase i trials. *Biometrics* 2004; 60(3): 661–669.

[23] Mozgunov P and Jaki T. An information theoretic phase i–ii design for molecularly targeted agents that does not require an assumption of monotonicity. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 2019; 68(2): 347–367.

[24] Cheung YK. Coherence principles in dose-finding studies. *Biometrika* 2005; : 863–873.

[25] Ladenstein R, Pötschger U, Hartman O et al. 28 years of high-dose therapy and sct for neuroblastoma in europe: lessons from more than 4000 procedures. *Bone marrow transplantation* 2008; 41: S118–S127.

[26] Pearson AD, Pinkerton CR, Lewis IJ et al. High-dose rapid and standard induction chemotherapy for patients aged over 1 year with stage 4 neuroblastoma: a randomised trial. *The lancet oncology* 2008; 9(3): 247–256.

[27] Cheung K. *dfcrm: Dose-Finding by the Continual Reassessment Method*, 2019. URL `https://CRAN.R-project.org/package=dfcrm`. R package version 0.2-2.1.

[28] Cheung YK. *Dose finding by the continual reassessment method*. CRC Press, 2011.

[29] Zhou H, Yuan Y and Nie L. Accuracy, safety, and reliability of novel phase i trial designs. *Clinical Cancer Research* 2018; 24(18): 4357–4364. DOI:10.1158/1078-0432.CCR-18-0168. URL `https://clincancerres.aacrjournals.org/content/24/18/4357`. `https://clincancerres.aacrjournals.org/content/24/18/4357.full.pdf`.

[30] O'Quigley J, Paoletti X and Maccario J. Non-parametric optimal design in dose finding studies. *Biostatistics* 2002; 3(1): 51–56.

[31] Mozgunov P, Jaki T and Paoletti X. A benchmark for dose finding studies with continuous outcomes. *Biostatistics* 2018; : 10.1093/biostatistics/kxy045.

[32] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2015. URL `https://www.R-project.org/`.

[33] Wages NA and Tait C. Seamless phase i/ii adaptive design for oncology trials of molecularly targeted agents. *Journal of Biopharmaceutical Statistics* 2015; 25(5): 903–920.

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

[35] Braun TM. Generalizing the tite-crm to adapt for early-and late-onset toxicities. *Statistics in Medicine* 2006; 25(12): 2071–2083.

[36] Liu S and Ning J. A bayesian dose-finding design for drug combination trials with delayed toxicities. *Bayesian analysis* 2013; 8(3): 703.

[37] Ick HJ, Liu S, Thall PF et al. Using data augmentation to facilitate conduct of phase i–ii clinical trials with delayed outcomes. *Journal of the American Statistical Association* 2014; 109(506): 525–536. DOI:10.1080/01621459.2014.881740. URL `https://doi.org/10.1080/01621459.2014.881740`. `https://doi.org/10.1080/01621459.2014.881740`.

[38] Wages NA, O'Quigley J and Conaway MR. Phase i design for completely or partially ordered treatment schedules. *Statistics in medicine* 2014; 33(4): 569–579.

Table 1: Operating characteristics in scenarios 1-3.2: proportions of selections at each dose level and respective average number of patients assigned (in brackets). The proportion of the true MTD selection is in bold. The proportion of early stopping for safety is given in column "Stop", the probability of selecting toxic doses as the MTD in "SelTox" and the percentage of patients assigned to the toxic doses in "%Tox". Results are based on 10^4 replications. CRM: Continual Reassessment Method; PO-CRM: Partial Ordering Continual Reassessment Method; mTPI: modified Toxicity Probability Interval design; BOIN: Bayesian Optimal Interval design; NMA: No Monotonicity Assumption design.

	d_1	d_2	d_3	d_4	d_5	d_6	Stop	SelTox	%Tox
Scenario 1	**0.30**	0.40	0.50	0.60	0.90	0.90			
Benchmark	**76.4**	20.5	0.3	0.1	0.0	0.0			
CRM	**48.5**	25.7	3.9	0.2	0.0	0.0	21.7	29.7	52.3
	(12.5)	(10.4)	(2.9)	(0.4)	(0.0)	(0.0)			
PO-CRM	**49.3**	25.0	3.6	1.6	0.0	0.0	20.4	30.3	53.6
	(12.2)	(10.1)	(2.8)	(1.2)	(0.1)	(0.0)			
mTPI	**65.4**	9.9	5.1	0.2	0.0	0.0	19.5	15.2	63.3
	(9.8)	(13.7)	(2.9)	(0.3)	(0.0)	(0.0)			
BOIN	**40.7**	29.1	5.1	0.3	0.0	0.0	24.9	34.4	60.6
	(10.3)	(12.0)	(3.3)	(0.4)	(0.0)	(0.0)			
NMA	**54.8**	23.6	4.9	2.1	0.1	0.0	14.5	30.7	52.5
	(13.4)	(8.8)	(2.3)	(2.2)	(1.5)	(0.0)			
Scenario 2	0.14	**0.30**	0.40	0.50	0.70	0.75			
Benchmark	14.1	**62.3**	20.6	3.0	0.0	0.0			
CRM	15.0	**53.7**	24.8	3.0	0.1	0.0	3.4	27.9	30.9
	(6.1)	(14.1)	(7.1)	(1.6)	(0.2)	(0.0)			
PO-CRM	14.1	**50.7**	22.4	8.7	0.9	0.0	3.2	32.0	35.6
	(5.7)	(13.1)	(6.8)	(3.0)	(0.5)	(0.0)			
mTPI	50.1	**26.0**	18.8	3.4	0.1	0.0	1.7	22.3	26.1
	(6.8)	(15.1)	(6.2)	(1.4)	(0.1)	(0.0)			
BOIN	18.7	**52.8**	21.5	4.3	0.1	0.0	2.5	26.0	31.1
	(5.6)	(14.7)	(7.1)	(1.8)	(0.2)	(0.0)			
NMA	26.1	**41.1**	18.7	11.0	2.7	0.0	0.4	32.4	41.5
	(6.7)	(10.9)	(4.6)	(4.7)	(3.1)	(0.1)			
Scenario 3.1	0.05	0.10	0.20	**0.30**	0.45	0.70			
Benchmark	0.0	1.4	27.9	**55.7**	15.1	0.0			
CRM	0.0	2.4	27.6	**52.8**	16.9	0.2	0.1	17.1	16.1
	(0.3)	(5.4)	(9.6)	(9.8)	(4.3)	(0.6)			
PO-CRM	0.0	2.4	33.9	**40.7**	22.4	0.7	0.0	23.0	19.1
	(0.3)	(5.0)	(10.0)	(9.0)	(5.2)	(0.6)			
mTPI	2.7	12.4	33.0	**38.4**	13.1	0.4	0.0	13.5	12.0
	(0.7)	(7.0)	(10.2)	(8.5)	(3.2)	(0.4)			
BOIN	0.3	4.4	29.9	**47.0**	17.8	0.5	0.0	18.4	15.6
	(0.3)	(5.8)	(9.8)	(9.4)	(4.1)	(0.5)			
NMA	2.5	7.1	27.5	**40.0**	21.5	1.3	0.0	22.8	26.3
	(1.7)	(5.5)	(6.3)	(8.6)	(6.7)	(1.2)			
Scenario 3.2	0.05	0.10	**0.30**	0.20	0.45	0.70			
CRM	0.0	8.1	**33.5**	35.7	22.3	0.3	0.1	22.6	15.9
	(0.5)	(7.2)	(10.7)	(6.8)	(4.3)	(0.5)			
PO-CRM	0.0	3.2	**44.7**	34.1	17.3	0.5	0.1	17.8	15.4
	(0.4)	(5.3)	(11.6)	(8.0)	(4.1)	(0.5)			
mTPI	2.8	28.4	**23.7**	31.6	13.0	0.4	0.0	13.4	12.1
	(0.7)	(10.1)	(9.3)	(6.2)	(3.3)	(0.4)			
BOIN	0.3	17.3	**27.0**	33.8	21.0	0.6	0.0	21.5	16.2
	(0.3)	(8.8)	(9.6)	(6.5)	(4.3)	(0.6)			
NMA	2.7	7.6	**38.5**	32.0	17.5	1.7	0.0	19.2	25.3
	(2.0)	(5.9)	(7.4)	(7.1)	(6.3)	(1.3)			

Table 2: Operating characteristics in scenarios 3.3-6: proportions of selections at each dose level and respective average number of patients assigned (in brackets). The proportion of the true MTD selection is in bold. The proportion of early stopping for safety is given in column "Stop", the probability of selecting toxic doses as the MTD in "SelTox" and the percentage of patients assigned to the toxic doses in "%Tox". Results are based on 10^4 replications. CRM: Continual Reassessment Method; PO-CRM: Partial Ordering Continual Reassessment Method; mTPI: modified Toxicity Probability Interval design; BOIN: Bayesian Optimal Interval design; NMA: No Monotonicity Assumption design.

	d_1	d_2	d_3	d_4	d_5	d_6	Stop	SelTox	%Tox
Scenario 3.3	0.05	0.10	0.20	0.45	**0.30**	0.70			
CRM	0.0	3.0	47.9	34.0	**13.5**	1.4	0.1	35.4	29.2
	(0.3)	(5.6)	(12.1)	(8.1)	(3.1)	(0.7)			
PO-CRM	0.0	3.3	32.4	27.8	**35.6**	0.9	0.1	28.6	28.4
	(0.3)	(5.3)	(10.0)	(8.0)	(5.8)	(0.5)			
mTPI	2.7	12.5	59.0	14.0	**11.4**	0.4	0.0	14.4	20.8
	(0.7)	(7.0)	(14.1)	(5.9)	(2.0)	(0.4)			
BOIN	0.3	4.5	59.4	20.6	**14.5**	0.7	0.0	21.3	26.6
	(0.3)	(5.9)	(13.6)	(7.4)	(2.3)	(0.6)			
NMA	2.7	8.0	24.2	23.7	**40.2**	1.2	0.0	24.9	30.9
	(1.6)	(5.5)	(5.9)	(7.6)	(7.8)	(1.7)			
Scenario 4	0.01	0.05	0.10	0.15	0.20	**0.30**			
Benchmark	0.0	0.0	0.8	6.5	24.0	**68.8**			
CRM	0.0	0.0	1.2	10.4	35.2	**53.2**	0.0	0.0	0.0
	(0.1)	(3.7)	(4.7)	(5.8)	(7.4)	(8.4)			
PO-CRM	0.0	0.0	2.8	15.2	27.6	**54.4**	0.0	0.0	0.0
	(0.1)	(3.5)	(5.0)	(6.5)	(6.6)	(8.4)			
mTPI	0.7	2.9	7.6	16.9	29.6	**42.3**	0.0	0.0	0.0
	(0.2)	(4.3)	(5.6)	(6.6)	(6.6)	(6.7)			
BOIN	0.1	0.3	2.2	11.0	30.6	**55.8**	0.0	0.0	0.0
	(0.0)	(3.7)	(4.9)	(6.2)	(7.0)	(8.1)			
NMA	0.1	1.0	7.5	17.4	23.5	**50.5**	0.0	0.0	0.0
	(0.5)	(4.0)	(3.8)	(6.2)	(7.3)	(8.2)			
Scenario 5	0.001	0.01	0.05	0.07	0.10	**0.15**			
Benchmark	0.8	0.3	3.2	4.0	13.6	78.2			
CRM	0.0	0.0	0.0	0.3	4.6	**95.0**	0.0	0.0	0.0
	(0.0)	(3.1)	(3.5)	(3.5)	(4.5)	(15.4)			
PO-CRM	0.0	0.0	0.1	1.1	3.3	**95.4**	0.0	0.0	0.0
	(0.0)	(3.1)	(3.6)	(4.1)	(4.1)	(15.2)			
mTPI	0.0	0.8	1.5	3.6	9.2	**84.9**	0.0	0.0	0.0
	(0.0)	(3.3)	(3.8)	(4.3)	(5.0)	(13.6)			
BOIN	0.0	0.0	0.1	0.8	4.5	**94.6**	0.0	0.0	0.0
	(0.0)	(3.1)	(3.6)	(4.0)	(4.7)	(14.6)			
NMA	0.0	0.0	1.5	3.4	7.0	**88.1**	0.0	0.0	0.0
	(0.3)	(3.3)	(2.3)	(4.4)	(5.5)	(14.2)			
Scenario 6	0.50	0.70	0.80	0.90	0.95	0.95			
Benchmark	99.9	0.1	0.00	0.00	0.00	0.00			
CRM	7.0	0.0	0.0	0.0	0.0	0.0	93.0	7.0	100.0
	(7.6)	(3.9)	(0.1)	(0.0)	(0.0)	(0.0)			
PO-CRM	7.5	0.1	0.0	0.0	0.0	0.0	92.4	7.6	100.0
	(7.7)	(3.9)	(0.1)	(0.0)	(0.0)	(0.0)			
mTPI	9.9	0.0	0.0	0.0	0.0	0.0	90.1	9.9	100.0
	(9.2)	(4.4)	(0.1)	(0.0)	(0.0)	(0.0)			
BOIN	9.4	0.0	0.0	0.0	0	0.0	90.6	9.4	100.0
	(9.1)	(4.3)	(0.1)	(0.0)	(0.0)	(0.0)			
NMA	13.1	0.2	0.0	0.0	0.0	0.0	86.7	13.3	100.0
	(10.5)	(2.9)	(0.1)	(0.1)	(0.1)	(0.0)			

Table 3: Operating characteristics in scenarios 7-8.2: proportions of selections at each dose level and respective average number of patients assigned (in brackets). The proportion of correct MTD selection is in bold. The proportion of early stopping for safety is given in column "Stop", the probability of selecting toxic doses as the MTD in "SelTox" and the percentage of patients assigned to the toxic doses in "%Tox". Results are based on 10^4 replications. CRM: Continual Reassessment Method; PO-CRM: Partial Ordering Continual Reassessment Method; mTPI: modified Toxicity Probability Interval design; BOIN: Bayesian Optimal Interval design; NMA: No Monotonicity Assumption design.

	d_1	d_2	d_3	d_4	d_5	d_6	Stop	SelTox	%Tox
Scenario 7.1	0.01	0.05	**0.20**	0.50	0.60	0.70			
Benchmark	0.0	0.2	**77.8**	21.6	0.4	0.0			
CRM	0.0	2.1	**61.9**	35.0	1.0	0.0	0.0	36.0	35.6
	(0.1)	(4.8)	(14.4)	(9.1)	(1.5)	(0.1)			
PO-CRM	0.0	2.6	**58.2**	26.7	12.4	0.0	0.0	39.1	39.4
	(0.1)	(4.8)	(13.3)	(8.0)	(3.7)	(0.1)			
mTPI	0.7	11.9	**71.4**	15.0	1.1	0.0	0.0	16.1	24.0
	(0.2)	(6.2)	(16.4)	(6.4)	(0.8)	(0.0)			
BOIN	0.1	3.9	**73.1**	21.6	1.3	0.0	0.0	22.9	29.9
	(0.0)	(5.1)	(15.9)	(8.0)	(0.9)	(0.1)			
NMA	0.7	4.8	**52.9**	23.9	17.2	0.4	0.0	41.5	48.0
	(1.9)	(5.5)	(8.2)	(8.6)	(5.5)	(0.3)			
Scenario 7.2	0.01	0.05	0.50	**0.20**	0.60	0.70			
CRM	0.0	33.2	47.5	**14.3**	5.0	0.0	0.0	52.5	49.5
	(0.8)	(11.4)	(13.5)	(2.9)	(1.3)	(0.1)			
PO-CRM	0.0	7.2	35.8	**55.4**	1.6	0.0	0.0	37.4	43.0
	(0.4)	(6.3)	(11.7)	(10.4)	(1.1)	(0.1)			
mTPI	0.7	76.1	8.2	**14.0**	1.0	0.0	0.0	9.2	26.8
	(0.2)	(19.1)	(7.3)	(2.7)	(0.7)	(0.0)			
BOIN	0.1	63.9	19.2	**14.7**	2.1	0.1	0.0	21.3	34.5
	(0.0)	(17.0)	(9.2)	(2.6)	(1.1)	(0.1)			
NMA	0.6	3.5	25.9	**65.0**	3.9	1.0	0.0	30.9	43.5
	(2.5)	(6.2)	(8.0)	(8.3)	(4.5)	(0.6)			
Scenario 8.1	0.01	0.05	0.10	**0.25**	0.45	0.60			
Benchmark	0.0	0.0	2.8	**72.8**	24.1	0.3			
CRM	0.0	0.0	6.8	**61.0**	30.6	1.4	0.0	32.1	28.5
	(0.1)	(3.7)	(6.0)	(11.7)	(7.2)	(1.4)			
PO-CRM	0.0	0.0	17.7	**47.6**	32.8	1.9	0.0	34.7	29.6
	(0.1)	(3.6)	(7.1)	(10.3)	(7.5)	(1.4)			
mTPI	0.7	2.9	19.7	**55.2**	20.1	1.4	0.0	21.5	20.1
	(0.2)	(4.3)	(7.7)	(11.9)	(5.3)	(0.8)			
BOIN	0.1	0.3	10.1	**61.1**	26.4	2.0	0.0	28.4	24.6
	(0.0)	(3.7)	(6.7)	(12.1)	(6.3)	(1.0)			
NMA	0.5	2.4	20.3	**46.5**	25.9	4.3	0.0	30.2	31.6
	(1.3)	(4.8)	(5.1)	(9.4)	(7.6)	(1.9)			
Scenario 8.2	0.01	0.05	**0.25**	0.10	0.45	0.60			
CRM	0.0	1.8	**16.2**	36.6	43.9	1.4	0.0	45.4	28.8
	(0.1)	(5.0)	(8.6)	(7.6)	(7.4)	(1.2)			
PO-CRM	0.0	0.2	**44.2**	22.0	31.5	2.0	0.0	33.6	26.1
	(0.1)	(3.8)	(11.3)	(7.0)	(6.5)	(1.3)			
mTPI	0.7	19.2	**18.2**	40.1	20.3	1.6	0.0	21.9	19.9
	(0.2)	(7.8)	(8.5)	(7.6)	(5.2)	(0.8)			
BOIN	0.1	7.9	**17.4**	38.9	33.6	2.1	0.0	35.7	25.4
	(0.0)	(6.3)	(8.3)	(7.8)	(6.6)	(1.0)			
NMA	0.6	2.1	**48.0**	21.0	23.8	4.5	0.0	28.2	31.3
	(1.5)	(5.1)	(8.3)	(5.7)	(7.3)	(2.1)			

Table 4: Operating characteristics in scenarios 8.3-9: proportions of selections at each dose level and respective average number of patients assigned (in brackets). The proportion of correct MTD selection is in bold. The proportion of early stopping for safety is given in column "Stop", the probability of selecting toxic doses as the MTD in "SelTox" and the percentage of patients assigned to the toxic doses in "%Tox". Results are based on 10^4 replications. CRM: Continual Reassessment Method; PO-CRM: Partial Ordering Continual Reassessment Method; mTPI: modified Toxicity Probability Interval design; BOIN: Bayesian Optimal Interval design; NMA: No Monotonicity Assumption design.

	d_1	d_2	d_3	d_4	d_5	d_6	Stop	SelTox	%.Tox
Scenario 8.3	0.01	0.05	0.10	0.45	**0.25**	0.60			
CRM	0.0	0.1	28.5	43.1	**22.4**	5.9	0.0	49.0	39.2
	(0.1)	(3.8)	(9.4)	(10.1)	(4.9)	(1.7)			
PO-CRM	0.0	0.1	11.9	35.1	**50.3**	2.5	0.0	37.6	36.8
	(0.1)	(3.7)	(7.0)	(9.9)	(8.2)	(1.2)			
mTPI	0.7	2.9	62.6	15.9	**16.4**	1.5	0.0	17.3	27.7
	(0.2)	(4.3)	(14.5)	(7.6)	(2.8)	(0.8)			
BOIN	0.1	0.3	51.8	26.7	**18.1**	3.0	0.0	29.7	33.9
	(0.0)	(3.7)	(13.2)	(9.0)	(2.8)	(1.2)			
NMA	0.5	2.8	10.0	28.7	**53.9**	4.1	0.0	32.9	37.3
	(1.1)	(4.6)	(4.5)	(8.6)	(8.6)	(2.6)			
Scenario 9.1	0.01	0.05	0.08	0.10	**0.25**	0.45			
Benchmark	0.0	0.0	0.5	2.3	**73.0**	24.2			
CRM	0.0	0.0	0.3	9.6	**60.5**	29.5	0.0	29.5	21.4
	(0.1)	(3.6)	(4.2)	(5.5)	(10.2)	(6.4)			
PO-CRM	0.0	0.0	3.8	15.8	**49.8**	30.6	0.0	30.6	22.3
	(0.1)	(3.5)	(4.8)	(6.1)	(8.8)	(6.7)			
mTPI	0.7	1.9	3.7	20.6	**52.2**	20.9	0.0	20.9	16.1
	(0.2)	(4.0)	(4.6)	(6.8)	(9.6)	(4.8)			
BOIN	0.1	0.2	0.7	11.1	**60.8**	27.2	0.0	27.2	19.7
	(0.0)	(3.7)	(4.1)	(6.2)	(10.1)	(5.9)			
NMA	0.2	1.5	9.0	22.9	**40.1**	26.4	0.0	26.4	20.4
	(0.7)	(4.2)	(4.2)	(5.7)	(9.0)	(6.1)			
Scenario 9.2	0.01	0.05	0.08	**0.25**	0.10	0.45			
CRM	0.0	0.0	3.0	**18.1**	36.3	42.6	0.0	42.6	24.3
	(0.1)	(3.6)	(5.0)	(6.9)	(7.2)	(7.3)			
PO-CRM	0.0	0.0	1.2	**41.7**	24.5	32.6	0.0	32.6	20.9
	(0.0)	(3.5)	(4.5)	(9.6)	(6.0)	(6.3)			
mTPI	0.7	1.9	18.7	**18.2**	38.3	22.1	0.0	22.1	16.3
	(0.2)	(4.0)	(7.3)	(7.4)	(6.2)	(4.9)			
BOIN	0.1	0.2	8.5	**19.6**	37.6	34.0	0.0	34.0	20.3
	(0.0)	(3.7)	(6.2)	(7.7)	(6.3)	(6.1)			
NMA	0.2	1.3	5.3	**48.6**	16.3	28.3	0.0	28.3	21.3
	(0.6)	(4.1)	(3.9)	(9.1)	(5.9)	(6.4)			