Original research article

Joint time-to-event partial order continual reassessment method and Joint time-to-event Bayesian logistic regression model: Statistical designs for dual agent phase I/II dose finding studies with late-onset toxicity and activity outcomes

Helen Barnett[1] (iD)**, Oliver Boix[2], Dimitris Kontos[2] and Thomas Jaki[3,4]** (iD)

Abstract

Dual agent dose-finding trials study the effect of a combination of more than one agent, where the objective is to find the Maximum Tolerated Dose Combination, the combination of doses of the two agents that is associated with a pre-specified risk of being unsafe. In a Phase I/II setting, the objective is to find a dose combination that is both safe and active, the Optimal Biological Dose, that optimises a criterion based on both safety and activity. Since Oncology treatments are typically given over multiple cycles, both the safety and activity outcome can be considered as late-onset, potentially occurring in the later cycles of treatment. This work proposes two model-based designs for dual-agent dose finding studies with late-onset activity and late-onset toxicity outcomes, the Joint time-to-event (TITE) partial order continual reassessment method and the Joint TITE Bayesian logistic regression model. Their performance is compared alongside a model-assisted comparator in a comprehensive simulation study motivated by a real trial example, with an extension to consider alternative sized dosing grids. It is found that both model-based methods outperform the model-assisted design. Whilst on average the two model-based designs are comparable, this comparability is not consistent across scenarios.

Keywords

dose-finding, late-onset toxicities, late-onset activity, dual agent, phase I trials, model-based

1 Introduction

A traditional Phase I dose-finding study uses binary observations of dose-limiting toxicities (DLTs), evaluating the safety of a novel agent by escalating dose assignment in order to find the Maximum Tolerated Dose (MTD), a dose level associated with a pre-specified probability of observing a DLT. It may however be of interest to study the effects of a combination of more than one agent, for example for synergistic effects of activity or safety. Here the objective is to find the MTD Combination, the combination of doses of each of the agents associated with the pre-specified probability of observing a DLT. Multiple designs exist for the dual agent Phase I dose-finding trial, for example the Bayesian Optimal Interval design[1]

[1]Lancaster University, Lancaster, UK
[2]Bayer AG, Leverkusen, Germany
[3]MRC Biostatistics Unit, University of Cambridge, Cambridge, UK
[4]University of Regensburg, Regensburg, Germany

Corresponding author:
Helen Barnett, Lancaster University, Lancaster, UK.
Email: helen.barnett@lancaster.ac.uk

(BOIN), Keyboard,[2] Partial Order Continual Reassessment Method[3] (POCRM), Waterfall[4] and Bayesian Logistic Regression Model[5] (BLRM), with a comprehensive comparison of model-assisted designs compared to model-based designs recently undertaken.[6]

It is often the case in oncology trials that there are multiple cycles of treatment, and therefore the trial duration would be infeasibly long if one had to wait until the entire follow-up period for the previous cohort had been completed before assigning the next cohort to a dose. Therefore, it is useful for a design to be able to incorporate delayed onset outcomes when only partial information is available, for example on the first cycle out of a three-cycle treatment programme. Few designs exist that incorporate delayed onset outcomes in dual agent trials, one such example being the time-to-event (TITE) version of the POCRM.[7]

In the cases where it is possible to collect measures of activity of the agents, as is often the case in oncology, seamless Phase I/II designs are possible since activity acts as a surrogate for efficacy. Here, the objective is to find the Optimal Biological Dose (OBD), a dose that is both safe and shows sufficient activity that optimises some utility criterion based on safety and activity, in line with the goal of the FDA's Project Optimus[8] 'to move forward with a dose-finding and dose optimisation paradigm across oncology that emphasises selection of a dose or doses that maximises not only the efficacy of a drug but the safety and tolerability as well.' A small number of designs exist that can evaluate both safety and activity with delayed-onset binary outcomes in the single agent setting, model-assisted approaches such as that from Liu et al.[9] the TITE-B,[10] and model-based approaches of the A_T/A_E Design[11] and the Joint TITE-CRM.[12] The inclusion of both safety and activity outcomes in dual agent trials with no consideration for late-onset outcomes has been used in various designs (see e.g. Wages and Conaway,[13] Shimamura et al.,[14] Yada and Hamada,[15] Zhang et al.,[16] Kakizume et al.,[17] Lu et al.[18] and Jaki et al.[19]). The model assisted TITE-comb-BOIN12[18] is currently the only design to address both late onset toxicity and late onset activity outcomes in the dual agent setting. However, there does not currently exist a model-based design that considers all of these factors.

In this work, we propose two model-based approaches to design a seamless Phase I/II dose finding trial in oncology, incorporating delayed-onset toxicity and delayed-onset activity outcomes for dual agents, hypothesising that the joint modelling approach will improve accuracy in the OBD identification. These novel model-based designs are compared to the model-assisted TITE-comb-BOIN12 design. A motivating example is given in Section 2, before Section 3 introduces the methodology for all three designs considered. Section 4 conducts a simulation study to compare the performance of the designs before concluding with a discussion in Section 5.

2 Motivating example

Combining several anticancer agents can enhance overall antitumour activity; however, it may also lead to increased toxicity. Combination therapies are frequently evaluated during the early clinical development of investigational oncology agents. This is particularly relevant for prostate cancer, which remains the most commonly diagnosed cancer in males and a leading cause of cancer-related deaths.

Target Radionuclide Therapy, which involves therapeutic agents based on radionuclides, holds significant promise in oncology. These new agents are also being investigated in combination with standard treatments for prostate cancer, such as hormone therapy. This approach is especially appealing and more easily planned when both agents are part of the internal portfolio, as preclinical data can be generated more readily, creating a commercially viable opportunity.

This exploration is driven by previous initiatives within Bayer's Targeted Alpha Therapy (TAT) platform. TAT is an emerging modality in the field of radiotherapy that combines tumour-targeting molecules with alpha particle-emitting radioisotopes, aiming to provide a novel approach to cancer treatment and potentially overcome resistance.[20]

In an early-phase study, a thorium-227 labelled antibody-chelator conjugate was combined with darolutamide, a synthetic nonsteroidal next-generation androgen receptor antagonist, in patients with metastatic castration-resistant prostate cancer.[21,22] In this context, the doses of both agents could be adjusted to identify the optimal dose combination that enhances overall antitumour activity while maintaining an acceptable safety profile.

3 Methodology

We propose two alternative approaches, both based on the Joint TITE-CRM,[12] a design for single agent therapies with late-onset activity and toxicity outcomes, which was shown to have good operating characteristics. The first proposed approach, the Joint TITE-POCRM, extends the Joint TITE-CRM to the dual agent setting by mapping the two-dimensional dosing grid into one dimension using the POCRM.[3] The second, the Joint TITE-BLRM, extends the Joint TITE-CRM to the dual agent setting by modelling the joint odds using the BLRM.[23] These are compared to the model assisted TITE-comb-BOIN12 design,[18] which brings together the TITE-BOIN12 design,[24] and the dual agent BOIN,[1] summarised in

Table 1. An example of dose labelling for the dual agent trial.

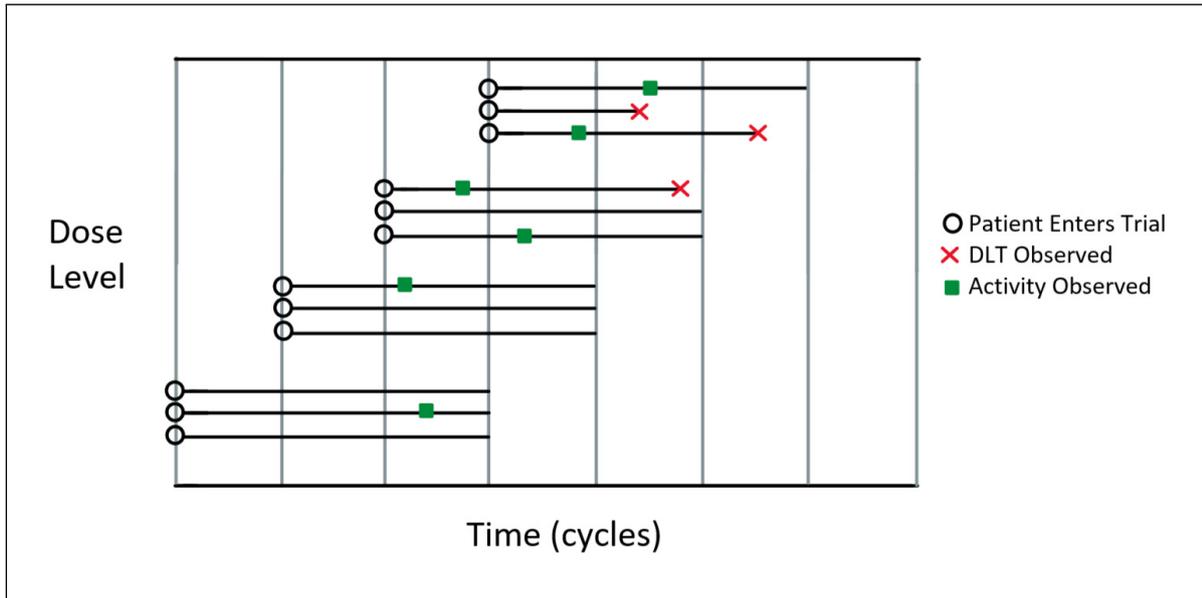|		Agent W_2				
		50 kBq/kg	75 kBq/kg	100 kBq/kg	125 kBq/kg	150 kBq/kg
Agent W_1	600 mg	d_{11}	d_{12}	d_{13}	d_{14}	d_{15}
	1200 mg	d_{21}	d_{22}	d_{23}	d_{24}	d_{25}

Figure 1. Visualisation of example trial with an observation period of three cycles and dose-escalation decisions taking place after one cycle.

this section, with further details given in the Supplemental Materials. Section 3.1 provides the trial setup that is applicable to all methods considered, then the following subsections introduce each of the approaches in turn. For brevity in this section, many equations that are identical in the cases of toxicity and activity are only given once, with relevant subscripts of T for toxicity that can be replaced with A in the case of activity.

3.1   Trial setup

3.1.1   Dosing grid. We consider two agents labelled agent W_1 and agent W_2; agent W_1 has I dose levels labelled $i = 1, \ldots, I$ and agent W_2 has J dose levels labelled $j = 1, \ldots, J$. The dose combinations are therefore labelled d_{ij}, making up an I by J dosing grid, with an example of such a grid for $I = 2$ and $J = 5$ shown in Table 1.

3.1.2   Trial procedure. In the motivating setting introduced in Section 2, a toxicity or activity outcome may be observed at any point during the trial, and therefore the trial procedure is set up to reflect this.

Patients are followed up for a total of τ treatment cycles, with the trial proceeding as follows:

1. The first cohort of patients is treated at the pre-specified starting dose combination.
2. After the previous cohort has been observed for one cycle of treatment, the next cohort is assigned to a dose combination based on admissibility and a criterion of utility, both incorporating activity and toxicity.
3. Repeat step 2 until a stopping rule is triggered.

Figure 1 illustrates how such a trial would progress, with each vertical line representing the start of a cycle. Patients may observe an activity response only, an activity response and then a DLT response, a DLT response only, or neither an activity nor DLT response, with all four types of outcome visualised here.

3.1.3 Admissibility and utility. In all designs, a cohort may only be assigned to dose combination d_{ij} if that dose combination is deemed to be admissible. This means that it is above a certain probability of being both safe and active, and is formally defined as:

$$\mathbb{P}(\pi_T(d_{ij}) < \phi_T) > q_T \quad \& \quad \mathbb{P}(\pi_A(d_{ij}) > \phi_A) > q_A$$

for some pre-defined thresholds q_T for toxicity and q_A for activity, where ϕ_T and ϕ_A are the target probabilities for toxicity and activity, respectively, for the follow-up period τ, and $\pi_T(d_{ij})$ and $\pi_A(d_{ij})$ are the probabilities of toxicity and activity at dose combination d_{ij}, respectively. This ensure that assignment is to a dose combination that is considered neither futile nor unsafe.

Once dose combinations are deemed admissible, the dose combination that optimises some utility criterion is chosen. For all designs, for fairness of comparability, the following utility as used by Barnett et al.[12] is used:

$$U(\pi_A, \pi_T) = \pi_A - \omega_1 \pi_T - \omega_2 \pi_T \mathbb{I}(\pi_T > \phi_T) \tag{1}$$

where the two weights, ω_1 and ω_2, reflect the trade-off between toxicity and activity. Note that this utility is different to the originally proposed utility used by the TITE-comb-BOIN12 design, which can be found in the Supplemental Materials.

3.2 Joint TITE-POCRM

The POCRM[3] maps the two-dimensional dosing grid into one dimension so that a single-agent dose-response model can be applied. In order to do this, one must choose what is known as the dose skeleton, the dose-response model and the ordering of the mapping itself.

Firstly, consider the dose skeleton, x. This can be interpreted for many models as the prior probability of observing a toxicity response on a given dose; a vector of increasing values between 0 and 1, with a length equal to the number of doses. This is used in the dose-response model as opposed to the real dose values, with the advantage of being invariant to units of measurement.

Next, consider the dose-response model. In the original proposal of the POCRM, the one parameter power model was used:

$$F(x, \alpha) = x^{\exp(\alpha)}$$

where α is the parameter of the model, assigned a $N(0, 1.34^2)$ prior. Although displaying reasonable operating characteristics, the two parameter logistic model allows for much more flexibility in the dose–response relationship:

$$F(x, \boldsymbol{\beta}) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1)$ is the parameter vector, with $(\beta_0, \log(\beta_1))$ being assigned a bivariate normal prior, $\begin{pmatrix} \beta_0 \\ \log(\beta_1) \end{pmatrix} \sim N_2(\begin{pmatrix} c_1 \\ c_2 \end{pmatrix}, \begin{pmatrix} v_1 & 0 \\ 0 & v_2 \end{pmatrix})$.

In this setting, since responses may be late-onset, we also consider the TITE of a response in the model:

$$G(x, \cdot, w) = wF(x, \cdot)$$

where w is a function of the TITE of a response.

Next, consider the mapping. In order to choose which dose combinations are mapped to which elements of the skeleton, one has to consider the ordering of the dose combinations. For example, one may map the grid in Table 1 into one dimension using the ordering:

$$R(d_{11}) \leq R(d_{21}) \leq R(d_{12}) \leq R(d_{22}) \leq R(d_{13}) \leq R(d_{23}) \leq R(d_{14}) \leq R(d_{24}) \leq R(d_{15}) \leq R(d_{25})$$

where $R(d_{ij})$ represents the toxicity response rate of dose d_{ij}. This is just one of many possible orderings. Using the dose skeleton of $(0.15, 0.19, 0.22, 0.26, 0.30, 0.34, 0.38, 0.42, 0.46, 0.50)$ would map d_{11} to 0.15, d_{21} to 0.19 and so on.

A one-dimensional dose-response model can then be applied. Since there are a large number of potential mappings that can exist, the POCRM considers only a subset of these, based on the ordering of the dosing grid. To choose the one dimensional ordering in which to apply the dose skeleton to the dosing grid, a set of so-called partial orderings

are considered. These are plausible orderings of the dose-combinations that adhere to the dose orderings. For example, increasing the dose level of one agent whilst keeping the other the same would be a "higher" dose. (e.g. $R(d_{12}) \geq R(d_{11})$) However, when increasing the dose level of one agent and decreasing the other (an off-diagonal move) one would not necessarily be able to specify which dose combination is "higher" (e.g. $R(d_{12}) \geq R(d_{21})$ or $R(d_{12}) \leq R(d_{21})$). Further restrictions may be placed on the set of partial orderings in order to define the subset to consider. These restrictions may be drawn from information about the agents themselves, or from a statistical perspective. It is worth noting that increasing the number of considered partial orderings substantially increases the computational cost of the design, and so including multiple orderings that are very similar may not yield any increase in performance. In general, one must consider enough orderings to capture the main differences that could occur, as originally argued by Wages et al.[3] in the suggestion of "good" orderings.

In this application, since we are considering both activity and toxicity endpoints, how they are incorporated into each part of the POCRM must be taken into account. The skeleton is not necessarily the same for activity and toxicity, but is fixed for each prior to conducting the study. The dose-response model must be for both toxicity and activity and the mapping must be done for activity and toxicity separately. Like the dose skeleton, the set of partial orderings is not necessarily the same for activity and toxicity, and hence the chosen ordering is also not necessarily the same.

The Joint TITE-POCRM therefore breaks down into two parts at each dosing decision:

- **Part 1:** Choose the partial ordering for toxicity and the partial ordering for activity with the highest associated posterior probabilities to be used for model fitting.
- **Part 2:** Use the chosen orderings for toxicity and activity from Part 1 to fit the Joint TITE-CRM. This is used to determine the next dose combination.

Since they are two distinct parts, it is not necessary that the models used in each part are the same. The gain in accuracy of using a two-parameter joint logistic model in Part 2 as opposed to a one-parameter independent power model is large, as this gives much more flexibility and allows the most use out of the observed data to estimate this relationship.[25] However in Part 1, not much is gained by the more complex model, which consequently substantially increases computational burden. This is because the objective is only to find the right ordering, and not precise estimation. Therefore it is proposed here to use one-parameter independent power models in Part 1 to determine the ordering, and a two parameter joint logistic model in Part 2 to determine the OBD.

Part 1: Choosing partial orderings

A set of M_T partial orderings for toxicity, labelled $m_T = 1, \ldots M_T$, and a set of M_A partial orderings for activity, labelled $m_A = 1, \ldots M_A$, are considered. The dose-response model for each partial ordering is referred to as a working model. For each set, a prior probability is assigned to each partial ordering within the set, these are $p(m_A)$ and $p(m_T)$ with $\sum_{m_A=1}^{M_A} p(m_A) = \sum_{m_T=1}^{M_T} p(m_T) = 1$. This probability represents the prior belief that the corresponding partial ordering is the correct ordering. The sets M_T and M_A are the "good" orderings recommended by Wages et al.,[3] further details of which are given in the Supplemental Materials.

Toxicity and activity are considered independently in this part, both following.[7] with the following dose-response models:

$$G_{P1}(x_{\ell,m_T}^{(T)}, w_\ell^{(T)}, \alpha^{(T)}) = w_\ell^{(T)} (x_{\ell,m_T}^{(T)})^{\exp(\alpha^{(T)})}$$

$$L_{P1,m_T}(\alpha^{(T)}|\Omega_{m_T}) = \prod_{\ell=1}^{n} G_{P1}(x_{\ell,m_T}^{(T)}, w_\ell^{(T)}, \alpha^{(T)})^{y_\ell^{(T)}} (1 - G_{P1}(x_{\ell,m_T}^{(T)}, w_\ell^{(T)}, \alpha^{(T)}))^{(1-y_\ell^{(T)})}$$

for up to and including patient n, where $x_{\ell,m_T}^{(T)}$ is the toxicity skeleton value for the dose given to patient ℓ under working model m_T, $y_\ell^{(T)}$ is the binary toxicity outcome for patient ℓ, $w_\ell^{(T)}$ is the TITE weight for toxicity for patient ℓ, $\alpha^{(T)}$ is the parameter of the dose-response model under working model m_T. Similarly for activity, replacing T with A.

The TITE weight for toxicity, $w^{(T)}$, is defined as the proportion of total follow-up, as used by Cheung and Chappell,[26] unless a DLT is observed in which case $w_\ell^{(T)} = 1$. In a similar fashion, the TITE weight for activity, $w^{(A)}$ is defined as the proportion of total follow-up, unless an activity outcome is observed, in which case $w_\ell^{(A)} = 1$. However, if a DLT is observed before an activity outcome can be observed, then the activity observation is censored by DLT time and so $w_\ell^{(A)} = $ (DLT time - entry time)/Total follow up time. This is the same specification as used by the Joint TITE-CRM.[12]

For activity and toxicity, a prior g is elicited on the parameter α, and the likelihoods are calculated for each of the considered working models. For each working model, the posterior probability of the corresponding partial ordering for

toxicity is therefore calculated as:

$$p(m_T|\Omega_n) = \frac{p(m_T)\int_{\alpha^{(T)}} L_{m_T}(\alpha^{(T)}|\Omega_n)g(\alpha^{(T)})d\alpha^{(T)}}{\sum_{m_T=1}^{M_T} p(m_T)\int_{\alpha^{(T)}} L_{m_T}(\alpha^{(T)}|\Omega_n)g(\alpha^{(T)})d\alpha^{(T)}}$$

and similarly for activity, again replacing T with A. For both activity and toxicity, the respective individual working models with the highest posterior probability, labelled m_A^* and m_T^*, are chosen to proceed to Part 2.

Part 2: Determining next dose combination

The chosen partial orderings from Part 1 are used in the following model, a joint two-parameter logistic model for both safety and activity outcomes, the Joint TITE-CRM.[12] The model for toxicity is outlined here, with the same used for activity, replacing T with A:

$$F(x_{\ell,m_T^*}^{(T)}, \boldsymbol{\beta^{(T)}}) = \frac{\exp(\beta_0^{(T)} + \beta_1^{(T)} x_{\ell,m_T^*}^{(T)})}{1 + \exp(\beta_0^{(T)} + \beta_1^{(T)} x_{\ell,m_T^*}^{(T)})}$$

where $\boldsymbol{\beta^{(T)}} = (\beta_0^{(T)}, \beta_1^{(T)})$. Similarly for activity, replacing T with A.

The same weights for toxicity and activity observations as used in part 1 ($w_\ell^{(T)}$ and $w_\ell^{(A)}$ respectively) are used in the weighted dose response models:

$$G_{P2}(x_{\ell,m_T^*}^{(T)}, w_\ell^{(T)}, \boldsymbol{\beta^{(T)}}) = w_\ell^{(T)} F(x_{\ell,m_T^*}^{(T)}, \boldsymbol{\beta^{(T)}})$$

The G_{P2} for activity and toxicity are related by a Gumbel Model with correlation parameter ψ considering all binary combinations of toxicity and activity outcomes, as in the Joint TITE-CRM, with full details given in the Supplemental Materials.

Priors are elicited on the parameters $\psi, \beta_0^{(A)}, \beta_1^{(A)}, \beta_0^{(T)}, \beta_1^{(T)}$ using the calibration procedure outlined in the Supplemental Materials, and the likelihood, also in the Supplemental Materials, is used to update the joint posterior for all of the parameters using MCMC methods. The posterior distributions of the probability of activity and toxicity on all cycles at each dose combination, respectively, π_A and π_T, are used to determine the admissibility and utility of each dose combination, and therefore the recommended next dose combination.

Out of the admissible dose combinations, the dose combination with the highest utility (Equation 1) is recommended.

3.3 Joint TITE-BLRM

The Joint TITE-BLRM uses the actual dose values of agent W_1, $d_i^{(W_1)}$ and agent W_2, $d_j^{(W_2)}$ in the model. The toxicity of the two agents is modelled, and their odds linked by an odds-multiplier (as recommended by Neuenschwander et al.[5]), and the same is done for activity. The dual agent TITE activity and toxicity models are then linked by a Gumbel model with correlation parameter ψ in a similar fashion to the Joint TITE-CRM.

The model for toxicity is outlined here, with the same used for activity, replacing T with A. The toxicity of agent W_1 is given as:

$$F(d_i^{(W_1)}, \boldsymbol{\beta^{(W_1,T)}}) = \frac{\exp(\beta_0^{(W_1,T)} + \beta_1^{(W_1,T)} d_i^{(W_1)})}{1 + \exp(\beta_0^{(W_1,T)} + \beta_1^{(W_1,T)} d_i^{(W_1)})}$$

where $\boldsymbol{\beta^{(W_1,T)}} = (\beta_0^{(W_1,T)}, \beta_1^{(W_1,T)})$, and the toxicity of agent W_2 is given as:

$$F(d_j^{(W_2)}, \boldsymbol{\beta^{(W_2,T)}}) = \frac{\exp(\beta_0^{(W_2,T)} + \beta_1^{(W_2,T)} d_j^{(W_2)})}{1 + \exp(\beta_0^{(W_2,T)} + \beta_1^{(W_2,T)} d_j^{(W_2)})}$$

where $\boldsymbol{\beta^{(W_2,T)}} = (\beta_0^{(W_2,T)}, \beta_1^{(W_2,T)})$, with their odds linked:

$$odds_{d_i^{(W_1)}, d_j^{(W_2)}}^{(T)} = odds_{d_i^{(W_1)}, d_j^{(W_2)}}^{(T),0} \exp(\eta^{(T)} d_i^{(W_1)} d_j^{(W_2)})$$

where

$$odds^{(T),0}_{d_i^{(W_1)}, d_j^{(W_2)}} = \frac{F(d_i^{(W_1)}, \boldsymbol{\beta}^{(W_1,T)})}{1 - F(d_i^{(W_1)}, \boldsymbol{\beta}^{(W_1,T)})} + \frac{F(d_j^{(W_2)}, \boldsymbol{\beta}^{(W_2,T)})}{1 - F(d_j^{(W_2)}, \boldsymbol{\beta}^{(W_2,T)})} + \frac{F(d_i^{(W_1)}, \boldsymbol{\beta}^{(W_1,T)}) F(d_j^{(W_2)}, \boldsymbol{\beta}^{(W_2,T)})}{(1 - F(d_i^{(W_1)}, \boldsymbol{\beta}^{(W_1,T)}))(1 - F(d_j^{(W_2)}, \boldsymbol{\beta}^{(W_2,T)}))}$$

The contribution to the likelihood is then weighted in the same way as in the Joint TITE-POCRM, with the same definition of weights *w*:

$$G_B(x_\ell^{W_1}, x_\ell^{W_2}, w_\ell^{(T)}, \boldsymbol{\beta}^{(W_1,T)}, \boldsymbol{\beta}^{(W_2,T)}, \eta^{(T)}) = w_\ell^{(T)} \frac{odds^{(T)}_{x_\ell^{W_1}, x_\ell^{W_2}}}{1 + odds^{(T)}_{x_\ell^{W_1}, x_\ell^{W_2}}}$$

where $x_\ell^{W_1}$ is the dosage of agent W_1 assigned to patient ℓ, and $x_\ell^{W_2}$ is the dosage of agent W_2 assigned to patient ℓ.

There are therefore 11 parameters in this model: $\beta_0^{W_2,A}$, $\beta_1^{W_2,A}$, $\beta_0^{W_1,A}$, $\beta_1^{W_1,A}$, $\beta_0^{W_2,T}$, $\beta_1^{W_2,T}$, $\beta_0^{W_1,T}$, $\beta_1^{W_1,T}$, $\eta^{(T)}$, $\eta^{(A)}$, ψ. Priors are elicited on all 11 parameters, again using the same calibration procedure as the Joint TITE-POCRM, detailed in the Supplemental Materials and the likelihood also outlined in the Supplemental Materials is used to update the joint posterior for all of the parameters using MCMC methods. In the same way as the Joint TITE-POCRM, the posterior distributions of the probability of activity and toxicity on all cycles at each dose combination, respectively, π_A and π_T, are used to determine the admissibility and utility of each dose combination, and therefore the recommended next dose combination. The same utility criterion (1) as used for the Joint TITE-POCRM is used for the Joint TITE-BLRM.

3.4 TITE-comb-BOIN12

As a model-assisted design, TITE-comb-BOIN12[18] uses a model at each dose level to make escalation decisions, rather than modelling the dose–response relationship as the previous two designs have. This design extends the original proposal of TITE-BOIN12 for single agent trials in line with the dual agent BOIN design.[1] In this comparison, the utility given in Equation 1 is used in order for the modelling approaches to be compared fairly since different utilities will naturally give different definitions of optimality. Further details of this model-assisted comparator are given in the Supplemental Materials, including the alternative utility.

4   Simulations

To assess and compare the relevant operating characteristics of the designs, simulations are conducted over a range of plausible scenarios. In each scenario, 1000 trial simulations are performed.

The setting is motivated by the example illustrated in Section 2, with two doses of Agent W_1 and five doses of Agent W_2. The starting dose is d_{22}, reflecting the fact that the lowest dose levels of each agent are fall-back doses. The procedure of the trial follows that outlined in Section 3.1.2.

Details of the values used in the priors of all methods are given in the Supplemental Materials.

4.1   Scenarios

Given the complexity of the setting, it is important to consider a wide range of relevant scenarios. Six safety scenarios and eight activity scenarios are combined to give 48 total scenarios. The individual safety and activity scenarios are defined so that different areas of the grid are active and safe. When combined, this gives the OBD in different positions in the grid. The six toxicity and activity scenarios are given in Table 2, with the probability of toxicity and activity for the whole observation window of $\tau = 3$ cycles quoted. The individual utility of the 48 scenarios are provided in the Supplemental Materials, as well as the alternative BOIN utility.

Table 2. Scenarios considered for simulation study, giving the probability of observing a toxicity/activity event at each dose combination.

	Agent W_1 (mg)	Agent W_2 (kBq/kg)				
		50	75	100	125	150
T1	600	0.03	0.07	0.11	0.15	0.2
	1200	0.05	0.09	0.13	0.25	0.3
T2	600	0.1	0.15	0.2	0.3	0.4
	1200	0.45	0.5	0.55	0.6	0.6
T3	600	0.05	0.08	0.15	0.2	0.45
	1200	0.1	0.12	0.3	0.4	0.5
T4	600	0.1	0.2	0.4	0.5	0.6
	1200	0.3	0.45	0.55	0.6	0.6
T5	600	0.3	0.45	0.5	0.55	0.6
	1200	0.4	0.5	0.55	0.6	0.6
T6	600	0.4	0.4	0.5	0.5	0.6
	1200	0.4	0.4	0.5	0.5	0.6

	Agent W_1 (mg)	Agent W_2 (kBq/kg)				
		50	75	100	125	150
A1	600	0.2	0.3	0.35	0.5	0.55
	1200	0.25	0.4	0.45	0.6	0.65
A2	600	0.3	0.32	0.38	0.4	0.46
	1200	0.34	0.36	0.42	0.44	0.48
A3	600	0.06	0.08	0.12	0.2	0.3
	1200	0.1	0.15	0.25	0.35	0.4
A4	600	0.05	0.2	0.3	0.4	0.5
	1200	0.1	0.25	0.35	0.45	0.55
A5	600	0.1	0.12	0.14	0.16	0.18
	1200	0.2	0.3	0.4	0.5	0.6
A6	600	0.1	0.1	0.1	0.1	0.1
	1200	0.1	0.1	0.1	0.1	0.2
A7	600	0.2	0.3	0.4	0.4	0.4
	1200	0.25	0.35	0.45	0.45	0.45
A8	600	0.2	0.3	0.4	0.3	0.2
	1200	0.3	0.35	0.45	0.35	0.25

Six for safety, labelled T1–T6, eight for activity labelled A1–A8. These are combined to give 48 total scenarios.

4.2 Data generation

Data in the form of event times for activity and toxicity are generated for each simulated patient response. Following Barnett et al.,[12] these event times t_A and t_T are generated from a bivariate log-normal model:

$$\begin{pmatrix} t_T \\ t_A \end{pmatrix} \sim Lognormal_2 \left(\begin{pmatrix} \mu_T \\ \mu_A \end{pmatrix}, \begin{pmatrix} \sigma_T^2 & -\frac{\sigma_T \sigma_A}{2} \\ -\frac{\sigma_T \sigma_A}{2} & \sigma_A^2 \end{pmatrix} \right)$$

with parameters $\mu_T, \mu_A, \sigma_T, \sigma_A$ calculated for each dose combination so that the probability of observing a toxicity event across the observation period equals that defined by the scenario specification, and similarly for activity. The pattern across cycles differs between activity and toxicity, with activity split equally across all three cycles, and 0.75 of the total probability of toxicity assigned to the first cycle. This is in line with the data generation used by Barnett et al.,[12] although for simplicity here we refer only to the true probabilities of response across all cycles in the definition of scenarios.

4.3 Rules

In a similar fashion to Barnett et al.,[12] a set of rules reflecting those applied in such a trial are used. Let $p_{1,d_{i*j*}}$ be the probability of observing a DLT response in cycle 1 on dose d_{i*j*}.

Enforcement rules restrict the set of allowable doses in order to protect the safety of patients in the trial. The dose skipping rule restricts escalation to unexplored doses, and the hard safety rule restricts escalation to explored doses. Stopping rules dictate the criteria for stopping the trial.

The parameters for the admissibility criteria for all designs are set at $q_T = q_A = 0.2$ and $\phi_T = 0.3$ and $\phi_A = 0.2$.

4.3.1 Enforcement.

1. **Dose skipping:** In this dual agent case, the dose skipping rule is defined by agent. Once dose d_{i*j*} has been explored, all doses d_{ij} such that $i \le i*$ an $j \le j* +1$, or $i \le i* +1$ and $j \le j*$ are allowable. This ensures that dose levels of each agent are not skipped, as well as restricting escalation in both agents at once.
2. **Hard safety**: The hard safety rule is defined by the observed number of DLT responses in the first cycle of treatment on explored doses. If there are 3 DLT responses out of 3 patients, 4 or more DLT responses out of 6 patients, or 5 or more DLT responses out of 9 patients on dose d_{i*j*}, then all doses d_{ij} such that $i \ge i*$ and $j* \ge j$, are excluded from further exploration. This is based on an exclusion criterion of $\mathbb{P}(p_{1,d_{i*j*}} > 0.3) > 0.95$ using a beta-binomial model with a $Beta(1, 1)$ prior.

4.3.2 Stopping.

1. **No admissible dose combinations**: If no dose combinations satisfy both admissibility criteria in Section 3.1.3, then the trial is stopped, deeming no dose combination admissible.
2. **Lowest dose combination deemed unsafe**: If $\mathbb{P}(p_{1,d_{11}} > 0.3) > 0.80$ and at least one cohort of patients has been assigned to dose d_{11}, the trial is stopped, deeming all dose combination unsafe.
3. **Highest dose combination deemed very safe**: If $\mathbb{P}(p_{1,d_{IJ}} \le 0.3) > 0.80$ and at least one cohort of patients has been assigned to dose d_{IJ}, the trial is stopped, deeming all doses under target safety.
4. **Sufficient information**: For a pre-defined cut-off value, C_{suff}, if a dose combination is recommended for the next cohort on which C_{suff} patients have already been assigned in the escalation, the trial is stopped. In this case, we use $C_{suff} = 30$.
5. **Hard safety**: If the d_{11} is considered unsafe according to the hard safety enforcement rule, the trial is stopped, deeming all dose combination unsafe.
6. **Maximum patients**: If the maximum number of patients ($n = n_{\max}$) have been recruited, the trial is stopped. In this case, we use $n_{\max} = 60$.

4.4 Results

To evaluate the merits of the three designs, their performance is measured according to criteria based both on their final recommendation of dose combination and the assignment of patients to the dose combinations. It is desirable for a design to give a high proportion of good dose combination recommendations, whilst also not exposing unnecessary levels of patients to unsafe dose combinations.
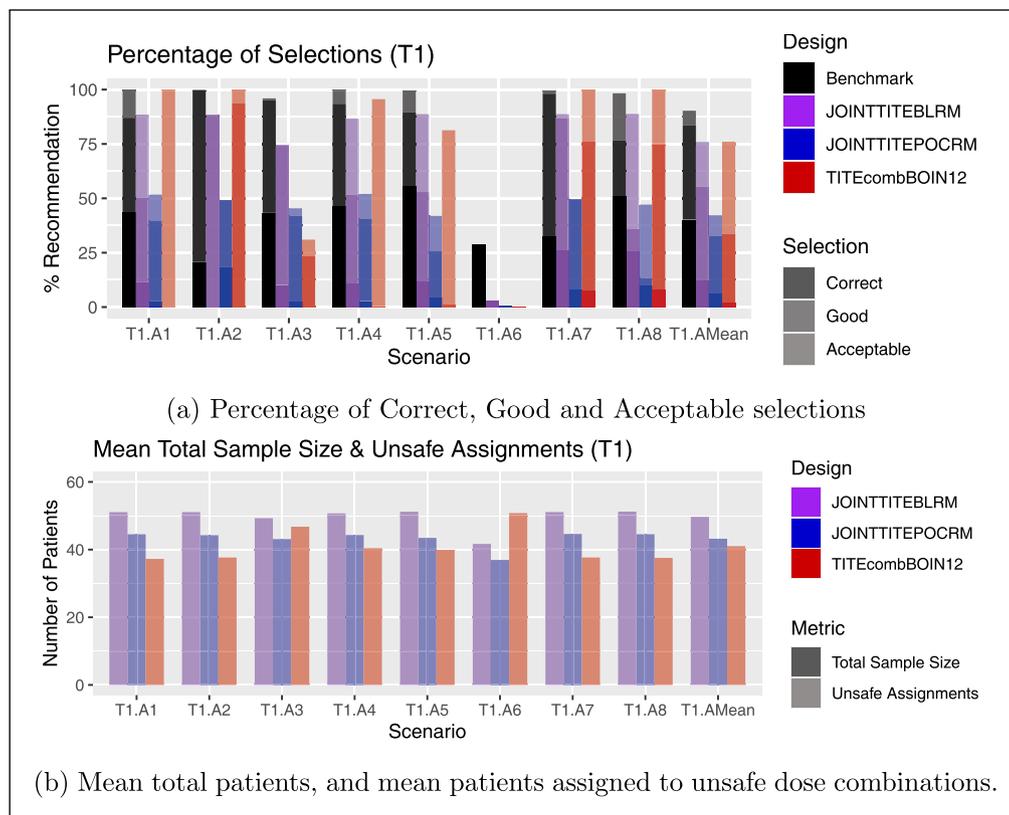
(a) Percentage of Correct, Good and Acceptable selections

(b) Mean total patients, and mean patients assigned to unsafe dose combinations.

Figure 2. Results for scenarios T1.A1–T1.A8, including an average over these six scenarios, labelled T1.AMean. (a) Percentage of Correct, Good and Acceptable selections; (b) Mean total patients, and mean patients assigned to unsafe dose combinations.

To compare the recommendations, we define certain classes of dose combinations. A dose combination is *safe* if the true probability of observing a toxicity is less than or equal to 0.3 and *unsafe* otherwise. A dose combination is *active* if the true probability of observing an activity event is greater than or equal to 0.2 and *futile* otherwise. If a dose is unsafe or futile, then it is *unacceptable*. If it is both safe and active, then it is *acceptable*. Within the class of acceptable dose combinations, the one dose combination with the highest true utility is defined as the *correct* dose combination, however when the utility is very similar across dose combinations with the highest utility, these dose combinations are referred to as *good*.

To gain an assessment of the performance, the selections are also compared to an optimal non-parametric benchmark that takes into account uncertainty of ordering.[27] This is modified in this context to also account for admissibility, with further details given in the Supplemental Materials.

It is worth noting that for the approaches to be fairly compared, we have used the same utility criterion for all approaches. The original proposal of the TITE-comb-BOIN12 design introduced an alternative utility criterion, which is evaluated in the Supplemental Materials. The decision was made to use the same utility criterion for all methods in this comparison as for the alternative BOIN utility criterion, in some scenarios, the definition of correct and good doses are not identical to the utility used here, and so direct comparisons of these metrics for different utility criteria are not meaningful.

Figures 2 to 7 display the results across the 48 scenarios, each figure representing one toxicity scenario, with its combination scenario with each activity scenario, and an average across these combinations. Within each figure, subfigure (a) concerns the selections, displaying the percentage of Correct, Good and Acceptable selections. Subfigure (b) concerns the assignments, displaying the average total sample size and the average number of patients assigned to unsafe doses. Figure 8 displays the mean of these results over the toxicity scenarios for each activity scenario, and an average over all 48 scenarios. Full details of the selections and assignments are available in the Supplemental Materials.

Since the overall average displayed in the final columns of Figures 8(a) and (b) show only a small difference across the three approaches, one might assume that there is comparable performance. However, there is a large amount of variation when one looks in more detail at the individual scenarios.
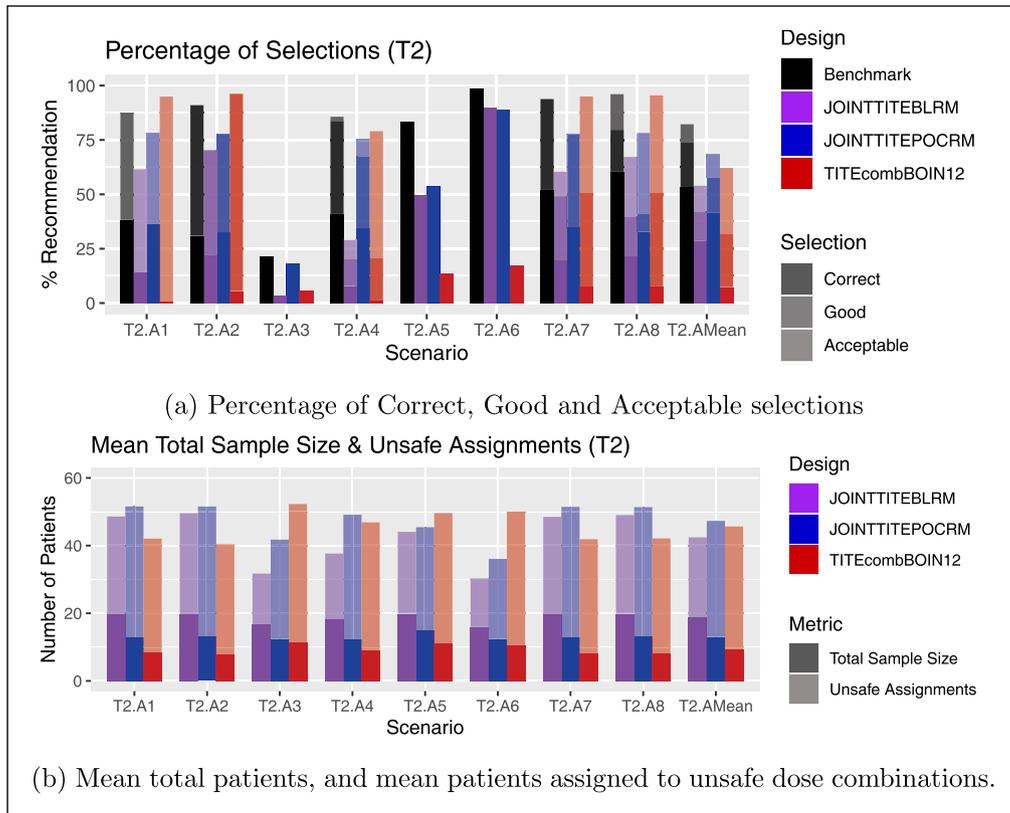
(a) Percentage of Correct, Good and Acceptable selections

(b) Mean total patients, and mean patients assigned to unsafe dose combinations.

Figure 3. Results for scenarios T2.A1–T2.A8, including an average over these six scenarios, labelled T2.AMean. (a) Percentage of Correct, Good and Acceptable selections; (b) Mean total patients, and mean patients assigned to unsafe dose combinations.

For example, in scenarios T1.A1, T1.A2, T1.A3, T1.A4 and T1.A5, the Joint TITE-BLRM considerably outperforms the Joint TITE-POCRM. In these scenarios, all dose combinations are safe, and the correct dose combination is at the highest level of Agent W_2, with good dose combinations surrounding it. In these scenarios, the somewhat poor performance of the Joint TITE POCRM is due to it incorrectly stopping early due to stopping rule 4.3.2., that the highest dose is deemed too safe. The TITE-comb-BOIN12 shows a poor performance in terms of correct selections and good, but a strong performance in terms of acceptable selections across activity scenarios 1–6. However, in activity scenarios 7 and 8, where the dose–activity relationship is a plateau and bell curve, the TITE-comb-BOIN12 design outperforms the model-based designs, due to this activity relationship not adhering to the monotonicity assumptions of the models. This trend is also seen throughout the other safety scenarios.

In contrast, it can be seen in Figure 3(a) in scenario T2.A1 and T2.A4, the Joint TITE-POCRM considerably outperforms the Joint TITE-BLRM. In both of these scenarios, dose combination d_{14} is the OBD. In T2.A4, the Joint TITE-BLRM incorrectly concludes that there is no admissible dose in 57% of simulations, despite a number of dose combinations being truly active and safe. The Joint TITE-POCRM however, shows a better performance. It is also apparent from Figure 3(b) that the Joint TITE-BLRM assigns more patients on average to unsafe dose combinations, despite a smaller average overall sample size. As a comparator, the TITE-comb-BOIN12 here shows a poor performance when there are truly no admissible dose combinations, being more willing to recommend an inactive dose combination.

Figure 4(a) shows some interesting contrasting behaviour of the two model-based approaches. In T3.A3 and T3.A5, the Joint TITE-BLRM shows superior performance in terms of correct selections, however in T3.A1 T3.A2 and T3.A4, the Joint TITE-POCRM shows superior performance in terms of correct selections, and the Joint TITE-BLRM fails to select the correct dose combination at all. Here the correct dose combination is on the lowest dose of Agent W_1, and the Joint TITE-BLRM seems to favour good doses on the higher dose level of Agent W_1. The TITE-comb-BOIN12 again is displaying behaviour whereby it is more willing to recommend inactive dose combinations.

Figure 5(a) again shows the contrast in the behaviour. The average performance across T4 is very similar in terms of correct selections, but the Joint TITE-BLRM has more good selections. However. for example T4.A4 and T4.A5
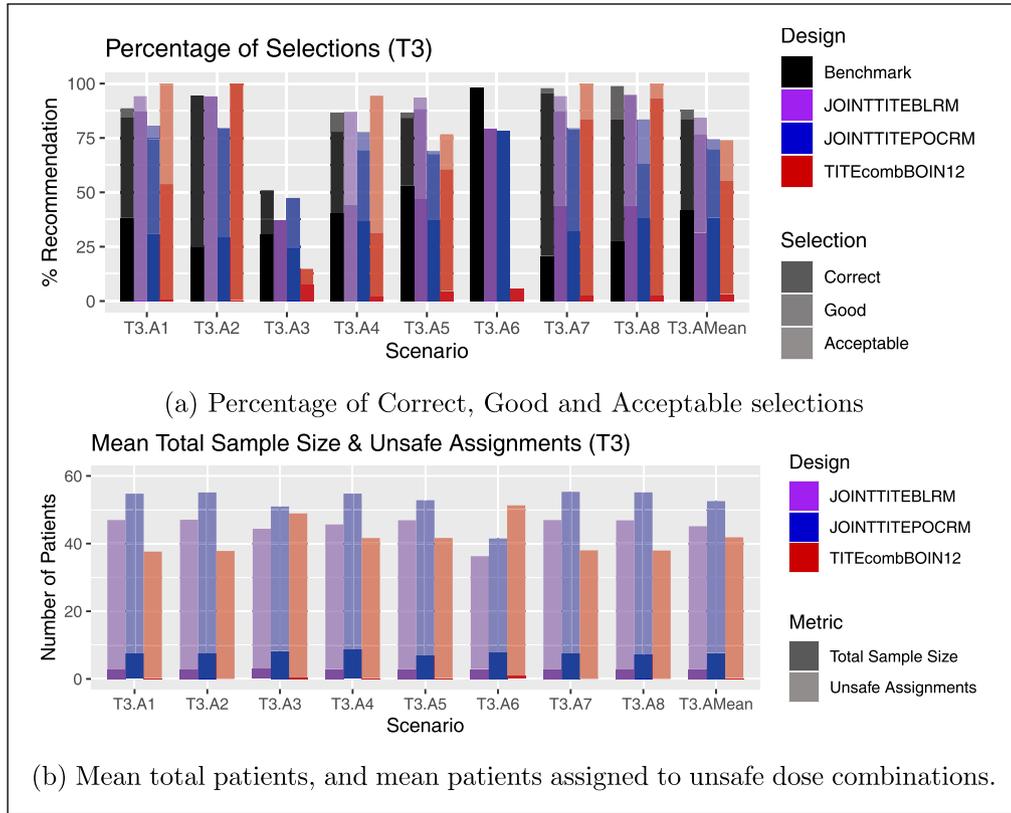
(a) Percentage of Correct, Good and Acceptable selections

(b) Mean total patients, and mean patients assigned to unsafe dose combinations.

Figure 4. Results for scenarios T3.A1–T3.A8, including an average over these six scenarios, labelled T3.AMean. (a) Percentage of Correct, Good and Acceptable selections; (b) Mean total patients, and mean patients assigned to unsafe dose combinations.

show very different performances. In T4.A4, the Joint TITE-BLRM shows a very poor performance. The only acceptable/good/correct dose is d_{12}, and the Joint TITE-BLRM only selects this dose combination in 3% of simulations, selecting the inactive d_{21} in 18% of simulations and no admissible dose in 74% of simulations. Whereas the Joint TITE-POCRM selects the correct dose combination in 30% of simulations. In contrast, in scenario T4.A5, dose d_{21} is the only acceptable/good/correct dose, and the Joint TITE-POCRM only selects this dose in 17% of simulations, selecting the inactive d_{12} in 5% of simulations and no admissible dose in 69% of simulations. The Joint TITE-BLRM however recommends the correct dose in 54% of simulations. This comparison is of particular interest, as from the starting dose of d_{22} being unsafe, these two scenarios illustrate the differences when it is the increase in Agent W_1 or Agent W_2 that is driving the toxicity. The Joint TITE-POCRM shows a more balanced performance, whereas the Joint TITE-BLRM shows a clear preference for Agent W_1.

Since in T5, only the lowest dose combination is safe, only in scenarios T5.A1 and T5.A2 are we seeking to find the correct dose combination, and in the other four T5 scenarios we seek to declare no admissible doses. In Figure 6, the Joint TITE-POCRM shows superior performance in T5.A5 and T5.A6, where the Joint TITE-BLRM is more likely to recommend the inactive dose combinations than the Joint TITE-POCRM. The Joint TITE-BLRM identifies the lowest dose combination as the OBD in scenarios T5.A1 and T5.A2 more than the Joint TITE-POCRM.

In the T6 scenarios, all dose combinations are unsafe, and therefore the correct recommendation is either stopping for no admissible dose, or for a safety stopping rule. Here, Figure 7 illustrates the performance of the methods in these scenarios. It is clear to see that the TITE-comb-BOIN12 design correctly stops in more simulations in T6.A1 and T6.A2, but also stops sooner across most scenarios, with much smaller mean sample sizes. The Joint TITE-POCRM and Joint TITE-BLRM show similar performances to each other in T6.A2, T6.A4 and T6.A6, however the across the rest of the scenarios the Joint TITE-POCRM shows superior performance. Both model-based methods show poorer performances in the scenarios where the lowest dose combinations are active.

Figure 8 shows the average results over toxicity scenarios, including an average over all 48 scenarios. Whilst conclusions cannot be drawn in cases where the averages are similar, it is of interest to note those activity scenarios where the
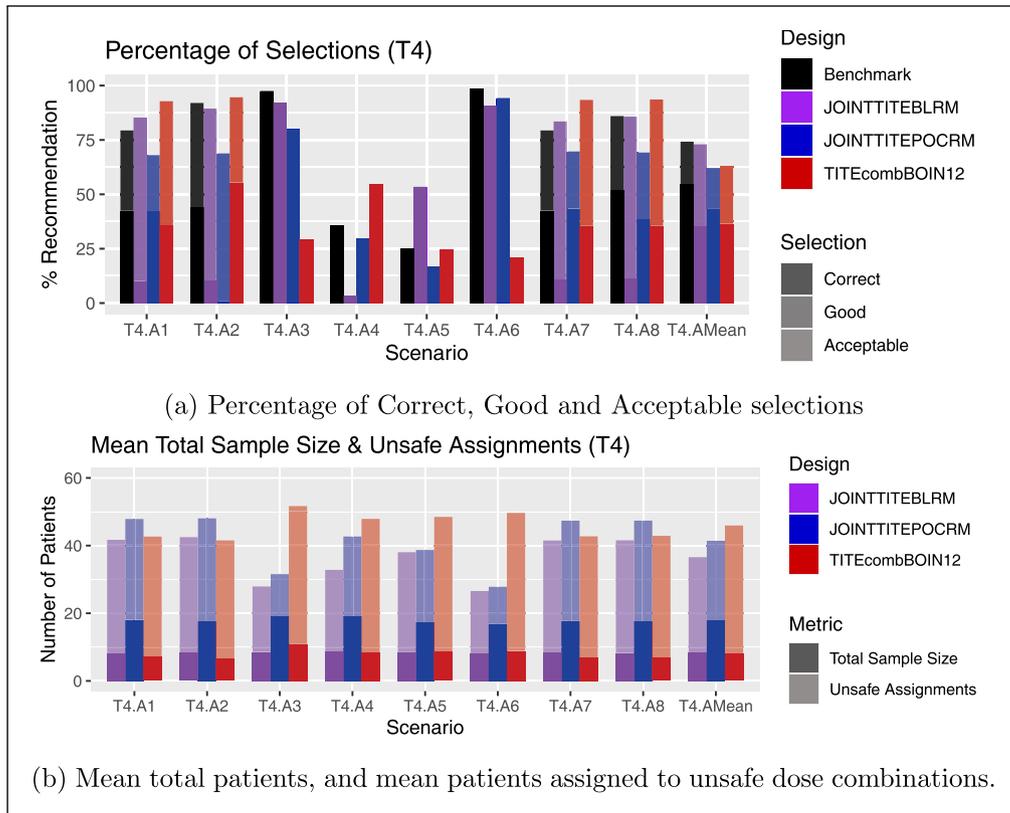
(a) Percentage of Correct, Good and Acceptable selections

(b) Mean total patients, and mean patients assigned to unsafe dose combinations.

Figure 5. Results for scenarios T4.A1–T4.A8, including an average over these six scenarios, labelled T4.AMean. (a) Percentage of Correct, Good and Acceptable selections; (b) Mean total patients, and mean patients assigned to unsafe dose combinations.

results are contrasting. Of note is A4, where the Joint TITE-POCRM shows superior performance over the Joint TITE-BLRM. Here, in three out of the six individual scenarios, the OBD was at the the 600 mg dose of Agent W_1, which the Joint TITE-BLRM finds harder to identify. The TITE-comb-BOIN12 shows superior performance and decreased sample size in A2, where all dose combinations are active, however shows an inferior performance and increased sample size in A6, where all but the highest dose combination is inactive. This is driven by the reluctance to stop for inactivity by the TITE-comb-BOIN12 design. As previously noted, the TITE-comb-BOIN12 design outperforms the model-based designs in activity scenarios 7 and 8, where the dose–activity relationship breaks the assumptions of the models.

4.5 Changing the size of the dosing grid

So far we have considered scenarios whereby the dosing grid has dimensions 2×5, Agent W_1 having two dose levels, 1200 mg and a fall-back of 600 mg, and Agent W_2 has 5 dose levels to explore. Now we extend the set of scenarios to include those where the dosing grid has dimensionality 3×3. Here we extend the range of Agent W_1 to include a higher dose of 1800 mg, and restrict the range of Agent W_2 to the lower three doses.

A smaller range of scenarios are considered in this setting, labelled S1–S6, with their definitions given in Table 3. These scenarios are chosen to give a range of relationships across the grid, and to further explore the results seen in the 2×5 grid.

The starting dose remains d_{22}, and the prior distribution require little amendments, only the POCRM requires the final value in the skeleton to be removed. The partial orders considered in the POCRM follow the same pattern as recommended by Wages et al.,[28] detailed in the Supplemental Materials.

The results for the simulation study of 3×3 grids are shown in Figure 9. There are noticeable differences between the Joint TITE-BLRM and Joint TITE-POCRM across this range of scenarios. In S1, the Joint TITE-BLRM recommends the correct dose in 29% of the simulations, compared to only 15% for the Joint TITE-POCRM. The Joint TITE-BLRM also has a higher percentage of good and acceptable selections.

In S2, again the Joint TITE-BLRM shows more promising behaviour, recommending a good dose combination in 59% of simulations compared to the 12% of the Joint TITE-POCRM. This stark difference is due to the POCRM incorrectly

Table 3. Toxicity and activity probabilities and utility values for scenarios S1–S6.

		Toxicity			Activity				Utility		
	Agent W_1 (mg)	Agent W_2 (kBq/kg)			Agent W_2 (kBq/kg)			Agent W_1 (mg)	Agent W_2 (kBq/kg)		
		50	75	100	50	75	100		50	75	100
S1	600	0.1	0.15	0.2	0.05	0.1	0.15	600	0.017	0.0505	0.084
	1200	0.15	0.2	0.3	0.2	0.25	0.3	1200	*0.1505*	*0.184*	*0.201*
	1800	0.2	0.3	0.45	0.35	0.4	0.45	1800	*0.284*	**0.301**	−0.7885
S2	600	0.05	0.09	0.11	0.05	0.08	0.10	600	0.034	0.050	0.064
	1200	0.07	0.13	0.25	0.10	0.12	0.15	1200	0.077	0.077	0.068
	1800	0.15	0.20	0.30	0.20	0.30	0.40	1800	*0.151*	*0.234*	**0.301**
S3	600	0.05	0.10	0.15	0.10	0.20	0.30	600	0.084	*0.167*	*0.251*
	1200	0.20	0.25	0.30	0.20	0.30	0.40	1200	*0.134*	*0.218*	**0.301**
	1800	0.40	0.45	0.50	0.30	0.40	0.50	1800	−0.922	−0.839	−0.755
S4	600	0.20	0.25	0.30	0.20	0.25	0.45	600	*0.134*	0.168	**0.351**
	1200	0.40	0.45	0.50	0.30	0.40	0.50	1200	−0.922	−0.839	−0.755
	1800	0.50	0.55	0.60	0.35	0.50	0.60	1800	−0.905	−0.772	−0.688
S5	600	0.05	0.10	0.20	0.05	0.15	0.30	600	0.034	0.117	**0.234**
	1200	0.15	0.25	0.35	0.10	0.25	0.40	1200	0.051	*0.168*	−0.806
	1800	0.30	0.40	0.45	0.20	0.35	0.45	1800	*0.101*	−0.872	−0.789
S6	600	0.05	0.20	0.40	0.05	0.20	0.35	600	0.034	*0.134*	−0.872
	1200	0.10	0.25	0.45	0.10	0.25	0.40	1200	0.067	*0.168*	−0.839
	1800	0.15	0.30	0.50	0.15	0.30	0.45	1800	0.101	**0.201**	−0.805

Acceptable dose combinations are highlighted in *italics*, good dose combinations are underlined and correct dose combinations are highlighted in **boldface**.
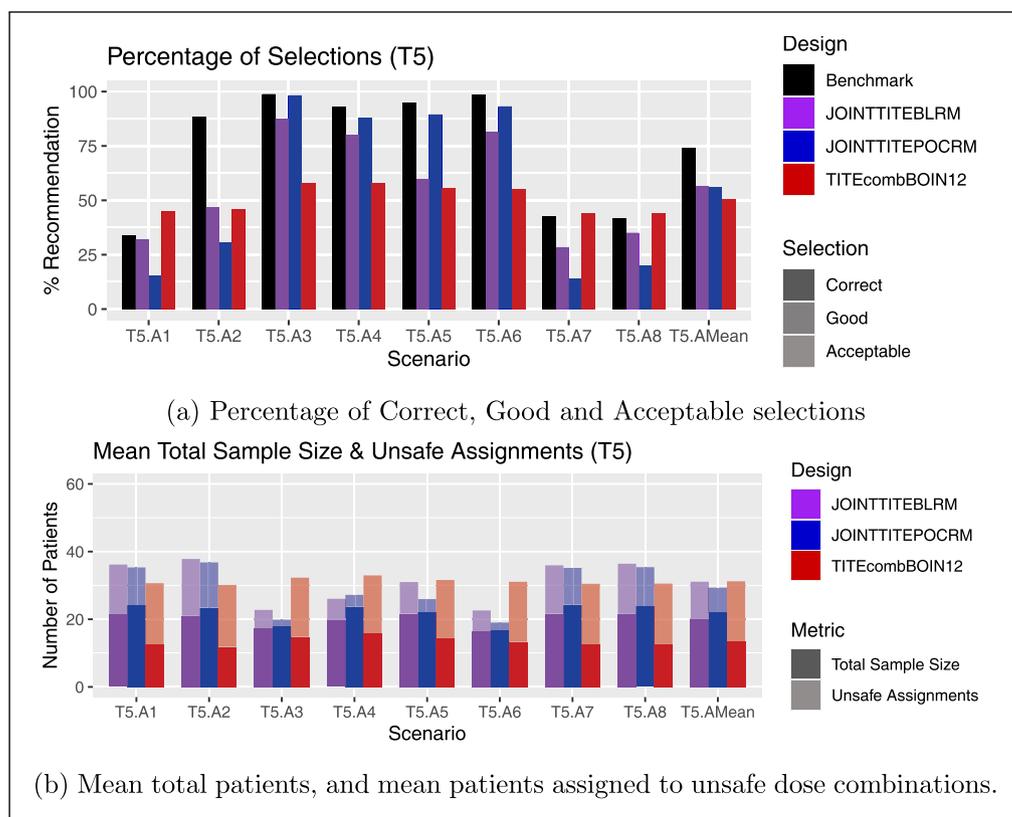
(a) Percentage of Correct, Good and Acceptable selections

(b) Mean total patients, and mean patients assigned to unsafe dose combinations.

Figure 6. Results for scenarios T5.A1–T5.A8, including an average over these six scenarios, labelled T5.AMean. (a) Percentage of Correct, Good and Acceptable selections; (b) Mean total patients, and mean patients assigned to unsafe dose combinations.

stopping the trial for no admissible doses or the highest dose being too safe in 74% of simulations. Here only the 1800 mg dose of Agent W_1 is active, and all doses are safe. The behaviour displayed is similar to that seen in scenario T1.A5, that the Joint TITE-POCRM struggles to identify good doses when only the highest dose level of Agent W_1 is active.

In S3–S6 however, the reverse is true, and the Joint TITE-POCRM show superior performance to the Joint TITE-BLRM. In S3, where the 1800 mg dose of Agent W_1 is unsafe, and only the lowest dose combination is inactive, the Joint TITE-POCRM recommends a good dose in 72% of simulations, compared to 41% for the Joint TITE-BLRM. Interestingly, the Joint TITE-POCRM favours recommending d_{13}, whereas the Joint TITE-BLRM favours recommending d_{21} and d_{22}. In S4, where only the 600 mg dose of Agent W_1 is safe, the BLRM incorrectly recommends the unsafe d_{21} in 38% of simulations, and also favours in recommendation the dose combination d_{11} only recommending the correct d_{13} in 5% of simulations, hence a poorer performance than the POCRM, which favours in recommendation the dose combination d_{12} at 30%, followed by the correct d_{13} in 28% of simulations.

In S5, the acceptable dose combinations lie along the off-diagonal of the grid, increasing utility as the dose of Agent W_2 increases. The Joint TITE-POCRM recommends the correct dose in 39% of simulations, whereas the Joint TITE-BLRM only recommends this dose in 1% of simulations, giving more favour to d_{31} and d_{22}, and has over twice as many simulations recommending no admissible dose than the Joint TITE-POCRM, at 28%. In S6, the highest dose of Agent W_2 in unsafe and the lowest dose of Agent W_2 is inactive, leaving only the middle dose. Here, the Joint TITE-BLRM recommend the correct dose of d_{32} in more simulations than the Joint TITE-POCRM (33% vs 16%), the POCRM recommends a good dose in 59% of simulations compared to the BLRM's 39%. The downfall of the Joint TITE-BLRM in this scenario is in its over-recommendation of d_{31}, which is inactive.

The TITE-comb-BOIN12 again shows a strong performance in terms of acceptable selections, however a poor performance in terms of good and correct selections, with a sample size in between the two model-based designs, and fewer patients assigned to unsafe doses.
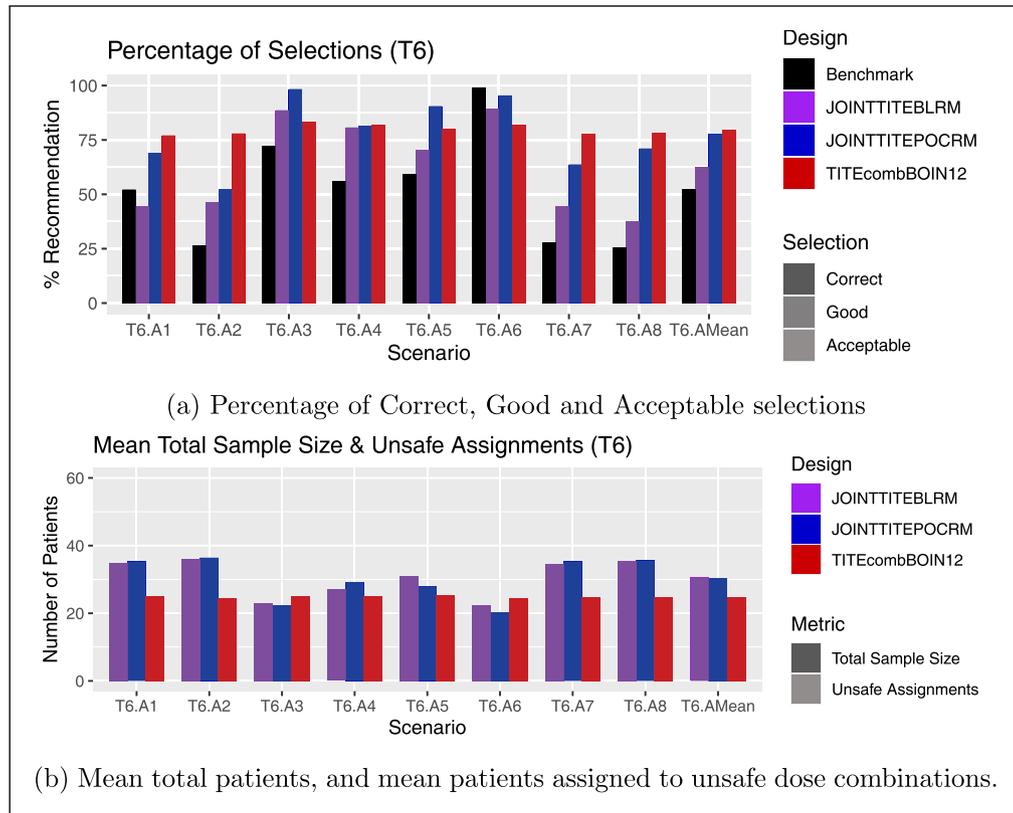
(a) Percentage of Correct, Good and Acceptable selections

(b) Mean total patients, and mean patients assigned to unsafe dose combinations.

Figure 7. Results for scenarios T6.A1–T6.A8, including an average over these six scenarios, labelled T6.AMean. (a) Percentage of Correct, Good and Acceptable selections; (b) Mean total patients, and mean patients assigned to unsafe dose combinations.

5 Discussion

In this work, two model-based approaches for a Phase I/II dose-finding trial with dual agents and late-onset endpoints are presented. One is based on the BLRM, and the other the POCRM. Both approaches show promise, but with contrasting behaviours in a subset of considered scenarios.

These contrasting behaviours are inherent due to the differences in nature of the two designs, accentuated by the dual TITE endpoint. Since the Joint TITE-POCRM has a fixed skeleton that can be mapped onto the dosing grid for the different partial orderings, this means that there is no fixed prior for each dose combination. In contrast, the Joint TITE-BLRM must by definition have a fixed prior for each dose combination. However, the Joint TITE-BLRM models interaction between the two agents and interaction between the two endpoints, whilst the Joint TITE-POCRM only models the interaction between the two endpoints in its second stage. Therefore both approaches offer flexibility in different ways, and this is illustrated in the simulation results presented here. In the case where there are two-dose levels of W_1, the Joint TITE-BLRM is better at identifying when the fall-back dose of Agent W_1 is inactive, and in general explores the higher dose of Agent W_1 well, only deviating to the lower dose if the higher dose is deemed unsafe. The POCRM on the other hand, explores the grid more, and will explore the lower dose of Agent W_1 more often in the scenarios explored here.

One obvious limitation of both methods is their computational intensity. The complex models require MCMC methods to compute posterior distributions for parameters. The large number of parameters on which priors must be elicited is especially problematic for the BLRM as the high dimensionality of the prior calibration greatly increases the computational load. The reliance on the prior calibration should also be noted. In this work, both methods were subject to the same method of calibration an hence both given the fair chance to perform. Within the results of this calibration, it was clear to see that the results were particularly sensitive to some changes in hyper-parameter values and robust to others. Therefore the prior specification must be considered carefully in the context of the trial.

The models themselves also rely on a number of assumptions. For example, the monotonicity of both the activity and toxicity with dose. We have explored scenarios where the dose–activity relationship is both a plateau and bell-shaped, finding that the model-assisted method outperforms both model-based methods in these cases, although it should be noted that the model-based methods still perform well. Whilst the framework of both approaches could be adapted to
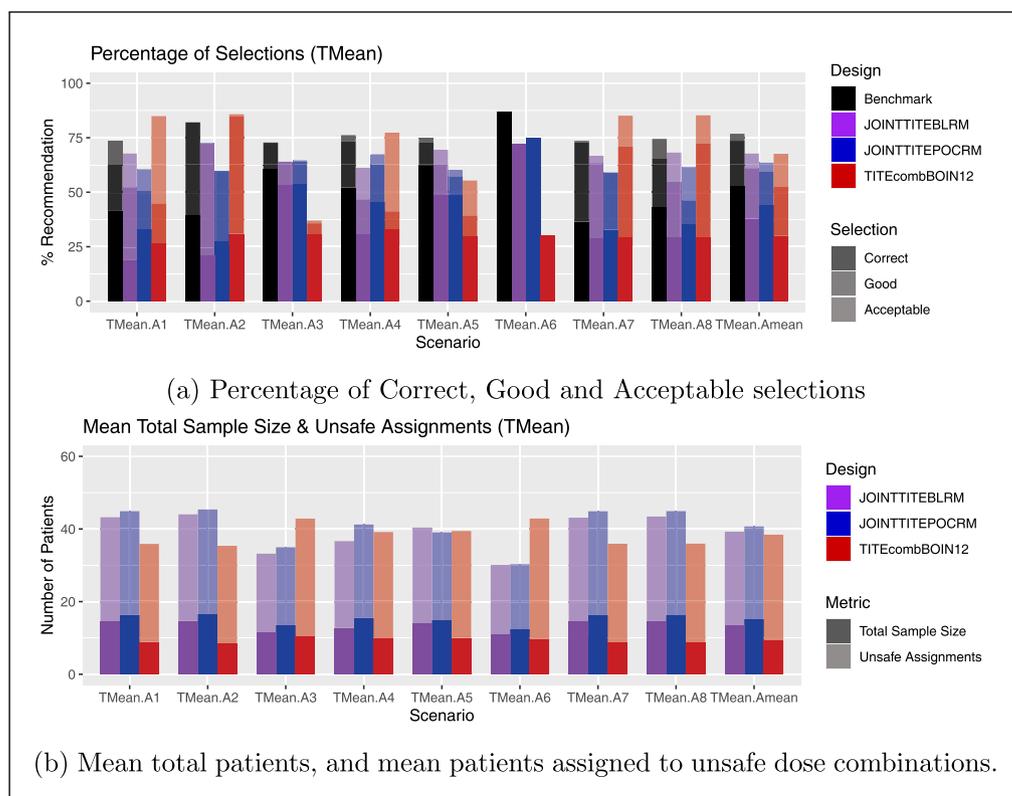
(a) Percentage of Correct, Good and Acceptable selections

(b) Mean total patients, and mean patients assigned to unsafe dose combinations.

Figure 8. Average results over toxicity scenarios, labelled TMean.A1–TMean.A8, including an average over all 48 scenarios, labelled TMean.AMean. (a) Percentage of Correct, Good and Acceptable selections; (b) Mean total patients, and mean patients assigned to unsafe dose combinations.

accommodate non-monotonicity, the way in which this should be done would depend upon any further assumptions on the dose–response relationship.

Whilst we endeavour to present a comprehensive study of the proposed methods, there are of course other extensions that could be considered. For example, sensitivity to the distribution of the TITE, number of cycles, cohort size and sample size could all be considered of interest and would be especially pertinent in application to a trial.

Funding

ORCID iDs

Helen Barnett ⓘD https://orcid.org/0000-0001-7466-7033
Thomas Jaki ⓘD https://orcid.org/0000-0002-1096-188X

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data availability

Software in the form of R code used to produce the presented results is available at https://github.com/helenyb/Dual_Agent_Joint_TITE_CRM

Supplemental material

Supplemental materials for this article are available online.

(a) Percentage of Correct, Good and Acceptable selections

(b) Mean total patients, and mean patients assigned to unsafe dose combinations.
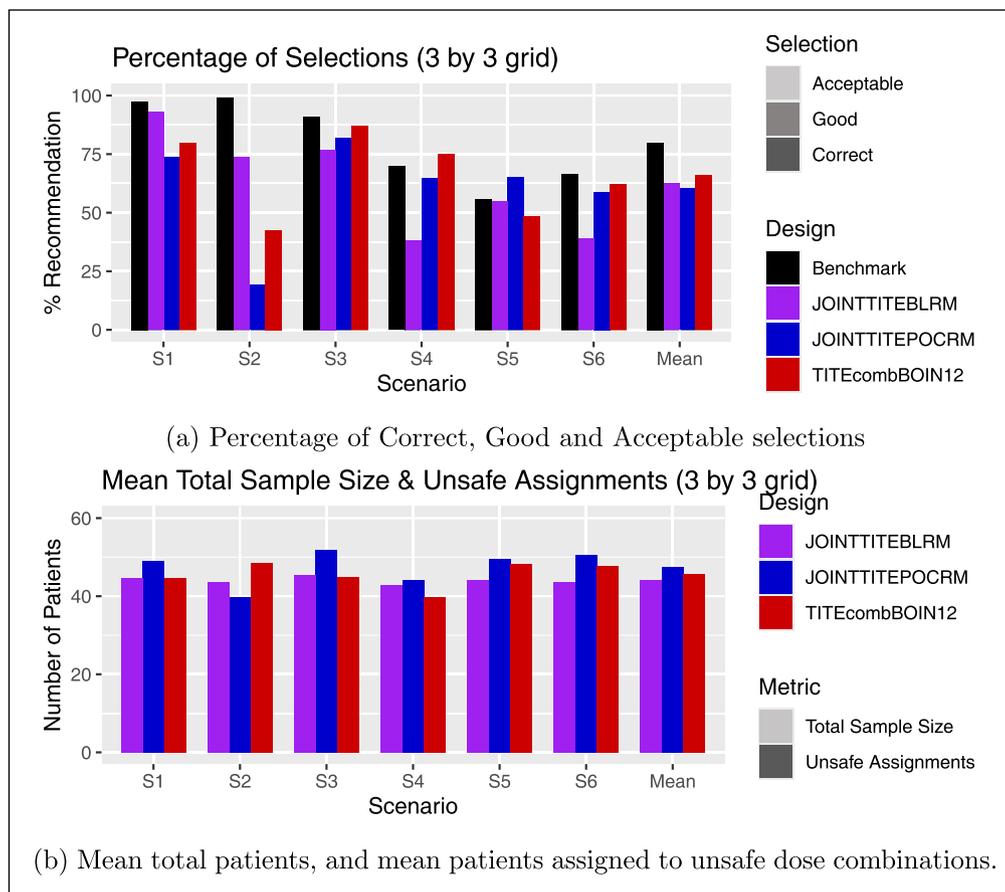
Figure 9. Results for the scenarios with 3 by 3 grids, including an average over these scenarios. (a) Percentage of Correct, Good and Acceptable selections; (b) Mean total patients, and mean patients assigned to unsafe dose combinations.

References

1. Lin R and Yin G. Bayesian optimal interval design for dose finding in drug-combination trials. *Stat Methods Med Res* 2017; **26**: 2155–2167.
2. Pan H, Lin R, Zhou Y, et al. Keyboard design for phase I drug-combination trials. *Contemp Clin Trials* 2020; **92**: 105972.
3. Wages NA, Conaway MR and O'Quigley J. Continual reassessment method for partial ordering. *Biometrics* 2011; **67**: 1555–1563.
4. Zhang L and Yuan Y. A practical bayesian design to identify the maximum tolerated dose contour for drug combination trials. *Stat Med* 2016; **35**: 4924–4936.
5. Neuenschwander B, Branson M and Gsponer T. Critical aspects of the Bayesian approach to phase I cancer trials. *Stat Med* 2008; **27**: 2420–2439.
6. Barnett H, George M, Skanji D, et al. A comparison of model-free phase I dose escalation designs for dual-agent combination therapies. *Stat Methods Med Res* 2024; **33**: 203–226.
7. Wages NA, Conaway MR and O'Quigley J. Using the time-to-event continual reassessment method in the presence of partial orders. *Stat Med* 2013; **32**: 131–141.
8. US Food & Drug Administration. Project optimus. https://www.fda.gov/about-fda/oncology-center-excellence/project-optimus (2024, accessed 2 December 2024).
9. Liu S and Johnson VE. A robust Bayesian dose-finding design for phase I/II clinical trials. *Biostatistics* 2016; **17**: 249–263.
10. Yan D, Tait C, Wages NA, et al. Generalization of the time-to-event continual reassessment method to bivariate outcomes. *J Biopharm Stat* 2019; **29**: 635–647.
11. Yuan Y and Yin G. Bayesian dose finding by jointly modelling toxicity and efficacy as time-to-event outcomes. *J R Stat Soc Ser C (Appl Stat)* 2009; **58**: 719–736.
12. Barnett H, Boix O, Kontos D, et al. Joint tite-CRM: a design for dose finding studies for therapies with late-onset safety and activity outcomes. *Stat Biopharm Res* 2024; **17**: 149–160.
13. Wages NA and Conaway MR. Phase I/II adaptive design for drug combination oncology trials. *Stat Med* 2014; **33**: 1990–2003.

14. Shimamura F, Hamada C, Matsui S, et al. Two-stage approach based on zone and dose findings for two-agent combination phase I/II trials. *J Biopharm Stat* 2018; **28**: 1025–1037.

15. Yada S and Hamada C. A bayesian hierarchal modeling approach to shortening phase I/II trials of anticancer drug combinations. *Pharm Stat* 2018; **17**: 750–760.

16. Zhang J, Yan F, Wages NA, et al. Local continual reassessment methods for dose finding and optimization in drug-combination trials. *Stat Methods Med Res* 2023; **32**: 2049–2063.

17. Kakizume T, Takeda K, Taguri M, et al. Boin-etc: a Bayesian optimal interval design considering efficacy and toxicity to identify the optimal dose combinations. *Stat Methods Med Res* 2024; **33**: 716–727.

18. Lu M, Zhang J, Yuan Y, et al. Comb-boin12: a utility-based Bayesian optimal interval design for dose optimization in cancer drug-combination trials. *Stat Biopharm Res* 2025; **17**: 266–276.

19. Jaki T, Barnett H, Titman A, et al. A seamless phase I/II platform design with a time-to-event efficacy endpoint for potential covid-19 therapies. *Stat Methods Med Res* 2024; **33**: 2115–2130.

20. Bidkar AP, Zerefa L, Yadav S, et al. Actinium-225 targeted alpha particle therapy for prostate cancer. *Theranostics* 2024; **14**: 2969.

21. Bayer. A Phase 1, open-label, first-in-human, multi-center, study to evaluate the safety, tolerability, pharmacokinetics, and anti-tumor activity of a thorium-227 labeled antibody-chelator conjugate, BAY 2315497 Injection alone, and in combination with darolutamide (BAY 1841788), in patients with metastatic castration resistant prostate cancer. NCT03724747. https://clinicaltrials.bayer.com/study/19445/, 2024.

22. US National Library of Medicine. Study to Evaluate the Safety, Tolerability, Pharmacokinetics, and Antitumor Activity of a Thorium-227 Labeled Antibody-chelator Conjugate Alone and in Combination With Darolutamide, in Patients With Metastatic Castration Resistant Prostate Cancer. NCT03724747. https://beta.clinicaltrials.gov/study/NCT03724747, 2024.

23. Neuenschwander B, Matano A, Tang Z, et al. A Bayesian industry approach to phase I combination trials in oncology. In: *Statistical methods in drug combination studies*. Boca Raton, FL: Chapman & Hall/CRC Press, 2015, pp.95–135.

24. Zhou Y, Lin R, Lee JJ, et al. Tite-boin12: a Bayesian phase I/II trial design to find the optimal biological dose with late-onset toxicity and efficacy. *Stat Med* 2022; **41**: 1918–1931.

25. Barnett H, Boix O, Kontos D, et al. Dose finding studies for therapies with late-onset toxicities: a comparison study of designs. *Stat Med* 2022; **41**: 5767–5788.

26. Cheung YK and Chappell R. Sequential designs for phase I clinical trials with late-onset toxicities. *Biometrics* 2000; **56**: 1177–1182.

27. Mozgunov P, Paoletti X and Jaki T. A benchmark for dose-finding studies with unknown ordering. *Biostatistics* 2021; **23**: 721–737.

28. Wages NA and Varhegyi N. POCRM: an R-package for phase i trials of combinations of agents. *Comput Methods Programs Biomed* 2013; **112**: 211–218.