A Bayesian adaptive design for dual-agent phase I-II oncology trials integrating efficacy data across stages

Received: 17 October 2022 | Revised: 19 April 2023 | Accepted: 20 April 2023

DOI: 10.1002/bimj.202200288

Biometrical Journal

RESEARCH ARTICLE

A Bayesian adaptive design for dual-agent phase I–II oncology trials integrating efficacy data across stages

José L. Jiménez[1] | **Haiyan Zheng**[2]

[1]Global Drug Development, Novartis Pharma AG, Basel, Switzerland

[2]MRC Biostatistics Unit, University of Cambridge, Cambridge, UK

Correspondence
José L. Jiménez, Global Drug Development, Novartis Pharma AG, Basel, Switzerland.
Email: jose_luis.jimenez@novartis.com

Present Address José L. Jiménez, Fabrikstrasse 2, 4056 Basel, Switzerland.

Funding information
Cancer Research UK, Grant/Award Number: RCCPDF/100008

This article has earned an open data badge "**Reproducible Research**" for making publicly available the code necessary to reproduce the reported results. The results reported in this article could fully be reproduced.

Abstract

Combination of several anticancer treatments has typically been presumed to have enhanced drug activity. Motivated by a real clinical trial, this paper considers phase I–II dose finding designs for dual-agent combinations, where one main objective is to characterize both the toxicity and efficacy profiles. We propose a two-stage Bayesian adaptive design that accommodates a change of patient population in-between. In stage I, we estimate a maximum tolerated dose combination using the escalation with overdose control (EWOC) principle. This is followed by a stage II, conducted in a new yet relevant patient population, to find the most efficacious dose combination. We implement a robust Bayesian hierarchical random-effects model to allow sharing of information on the efficacy across stages, assuming that the related parameters are either exchangeable or nonexchangeable. Under the assumption of exchangeability, a random-effects distribution is specified for the main effects parameters to capture uncertainty about the between-stage differences. The inclusion of nonexchangeability assumption further enables that the stage-specific efficacy parameters have their own priors. The proposed methodology is assessed with an extensive simulation study. Our results suggest a general improvement of the operating characteristics for the efficacy assessment, under a conservative assumption about the exchangeability of the parameters a priori.

KEYWORDS

drug combination, information borrowing, meta-analytic-combined, phase I–II, seamless designs

1 | INTRODUCTION

The primary objective of early-phase clinical trials is to identify a dose that is safe and efficacious. Seamless phase I–II clinical trial designs are efficient approaches to study these two aspects in a single protocol. In the literature, we find two types of seamless phase I–II designs: one-stage and two-stage. The former usually estimates the joint probability of toxicity and efficacy using the accumulating data to recommend a best-suited dose to patients in the next cohort (Ivanova, 2003;

Liu et al., 2018; Thall & Cook, 2004; Yuan & Yin, 2009). This setting is favored when the efficacy outcome can be observed relatively soon after administration of the dose, for example, after one or two cycles of therapy. By contrast, two-stage phase I–II designs come into play when efficacy cannot be ascertained in a short period of time. Specifically, stage I would commonly focus on toxicity considerations alone for dose (de-)escalation, despite that efficacy data are collected. This is then followed by stage II, where the evaluation of efficacy is the priority. A considerable amount of statistical literature has been written for the two-stage type (Jiménez et al., 2020; Jiménez & Tighiouart, 2022; Le Tourneau et al., 2009; Rogatko et al., 2008; Tighiouart, 2019).

Our work is motivated by the cisplatin–cabazitaxel clinical trial design proposed by Tighiouart (2019). In their proposed phase I–II study design, stage I was inspired by a conducted phase I trial Lockhart et al. (2014) in patients with advanced solid tumors, where a single maximum tolerated dose (MTD) of cisplatin/cabazitaxel 15/75 mg/m^2 was recommended. Based on the available results and other preliminary efficacy data, it was hypothesized by the clinical team that there could be a series of tolerable and efficacious dose combinations for prostate cancer. Over the last years, clinical trials with drug combinations have received a fair amount of attention. This interest is motivated by the fact that drug combinations are able to induce synergistic treatment effects by simultaneously inhibiting resistance mechanisms and targeting multiple pathways. In stage II of the cisplatin–cabazitaxel trial (Tighiouart, 2019), 30 additional patients were enrolled to identify the dose combinations with high probability of efficacy, along the MTD curve estimated from stage I data. Another characteristic of the cisplatin–cabazitaxel trial (Tighiouart, 2019) was that each stage has a different patient population, being the population in stage II the one of interest for the clinical team. Then, since the patient populations were not exactly the same in stages I and II, it was hypothesized that the dose-efficacy profiles could differ. Consequently, the dose-efficacy relationship was estimated using stage II data alone. The cisplatin–cabazitaxel trial (Tighiouart, 2019) can further be improved in two ways: (i) uncertainty about the estimated MTD curve should better be taken into account in stage II (i.e., the MTD curve may further be updated during stage II) and (ii) efficacy data from stage I could be used for the final analysis by the end of stage II. The first limitation was recently addressed (Jiménez & Tighiouart, 2022), allowing for a continuous update of the MTD curve throughout the entire phase I–II design. The novelty of the present article lies in addressing the second aspect; that is, we aim to integrate the efficacy information from both stages without neglecting the potential heterogeneity caused by the change of patient population across stages.

A robust Bayesian hierarchical model is fitted to allow combining the efficacy data across stages. The associated parameters are assumed to be either exchangeable or nonexchangeable. In our specific application, the benefit of borrowing is expected to lead to a precision improvement of the model parameter estimates and a reduction number of patients treated at subtherapeutic dose combinations when there is a consistency between the efficacy profiles across stages. For cases of data inconsistency, the stage I efficacy data need to be discounted effectively.

The manuscript is organized as follows. In Section 2, we review the cisplatin–cabazitaxel trial, which serves as a motivating example for the present work, as well as the proposed marginal dose-toxicity and dose-efficacy models to fit the trial. In Section 3, we introduce the proposed dose-finding algorithm for stages I and II, whereas in Section 4, we present a simulation study to evaluate the operating characteristics of the design, with focus on stage II. We provide concluding remarks in Section 5.

2 | MOTIVATING EXAMPLE AND STATISTICAL MODELS

2.1 | The cisplatin–cabazitaxel trial and data collection

The original cisplatin–cabazitaxel trial enrolled patients with metastatic, castration resistant prostate cancer. A combination of continuous doses ranging from 10 to 25 mg/m^2 for cisplatin and from 50 to 100 mg/m^2 for cabazitaxel were administered intravenously every 3 weeks. As informed by a precedent study (Lockhart et al., 2014), three specific combinations of cisplatin/cabazitaxel, 15/75, 20/75, and 25/75 mg/m^2 were evaluated. In stage I, the study enrolls 30 patients using conditional escalation with overdose control (EWOC) algorithm (Tighiouart et al., 2017) to estimate the MTD curve. In stage II, the study enrolls another 30 patients from the same population of patients but with visceral metastasis to identify dose combinations with high probability of efficacy along the MTD curve estimated at the end of stage I. These patients are allocated to dose combinations along the MTD curve using a Bayesian adaptive design after modeling the dose-efficacy curve with cubic splines.

The recommended MTD was 15/75 mg/m^2 on the basis of data from 24 patients (i.e., nine evaluable patients in phase I and 15 patients in the expansion cohort) where only 2 out of 18 patients treated at the recommended MTD had dose

limiting toxicity (DLT). Considering the low toxicity rate at the MTD reported (Lockhart et al., 2014), as well as other (unpublished) preliminary efficacy data, the clinicians who contributed to the design of the cisplatin–cabazitaxel trial hypothesized that a series of tolerable dose combinations that could be efficacious in prostate cancer could exist.

In this article, we regard the potential differences between stage I and stage II efficacy profiles from a different perspective. More specifically, we are motivated to establish a robust model formally accounting for such uncertainty, so that we can enhance the conduct and analysis of the stage II when there is a certain level of similarity between the efficacy profiles across the patient populations (the stages), as well as to discount the stage I efficacy data in case of dissimilarity.

2.2 | Problem formulation

Let x and y be the respective dose levels, on their original continuous scales, of two compounds (labeled X and Y) of interest, and further, $\{X_{\min}, Y_{\min}, X_{\max}, Y_{\max}\}$ be the lower and upper bounds. The measurement scales of x and y might differ from each other substantially. To avoid one variable being overly influential in the risk of toxicity, we standardize the doses using the transformations $h_1(x) = (x - X_{\min})/(X_{\max} - X_{\min})$ and $h_2(y) = (y - Y_{\min})/(Y_{\max} - Y_{\min})$, so that the standardized doses fall within the interval of $[0, 1]$. Thus, the dose combination (0,0) corresponds to the lowest dose combination available in the trial and not to a lack of dose combination administration. For ease of notation, we retain the notation of x and y to denote the standardized dose levels.

Let $Z \in \{0, 1\}$ be the binary indicator of DLT where $Z = 1$ represents the presence of a DLT and $Z = 0$ otherwise. Likewise, let $E \in \{0, 1\}$ be the binary indicator of treatment response where $E = 1$ represents a positive response, and $E = 0$ otherwise. In this article, following the motivating trial, we assume that only the DLT can be observed rapidly after drug administration (e.g., after one cycle of therapy), whereas it takes three cycles or more for the efficacy outcome to be observable. Following Tighiouart (2019), let $\theta_T = 0.33$ be the target probability of DLT and $p_0 = 0.15$ be the probability of efficacy of the standard of care treatment. When employing synergistic cytotoxic agents, it is common to assume that both the dose-toxicity and dose-efficacy relationship are monotonically increasing functions. This implies that the optimal dose combination (i.e., the dose combination with most desirable benefit-risk trade-off) will lie in the MTD set, defined as $\mathcal{M} = \{(x, y) : P(Z = 1|x, y) = \theta_T\}$, that is, any dose combination (x, y) with probability of DLT equal to θ_T. A formal definition of the optimal dose combination is given in Section 3. Given the two-stage formulation of this design, let $S \in \{1, 2\}$ be the stage enrollment indicator to stage I and stage II, respectively, and let $\boldsymbol{D}_{S,i} = \{(Z_i, E_i, x_i, y_i)\}$ be the data collected in stage S for the ith patient.

2.3 | A marginal dose-toxicity model

We assume that the binary outcomes of toxicity and efficacy are independent (Cai et al., 2014; Ivanova et al., 2009; Lyu et al., 2019). Alternatively, one could also account for the relationship between toxicity and efficacy either with the use of a copula (Thall & Cook, 2004) or with a latent variable approach (Liu et al., 2018; Lin et al., 2020). However, this would add an additional layer of complexity to the design that is not in the scope of the article. Let the model for the marginal probability of DLT be

$$\pi_T(x, y) = P(Z = 1|x, y) = F(\alpha_0 + \alpha_1 x + \alpha_2 y + \alpha_3 xy), \quad \alpha_1, \alpha_2, \alpha_3 > 0, \tag{1}$$

where $F(.)$ is the cumulative distribution function of the logistic distribution, that is, $F(u) = 1/(1 + e^{-u})$. The parameters in this model can be interpreted as follows: (i) α_0 determines the probability of DLT at the lowest dose combination available in the trial, that is, $(x = 0, y = 0)$, (ii) α_1 and α_2 determine the contribution of compounds X and Y to the overall probability of DLT, and (iii) α_3 captures the potential increase in the probability of DLT due to drug–drug interaction.

Note that in Model (1), because the number of attributable DLTs is expected to be very low given the cytotoxic nature of cisplatin and cabazitaxel, we do not take into account toxicity attributions (Jimenez et al., 2019).

We reparameterize the marginal probability of DLT defined in Model (1) in terms of parameters that clinicians can easily interpret (Tighiouart et al., 2017). Let ρ_{uv} denote the joint probability of DLT when the levels of agents $X = u$ and $Y = v$, with $u \in \{0, 1\}$, and $v \in \{0, 1\}$, so that $\alpha_0 = F^{-1}(\rho_{00})$, $\alpha_1 = (F^{-1}(\rho_{10}) - F^{-1}(\rho_{00}))$, and $\alpha_2 = (F^{-1}(\rho_{01}) - F^{-1}(\rho_{00}))$.

The MTD thus has the form of

$$\mathcal{M} = \left\{ (x, y) : y = \frac{(F^{-1}(\theta_T) - F^{-1}(\rho_{00})) - (F^{-1}(\rho_{10}) - F^{-1}(\rho_{00}))x}{(F^{-1}(\rho_{01}) - F^{-1}(\rho_{00})) + \alpha_3 x} \right\}. \tag{2}$$

Following Tighiouart (2019), we use informative prior distributions based on the results of Lockhart et al. (2014) so that ρ_{10}, ρ_{01}, and α_3 are independent a priori with $\rho_{01} \sim \text{Beta}(1.4, 5.6)$, $\rho_{10} \sim \text{Beta}(1.4, 5.6)$, and conditional on (ρ_{01}, ρ_{10}), $\rho_{00}/\min(\rho_{01}, \rho_{10}) \sim \text{Beta}(0.8, 7.2)$. Also, let the interaction parameter $\alpha_3 \sim \text{Gamma}(0.8, 0.0384)$, that is, a shape of 0.8 and a rate of 0.0384. These prior distributions imply that the combination of cisplatin/cabazitaxel 15/75 mg/m^2 has a probability of DLT approximately equal to 0.33. The posterior distribution of the dose-toxicity model parameters is defined as $p(\boldsymbol{\rho}, \alpha_3 | \boldsymbol{D}_1) \propto p(\boldsymbol{\rho}, \alpha_3) \times \mathcal{L}(\boldsymbol{D}_1 | \boldsymbol{\rho}, \alpha_3)$, where $\boldsymbol{\rho} = \{\rho_{00}, \rho_{01}, \rho_{10}\}$ and where $\boldsymbol{D}_1$ corresponds to the data from stage 1.

2.4 | A marginal dose-efficacy model

We now shift our focus to estimate the dose-efficacy relationship in a dual-dimensional plane; for stage $S = \{1, 2\}$, we stipulate the stagewise dose-efficacy data model as

$$\pi_E^S(x, y) = P(E = 1 | x, y, S) = F(\beta_{0S} + \exp(\beta_{1S})x + \exp(\beta_{2S})y + \beta_{3S}xy), \tag{3}$$

where $F(.)$ remains to be the cumulative distribution function of the logistic distribution. Because the motivating trial employs cytotoxic agents, we assume that the probability of efficacy does not decrease with the dose of any agent when the other agent is held constant. To ensure this property, we apply the exponential function to β_{1S} and β_{2S} since $\exp(u) > 0$. We also assume that $\beta_{3S} > 0$, which means that there is a synergistic effect due to the interaction of the two compounds (Gasparini, 2013). The parameters in this model can be interpreted as follows: (i) β_0 determines the probability of efficacy at the lowest dose combination available in the trial, that is, $(x = 0, y = 0)$, (ii) β_1 and β_2 determine the contribution of compounds X and Y to the overall probability of efficacy, and (iii) β_3 captures the potential increase in the probability of efficacy due to drug–drug interaction.

Let $\boldsymbol{\Psi}_S = (\beta_{1S}, \beta_{2S})$ denote the main effects of the treatment specific to stage. We consider a *meta-analytic-combined* (MAC) approach (Neuenschwander et al., 2016) to establish a Bayesian predictive distribution for $\boldsymbol{\Psi}_2 \mid \boldsymbol{D}_1, \boldsymbol{D}_2$. This would allow the investigator to estimate main effects of drugs X and Y, using the efficacy data from both stages.

We assume a normal-normal hierarchical model to relate the stagewise main effects of efficacy for the dual agent. Specifically, at stage $S = 1$,

$$\boldsymbol{\Psi}_1 \mid \boldsymbol{\mu}, \boldsymbol{\Phi} \sim \text{BVN}(\boldsymbol{\mu}, \boldsymbol{\Phi}).$$

Continuing the phase I–II trial to stage $S = 2$ in a new population, we introduce a nonexchangeability distribution and stipulate that

$$\begin{aligned} \boldsymbol{\Psi}_2 \mid \boldsymbol{\mu}, \boldsymbol{\Phi} &\sim \text{BVN}(\boldsymbol{\mu}, \boldsymbol{\Phi}) \qquad \text{with probability } \omega, \\ \boldsymbol{\Psi}_2 &\sim \text{BVN}(\boldsymbol{m}_0, R_0) \qquad \text{with probability } 1 - \omega, \end{aligned} \tag{4}$$

with

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \qquad \boldsymbol{\Phi} = \begin{pmatrix} \tau_1^2 & \xi\tau_1\tau_2 \\ \xi\tau_1\tau_2 & \tau_2^2 \end{pmatrix}, \tag{5}$$

and

$$\mathbf{m}_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \qquad \mathbf{R}_0 = \begin{pmatrix} 10^2 & \zeta 100 \\ \zeta 100 & 10^2 \end{pmatrix}. \tag{6}$$

The variance terms in $\boldsymbol{\Phi}$ represent between-stage heterogeneity. Mind that in each stage of the cisplatin–cabazitaxel trial, we have two distinct and well-designated populations. Consequently, between-stage heterogeneity and between-population heterogeneity cannot be disentangled. If our trial would involve multiple populations in each stage, we would need additional random-effects distributions to account for the between-population differences (Zheng et al., 2020, 2021).

The values of $\mathbf{m}_0$ and $\mathbf{R}_0$ are selected so that they induce weakly informative prior distributions over the parameters in $\boldsymbol{\Psi}_2$. This Bayesian hierarchical random-effects model is completed by the following hyperpriors:

$$\mu_1 \sim N(\eta_1, s_1^2), \quad \mu_2 \sim N(\eta_2, s_2^2), \quad \tau_1 \sim \text{HN}(z_1), \quad \tau_2 \sim \text{HN}(z_2), \quad \xi \sim U(0, 0.5), \quad \zeta \sim U(0, 0.5),$$

where HN(z) denotes a half-normal distribution formed by truncating a $N(0, z^2)$, so it covers the interval $(0, \infty)$. We select HN(0.5) anticipating for substantial between-stage heterogeneity in the main effects model parameters. Other viable choices of HN(z) and the indication have been noted (Roychoudhury & Neuenschwander, 2020; Zheng et al., 2020). The value $z = 0.5$ serves as a weakly informative prior distribution, although the values of z_1 and z_2 can be justified appropriately for the user's own case, with evidence suggesting the similarity or dissimilarity of efficacy in such two-patient populations.

The specification of ω requires, in practice, the input of subject-matter experts and needs to be fixed a priori. We place weakly informative prior distributions over the dose-efficacy model parameters:

$$\beta_{01} \sim N(-1.8, 3.16^2), \quad \beta_{02} \sim N(-1.8, 3.16^2),$$
$$\beta_{31} \sim \text{Gamma}(0.1, 0.1), \quad \beta_{32} \sim \text{Gamma}(0.1, 0.1).$$

For illustration purposes, we set

$$\eta_1 = 0, \quad s_1 = 3.16, \quad \eta_2 = 0, \quad s_2 = 3.16, \quad z_1 = 0.5, \quad z_2 = 0.5,$$

to implement the model. Overall, the weakly informative prior distributions we select in this article translate into the median probability of efficacy estimates with 95% credible intervals displayed in Table S1 of the Supporting Information.

3 | AN INTEGRATED PHASE I–II DESIGN FOR DOSE FINDING

Stage I will enroll a total of $N_1 = C_1 \times m_1$ patients, where C_1 denotes the total number of cohorts in phase I with each of the size m_1. Stage II will enroll a total of $N_2 = n_2 + C_2 \times m_2$ patients, where n_2 is the number of patients in the first cohort of stage II, C_2 the additional number of cohorts, and m_2 its size.

In the original cisplatin–cabazitaxel trial, stage I efficacy data were entirely discarded, and therefore, at the beginning of stage II, an initial cohort n_2 was used to collect efficacy data homogeneously across the entire MTD curve. In this article, we choose to keep n_2 as a short run-in period that can inform the data (in)consistency, and thus, determine the degree of information sharing.

Let $N = N_1 + N_2$ be the total number of patients that the entire study will enroll and $\widehat{\mathcal{M}}_{\boldsymbol{D}_1}$ be the estimated MTD set based on data from stage I. We select $m_1 = 2$, $m_2 = 5$, and $n_2 = 10$. At the end of stage II, we test the following null and alternative hypotheses

$$\begin{aligned} H_0 &: \pi_E^{S=2}(x, y) \leq p_0 \text{ for all } (x, y) \in \widehat{\mathcal{M}}_{\boldsymbol{D}_1} \quad \text{versus} \\ H_1 &: \pi_E^{S=2}(x, y) > p_0 \text{ for some } (x, y) \in \widehat{\mathcal{M}}_{\boldsymbol{D}_1}, \end{aligned} \tag{7}$$

and we reject the null hypothesis if

$$\underset{(x,y)\in\widehat{\mathcal{M}}_{\boldsymbol{D}_1}}{\arg\max} \left(P(\pi_E^{S=2}(x, y) > p_0 | \boldsymbol{D}_1, \boldsymbol{D}_2) \right) > \delta_u, \tag{8}$$

where $\delta_u = 0.4$ is a prespecified design parameter. Moreover, the dose combination

$$(x, y)_{opt} = \underset{(x,y)\in\widehat{\mathcal{M}}_{\boldsymbol{D}_1}}{\text{argmax}} \, P(\pi_E^{S=2}(x, y) > p_0 | \boldsymbol{D}_1, \boldsymbol{D}_2), \tag{9}$$

is recommended as the optimal dose combination and is selected for further phase IIb or III studies.

ALGORITHM 1 Stage I and stage II algorithms

STAGE I

- In the first cohort ($c_1 = 1$) patients 1 an 2 receive the dose combination $(x_1, y_1) = (x_2, y_2) = (0.33, 0.5)$

- In the second cohort ($c_1 = 2$) patients 3 and 4 receive doses (x_3, y_3) and (x_4, y_4), respectively, where $y_3 = y_1$, $x_4 = x_2$, x_3 is the αth percentile of $\lambda(\Gamma_{X|Y=y_1}|\boldsymbol{D}_{1,2})$, y_4 is the αth percentile of $\lambda(\Gamma_{Y|X=x_2}|\boldsymbol{D}_{1,2})$.

for $c_1 = 3 : C_1$ **do**

if c_1 is an even number **then**

- Patient $2c_1 - 1$ receives the dose combination (x_{2c_1-1}, y_{2c_1-3}) where $x_{2c_1-1} = \Lambda^{-1}_{\Gamma_{X|Y=y_{2c_1-3}}}(\alpha|\boldsymbol{D}_{1,2c_1-2})$

- Patient $2c_1$ the dose combination (x_{2c_1-2}, y_{2c_1}), where $y_{2c_1} = \Lambda^{-1}_{\Gamma_{Y|X=x_{2c_1-2}}}(\alpha|\boldsymbol{D}_{1,2c_1-2})$

else

- Patient $2c_1 - 1$ receives the dose combination (x_{2c_1-3}, y_{2c_1-1}) where $y_{2c_1-1} = \Lambda^{-1}_{\Gamma_{Y|X=x_{2c_1-3}}}(\alpha|\boldsymbol{D}_{1,2c_1-2})$

- Patient $2c_1$ receives the dose combination (x_{2c_1}, y_{2c_1-2}) where $x_{2c_1} = \Lambda^{-1}_{\Gamma_{X|Y=y_{2c_1-2}}}(\alpha|\boldsymbol{D}_{1,2c_1-2})$

end if

end for

STAGE II

- Calculate the posterior median of the parameters $\rho_{00}, \rho_{10}, \rho_{01}$ and α_3 given data $\boldsymbol{D}_1$, that is, $(\widehat{\rho}_{00}, \widehat{\rho}_{10}, \widehat{\rho}_{01}, \widehat{\alpha}_3)$.

- Calculate estimated MTD set $\widehat{\mathcal{M}}_{\boldsymbol{D}_1} = \{(x, y) : y = (\frac{F^{-1}(\theta_Z) - F^{-1}(\widehat{\rho}_{00}) - (F^{-1}(\widehat{\rho}_{10}) - F^{-1}(\widehat{\rho}_{00}))x}{(F^{-1}(\widehat{\rho}_{01}) - F^{-1}(\widehat{\rho}_{00})) + \widehat{\alpha}_3 x})\}$

- Allocate n_2 patients to dose combinations equally spaced along the estimated MTD curve $\widehat{\mathcal{M}}_{\boldsymbol{D}_1}$.

- Calculate the posterior median of the parameters $(\widehat{\beta}_{02}, \widehat{\beta}_{12}, \widehat{\beta}_{22}, \widehat{\beta}_{32})$ using the MAC approach given data $\boldsymbol{D}_1, \boldsymbol{D}_2$.

for $c_2 = 1 : C_2$ **do**

- Generate a sample of dose combinations of size m_2 that belong to $\widehat{\mathcal{M}}_{\boldsymbol{D}_1}$ from the (estimated) standardized density $\widehat{\pi}_E^{S=2}(x, y)$, and assign it to the subsequent cohort of m_2 patients.

- Calculate the posterior median of the parameters $(\widehat{\beta}_{02}, \widehat{\beta}_{12}, \widehat{\beta}_{22}, \widehat{\beta}_{32})$ using the MAC approach given data $\boldsymbol{D}_1, \boldsymbol{D}_2$.

end for

As previously mentioned, stage I is based on the EWOC principle (Babb et al., 1998; Shi & Yin, 2013; Tighiouart et al., 2005, 2010, 2017; Tighiouart & Rogatko, 2012) where the posterior probability of overdosing the next cohort of patients is bounded by a feasibility bound α. For the definition of the algorithm, let $\lambda(\Gamma_{X|Y=y}|\boldsymbol{D}_1)$ represent the posterior distribution of the MTD of drug X, given that the level of drug Y is equal to y (i.e., given that Y is fixed) based on stage I data $\boldsymbol{D}_1$ (see Equation (2) for the definition of the MTD). Also, let $\Lambda^{-1}_{\Gamma_{X|Y=y}}(\alpha|\boldsymbol{D}_1)$ denote the αth percentile of $\lambda(\Gamma_{X|Y=y}|\boldsymbol{D}_1)$. In a cohort with two patients, the first one would receive a new dose of compound X given that the dose y of compound Y that was previously assigned. The other patient would receive a new dose of compound Y, given that dose x of compound X was previously assigned. These steps are described in Stage I of Algorithm 1. Using EWOC, these new doses are at the αth percentile of the conditional posterior distribution of the MTD combinations. The feasibility bound α increases from 0.25 up to 0.5 in increments of 0.05 (see Wheeler et al., 2017). Accrual continues until the maximum sample size in stage I is reached or the trial is stopped early for safety.

Stage II follows the response-adaptive randomization principle. This type of Monte Carlo algorithm uses the current parameter estimates to sample a cohort of m_2 dose combinations from the estimated dose-efficacy standardized density of $\widehat{\pi}_E^{S=2}(x, y)$ along the estimated MTD curve. Note that $\widehat{\pi}_E^{S=2}(x, y)$ uses the Bayes estimates of the dose-efficacy model parameters. As explained in Jiménez and Tighiouart (2022), because stage II selects doses on the estimated MTD curve $\widehat{\mathcal{M}}_{\boldsymbol{D}_1}$, and there is a one-to-one correspondence between $(x, y) \in \widehat{\mathcal{M}}_{\boldsymbol{D}_1}$, we may write $\widehat{\pi}_E^{S=2}(x, y) = \widehat{\pi}_E^{S=2}(x)$ for $(x, y) \in \widehat{\mathcal{M}}_{\boldsymbol{D}_1}$. In other words, by having the value of x, using the definition of the MTD in Equation (2), we can easily obtain the corresponding value of y. Thus, to facilitate the definition of the standardize density function, instead of writing $\widehat{\pi}_E^{S=2}(x, y)$, we simply write $\widehat{\pi}_E^{S=2}(x)$. The standardized density of the estimated efficacy curve is $\tilde{\pi}_E^{S=2}(x) = \frac{\widehat{\pi}_E^{S=2}(x)}{\int_{x \in X'} \widehat{\pi}_E^{S=2}(x) dx}$. A rejection sampling algorithm is then used to sample m_2 dose combinations from this density. These steps are described in Stage II of Algorithm 1.

The dose finding algorithm contains the following stopping rules for safety and futility:

- **Futility stopping rule**
  For ethical considerations and to avoid exposing patients to subtherapeutic dose combinations, we would stop the trial for futility if

$$\underset{(x,y)\in\widehat{\mathcal{M}}_{\boldsymbol{D}_1}}{\arg\max}\ \left(\pi_E^{S=2}(x,y) > p_0|\boldsymbol{D}_1,\boldsymbol{D}_2)\right) < \delta_0,$$

  where δ_0 is a prespecified threshold. For the purposes of illustration in this article, we choose $\delta_0 = 0.1$. Mind that this stopping rule applies only after the run-in cohort of n_2 patients in stage II.
- **Safety stopping rule**
  The design contains two stopping rules for safety, one for stage I and a less stringent one for stage II. During stage I, we would stop the trial if

$$P(\pi_T(x=0,y=0) > (\theta_T + 0.1) \mid \boldsymbol{D}_1) > \delta_{\theta_1}, \tag{10}$$

  where $\delta_{\theta_1} = 0.5$. In contrast, during stage II, we would stop the trial if

$$P(\Theta > (\theta_T + 0.1) \mid \boldsymbol{D}_2) > \delta_{\theta_2}, \tag{11}$$

  where Θ represents the rate of DLTs for both stages of the design regardless of dose and $\delta_{\theta_2} = 0.9$ represents the confidence level (i.e., 90%) that a prospective trial results in an excessive DLT rate. A noninformative Jeffrey's prior Beta(0.5,0.5) is placed on the parameter Θ.

4 | SIMULATION STUDY

4.1 | Operating characteristics

In this section, we present a simulation study that will assess the operating characteristics of our design. Since we apply an already established dose-escalation procedure in stage I, we concentrate on evaluating the design's operating characteristics for the stage II, which leverages efficacy data from stage I. We report the simulation results according to the following metrics:

- Distribution of the recommended optimal dose combinations.
- Proportion of recommended optimal dose combinations with true probability of efficacy above p_0. For simplicity, this metric is referred as the *percentage of correct recommendation*.
- (Approximated) Bayesian power (or type-I error probability under H_0):

$$\text{Power} \approx \frac{1}{J}\sum_{j=1}^{J}\mathbf{1}\left\{\max_{(x,y)\in\widehat{\mathcal{M}}_{\boldsymbol{D}_1}}\left(P\left(\pi_{E,j}^{S=2}(x,y) > p_0|\boldsymbol{D}_1,\boldsymbol{D}_2\right)\right) > \delta_u\right\}, \tag{12}$$

  where "$\mathbf{1}(.)$" represents an indicator function and J represents the total number of simulated trials, and $\pi_{E,j}^{S=2} = F(\beta_{02}^{(j)} + \exp(\beta_{12}^{(j)})x + \exp(\beta_{22}^{(j)})y + \beta_{32}^{(j)}xy)$. Under null scenarios as defined in (7), the above formula represents the (approximated) Bayesian type-I error probability.
- Average posterior probability of early stopping for futility and safety.
- Proportion of patients in stage II allocated to dose combinations with true probability of efficacy above p_0.

4.2 | Scenarios

In stage I, we construct two dose-toxicity scenarios considered as highly plausible by the principal investigator of the motivating trial (Jiménez & Tighiouart, 2022; Tighiouart, 2019). The true dose-toxicity model parameters are presented

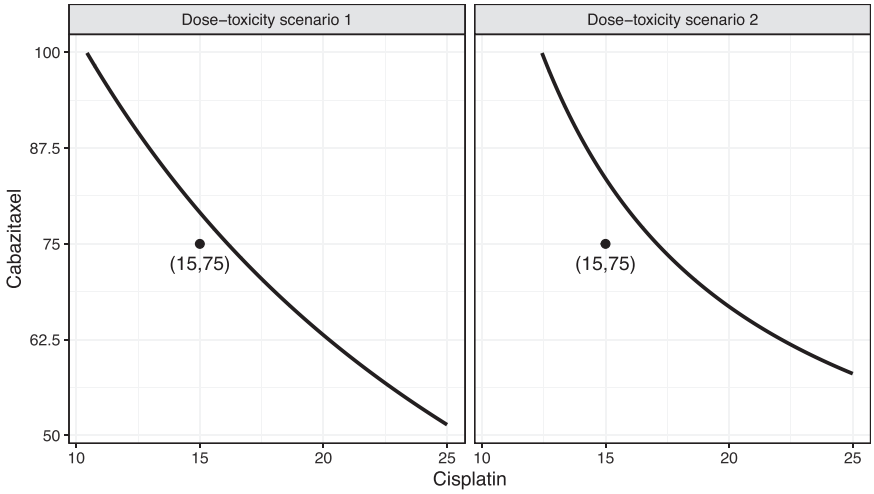

FIGURE 1 MTD curves obtained with the dose-toxicity model parameter values presented in Table S2 in the Supporting Information. The point at the cisplatin/cabazitaxel 15/75 mg/m^2 combination represents the MTD found by Lockhart et al. (2014).

in Table S2 in the Supporting Information, and displayed in Figure 1. Furthermore, we assume that stage II has the same dose-toxicity profile as stage I (i.e., the dose-toxicity profiles do not vary across patient populations) (Jiang et al., 2021; Jiménez et al., 2020; Tighiouart, 2019). The target probability of DLT is $\theta_T = 0.33$.

For each of the dose-toxicity scenarios, we consider two different stage II dose-efficacy profiles that place the dose combination with the highest efficacy in opposite locations. In terms of the stage I dose-efficacy profiles, we consider the following three hypothetical situations:

1. The stage I and stage II dose-efficacy profiles are perfectly consistent. For reading purposes, we refer to this profile as "complete agreement between stage I and stage II dose-efficacy profiles" or simply as "CA," which is short for "Complete Agreement."
2. The stage I and stage II dose-efficacy profiles point to the same dose combination with highest efficacy, but the probabilities of efficacy are different across stages. For reading purposes, we refer to this profile as "partial agreement between stage I and stage II dose-efficacy profiles" or simply as "PA," which is short for "Partial Agreement."
3. The stage I and stage II dose-efficacy profiles are completely different, and place the dose combination with the highest efficacy in different locations. For reading purposes, we refer to this profile as "complete disagreement between stage I and stage II dose-efficacy profiles" or simply as "CD," which is short for "Complete Disagreement."

To reflect low to high levels of prior confidence in the efficacy data consistency across stages, we run the simulations per scenario with the prior probability of exchangeability $\omega = 0, 0.25, 0.5, 0.75, 1$. For scenarios under the alternative hypothesis H_1, we assume an effect size of 0.25 (i.e., in all stage II dose-efficacy profiles, the highest probability of efficacy is equal to $p_0 + 0.25 = 0.4$, with $p_0 = 0.15$). For scenarios under H_0, the highest probability of efficacy in stage II is equal to p_0.

Overall, we have a large number of comparisons, given that for each dose-toxicity profile, there are two different stage II dose-efficacy profiles, each coupled with three different stage I dose-efficacy profiles (i.e., CA, PA, and CD). We have scenarios under H_1 and H_0, and furthermore, five different values of ω. To facilitate the communication of the simulation results over a large number of scenarios, we label the configuration of dose-toxicity and dose-efficacy profiles by scenarios A–H as follows:

- Dose-toxicity profile 1 + stage II dose-efficacy profile 1 under H_1 = scenario A,
- Dose-toxicity profile 1 + stage II dose-efficacy profile 2 under H_1 = scenario B,
- Dose-toxicity profile 2 + stage II dose-efficacy profile 1 under H_1 = scenario C,
- Dose-toxicity profile 2 + stage II dose-efficacy profile 2 under H_1 = scenario D,
- Dose-toxicity profile 1 + stage II dose-efficacy profile 1 under H_0 = scenario E,
- Dose-toxicity profile 1 + stage II dose-efficacy profile 2 under H_0 = scenario F,
- Dose-toxicity profile 2 + stage II dose-efficacy profile 1 under H_0 = scenario G,
- Dose-toxicity profile 2 + stage II dose-efficacy profile 2 under H_0 = scenario H.

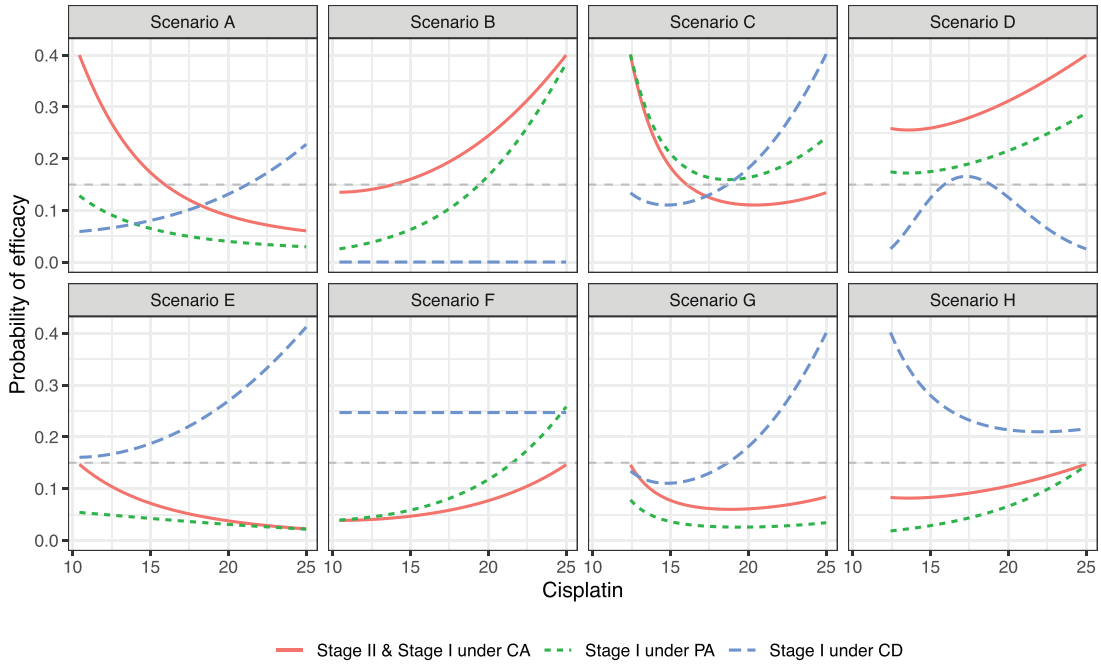

FIGURE 2 True dose-efficacy profiles favoring the alternative hypothesis H_1 under each dose-toxicity scenarios varying with the dose of cisplatin. In each efficacy scenario, we have the true stage II efficacy profile (red), and three-stage I efficacy scenarios: (i) one that is exactly like the stage II dose-efficacy profile (in red), (ii) one in which the optimal dose combination is the same but the efficacy profile is slightly different (green), and (iii) one that is completely different to the stage II dose-efficacy profile (blue). The gray line represents the threshold $p_0 = 0.15$.

The true dose-efficacy profile per scenario, with specification of model parameters, is given in Table S3 in the Supporting Information, and displayed graphically in Figure 2.

The sample sizes for Stages I and II are $N_1 = N_2 = 30$ and we simulated $J = 1000$ trials using Algorithm 1. The DLT and efficacy responses were generated from models (1) and (3), respectively.

4.3 | Results

As discussed in Section 2.4, the value $\omega = 0$ implements no borrowing of information. In other words, the treatment efficacy in Equation (4) would be estimated using data from one stage solely, leading to a complete discard of stage I efficacy data. It is of interest to quantify the improvement achieved by allowing the combination of efficacy data based on the assumption of full exchangeability (with $\omega = 1$) or partial exchangeability (with $0 < \omega < 1$).

In Figure 3, we display the power and type-I error values obtained at different values of $\omega > 0$ with respect to $\omega = 0$. Under H_0 (i.e., scenarios E-H), if we choose low to medium values of ω (e.g., $0 < \omega \leq 0.25$), the type-I error varies between 0 and 0.049 with respect to $\omega = 0$, depending on the scenario and the level of agreement. That is, the type-I error remains very close to its reference value (i.e., with $\omega = 0$). With larger values of ω (i.e., $\omega > 0.25$), the type-I error varies between 0.012 and 0.106 with respect to $\omega = 0$. Under H_1 (i.e., scenarios A–D), the power increases as the value of ω, with differences, with respect to $\omega = 0$, up to 0.121. We notice that for values of $\omega \leq 0.25$, the power gain is already notable. The numerical results of power and type-I error for all values of ω are presented in the Supporting Information (Figure S1). In scenarios under H_1 with $\omega = 0$, the power ranges between 0.66 and 0.93, whereas in scenarios under H_0 with $\omega = 0$, the type-I error ranges between 0.11 and 0.21. These power and type-I error results are consistent with those reported in previous publications (Jiménez et al., 2020; Jiménez & Tighiouart, 2022; Tighiouart, 2019).

In Figure 4, we present the distribution of the recommended optimal dose combinations across all scenarios, agreement levels, and values of ω. Overall, we see how the design correctly identifies the most efficacious region of the MTD curve by allocating there the majority of patients. At $\omega = 0$, we notice that the level of dispersion of the distribution is generally higher than with $\omega > 0$. As we increase the value of ω, the dispersion shrinks toward the mode of the distribution. This

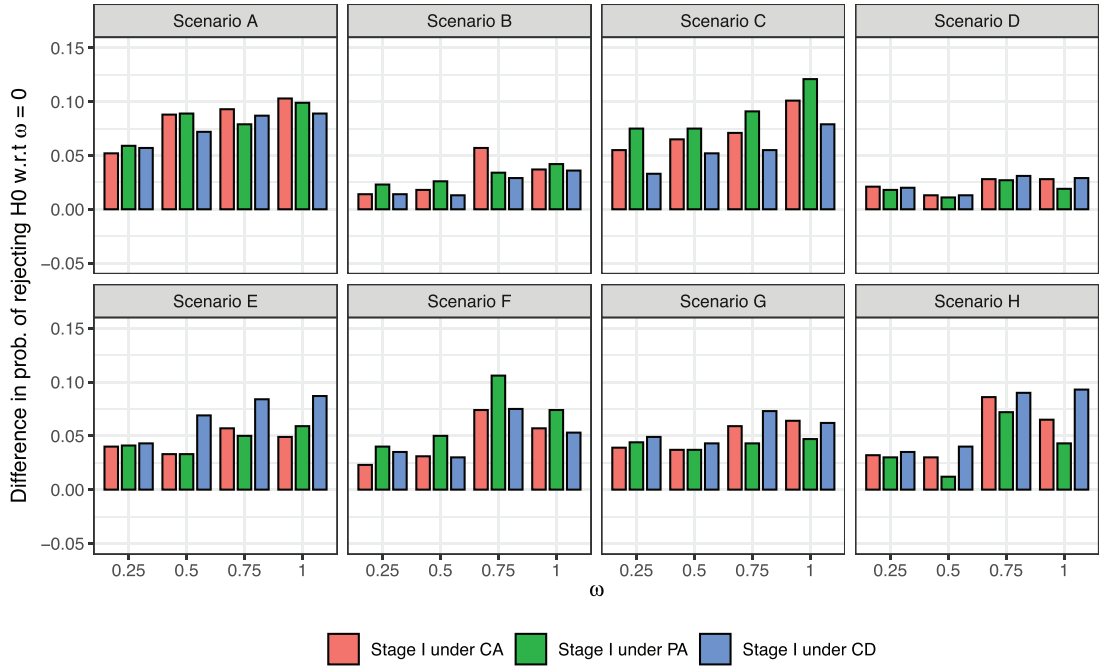

FIGURE 3 Differences in the probability of rejecting H_0 with respect to $\omega = 0$ in scenarios under the H_1 (i.e., power) and under the H_0 (i.e., type-I error). Scenarios A–D and E–H correspond to settings under H_1 and H_0, respectively.

behavior is manifested under the levels of CA and PA. These results are as expected: our model effectively discounts efficacy data from stage I if it is not consistent with the efficacy data observed in stage II.

In Figure 5, we present the difference in the recommended optimal dose combinations with true probability of efficacy above p_0 (also known as the proportion of correct recommendation), between models with $\omega > 0$ and $\omega = 0$. In settings with $\omega = 0$, the proportion of correct recommendation ranges between 81% and 100%, which is consistent with the values reported in previous publications (Jiménez et al., 2020; Jiménez & Tighiouart, 2022; Tighiouart, 2019). For values of $\omega > 0$, such proportion varies between −0.49% and 7.95%. In scenario D, the proportion of correct recommendation with $\omega = 0$ is already practically 100%, which remains the same for values of $\omega > 0$. This explains why the difference in the proportion of correct recommendation between $\omega = 0$ and $\omega > 0$ is approximately 0.

In Figure 6, we show the differences in terms of the probability of early stopping for futility, under both H_1 and H_0. At $\omega = 0$, the probability of early stopping for futility ranges between 0.012 and 0.126 under H_1, and between 0.478 and 0.612 under H_0, depending on the scenario. Under H_1, by increasing ω, we see that the probability of early stopping for futility varies between −0.058 and 0.008, with respect to $\omega = 0$. Under H_0, by increasing ω, we see that the probability of early stopping for futility varies between −0.167 and −0.060, with respect to $\omega = 0$. It is worth mentioning that in scenarios under H_1, the probabilities of early stopping for futility with $\omega = 0$ are already low, and therefore, it is reasonable that allowing for robust sharing of efficacy data across stages does not have a major impact on the probability of early stopping for futility. On the other hand, in scenarios under H_0, the probability of early stopping for futility with $\omega = 0$ is, as expected, higher and a big decrease would be problematic. However, we see that by selecting a conservative value of ω, such as $\omega = 0.25$, the decrease in the probability of early stopping is usually lower than 0.1.

In Table S5 of the Supporting Information, we show the average sample sizes obtained when applying this early stopping rule. Under H_1, results show that the observed decrease in the probability of early stopping for futility caused by the increment of ω translated into an average sample size increase of zero to one patients with respect to $\omega = 0$. Under H_0, the increment of ω translated into an average sample size increase of zero to two patients with respect to $\omega = 0$.

In Figure S2 of the Supporting Information, we present the difference in the proportion of patients allocated to dose combinations with true probability of efficacy above p_0 in stage II, between models with $\omega > 0$ and $\omega = 0$. With $\omega = 0$, the proportion of patients allocated to dose combinations with true probability of efficacy above p_0 in stage II ranges between 40% and 95%, which is consistent with the values reported by Jiménez et al. (2020), Jiménez and Tighiouart (2022), and Tighiouart (2019). As we increase the value of ω, this proportion increases between 0.25% and 6.17%.

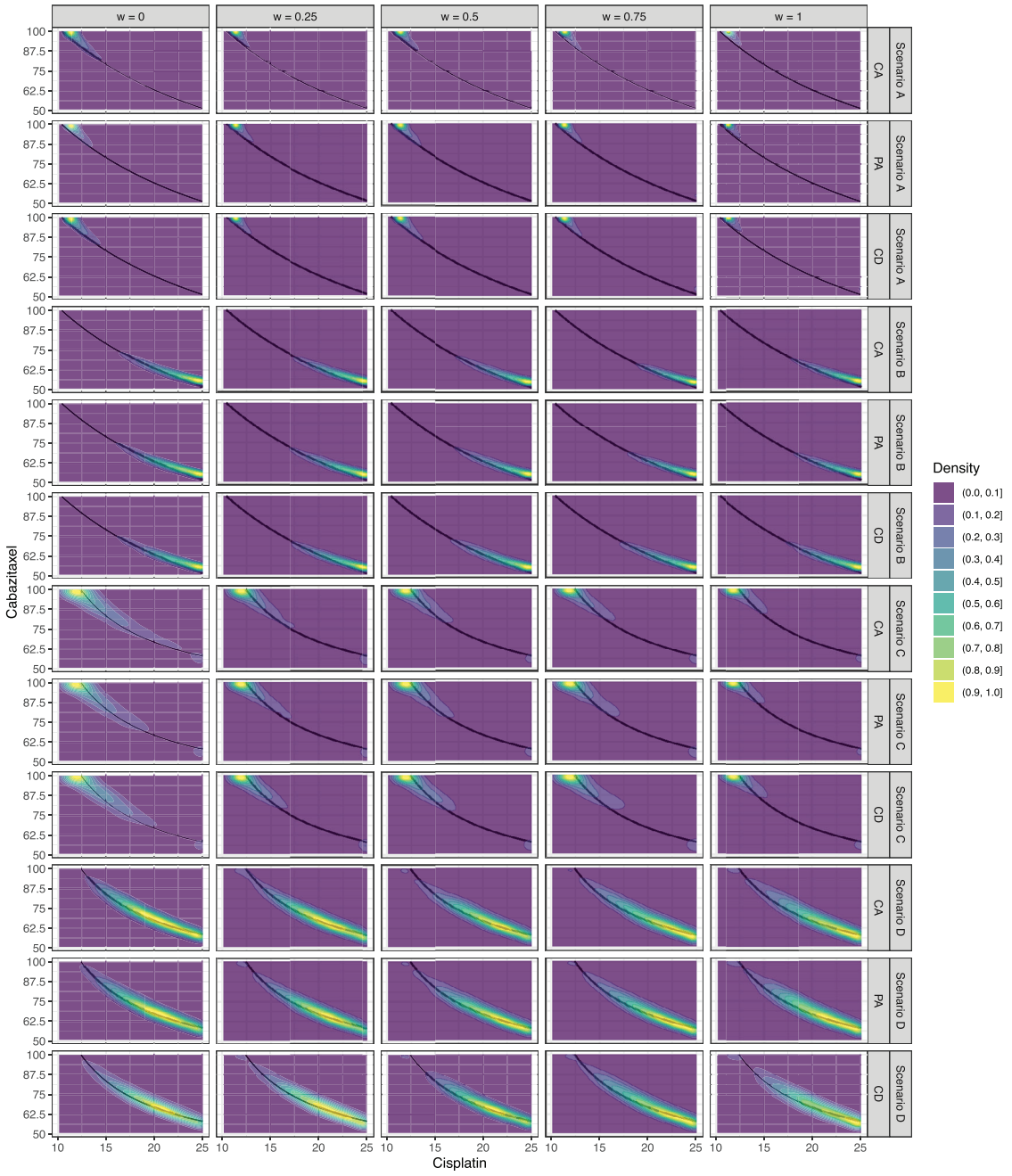

FIGURE 4 Distribution of the recommended optimal dose combinations in scenarios under the alternative hypothesis (A–D) and levels of agreement CA, PA, and CD. The black curve represents the MTD curve.

In terms of safety, we observe that scenarios A, B, E, and F (i.e., dose-toxicity scenario 1 in Figure 1) have an overall (i.e., stage I + stage II) average DLT rate between 27% and 35%, depending on the scenario, with an average proportion of trials with DLT rate above $\theta_T + 0.1$ of 0%. Stage II alone in these scenarios has an average DLT rate between 28% and 43%, with an average proportion of trials with DLT rate above $\theta_T + 0.1$ of 7% and 42%. Scenarios C, D, G, and H (i.e., dose-toxicity scenario 2 in Figure 1) have an overall (i.e., stage I + stage II) average DLT rate between 26% and 35%, depending on the scenario, with an average proportion of trials with DLT rate above $\theta_T + 0.1$ of 0%. Stage II alone in these scenarios has an average DLT rate between 28% and 43%, with an average proportion of trials with DLT rate above $\theta_T + 0.1$ between 7% and 43%. Because toxicity data are not shared across stages, these values are constant across all values of ω. In Figure S3 of the Supporting Information, we display the probability of early stopping for safety. In scenarios A, B, E, and F, this probability is close 0.25, whereas in scenarios C, D, G, and H is close to 0.05. These values are consistent with the informative prior distributions for the dose-toxicity models and the distance between the dose combination 15/75 mg/m^2, which has a prior

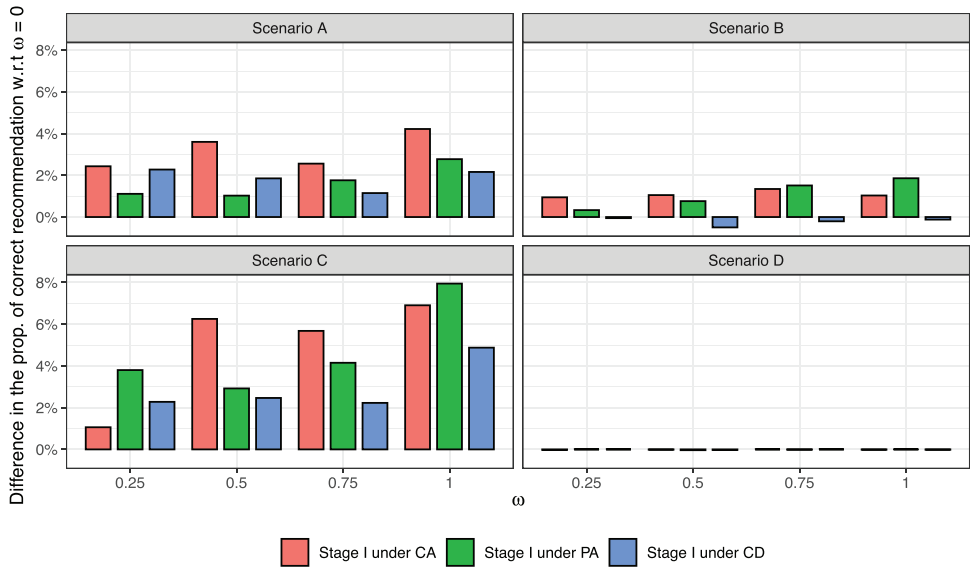

FIGURE 5 Difference in the proportion of correct dose combination recommendation between models with $\omega > 0$ and $\omega = 0$ in settings under H_1.

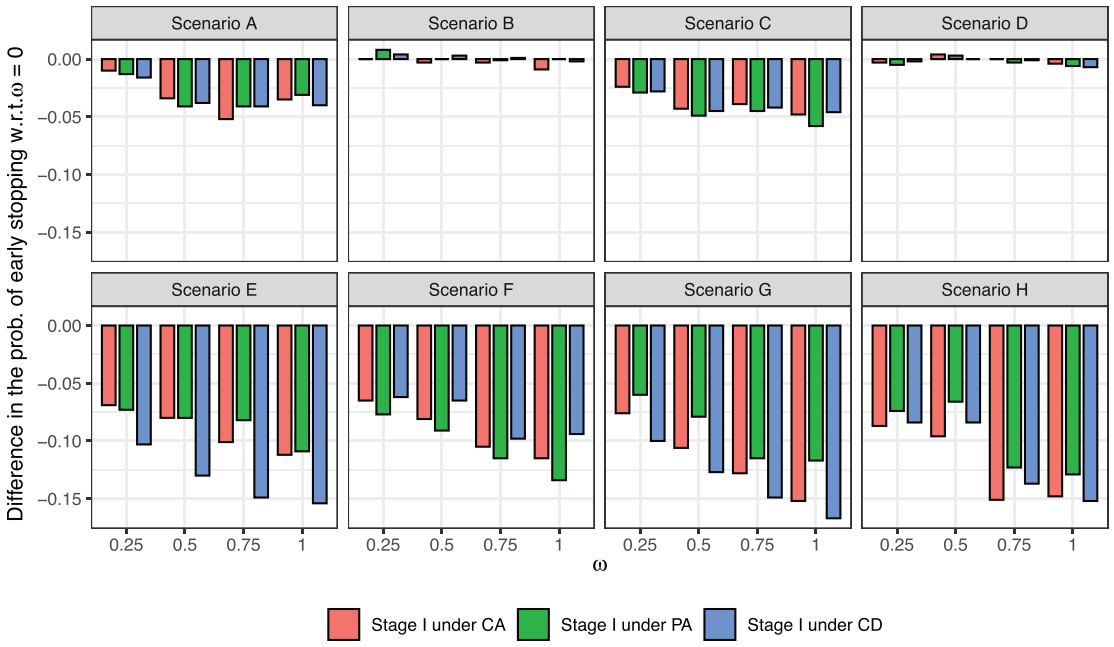

FIGURE 6 Difference in the proportion of trials with early stopping for futility between models with $\omega > 0$ and $\omega = 0$ under H_1 (A–D) and H_0 (E–H).

probability if DLT of approximately 0.33, and the true MTD curves. Also, because toxicity data are not shared across stages, we do not present the average sample sizes as they are in line with those reported by Jiménez et al. (2020) in a similar setting.

5 | DISCUSSIONS

Motivated by a real phase I–II trial that combines continuous dose levels of cisplatin and cabazitaxel involving two different populations of patients with advanced prostate cancer, in this paper, we present a phase I–II design in two stages

that allows robust integration of efficacy data across relevant patient populations. The main contribution of this article lies in the formal consideration about the uncertainty around the potentially different dose-efficacy profiles across stages. We propose to employ a robust Bayesian hierarchical random-effects model to allow sharing of information on the efficacy across stages, assuming that the related parameters are either exchangeable or nonexchangeable. In other words, the key idea is to exploit any potential similarities between the dose-efficacy profiles so as to borrow information, while avoiding too optimistic borrowing under the presence of data inconsistency across stages. This proposal requires specification of the prior probability that the main effects set of parameters is exchangeable across stages. We denote this prior probability by ω, which in practice is selected by subject-matter experts. When $\omega = 0$, the design estimates the stage II dose-efficacy profile independently from stage I (Tighiouart, 2019; Jiménez et al., 2020). In this article, we focus on analyzing the operating characteristics of stage II, and we study how these vary, with respect to $\omega = 0$, as we increase the prior probability of exchangeability ω under different stage I dose-efficacy profiles. For a detailed evaluation of the stage I operating characteristics, we refer the reader to Tighiouart (2019), Jiménez et al. (2020), and Jiménez and Tighiouart (2022).

The selection of the dose-efficacy data model is closely related to the type of compound investigated in a phase I–II clinical trial. With cytotoxic agents, the monotonicity assumption is expected to hold also from an efficacy perspective (i.e., a compound will have greater activity as the dose increases). Thus, a linear model such as the one defined in (3) will be sufficient to capture the dose-efficacy relationship. However, with other types of compounds such as molecularly targeted therapies, more flexible modeling approaches may be needed to capture dose-efficacy relationships where the probability of efficacy may not even increase with the dose.

We have limited the simulations to the two main dose-toxicity scenarios considered by the principal investigator of the motivating example. In each of these dose-toxicity profiles, we have studied two different stage II dose-efficacy profiles, each one accompanied with three stage I dose-efficacy profiles that have different levels of similarity with respect to the stage II dose-efficacy profile. Also, because we allow the main effects set of parameters to be exchangeable across stages, similarity, or agreement across stages is based only on these two parameters. However, depending on the application and the definition of the dose-efficacy profile, this work could be extended by tweaking the `JAGS` code, which we have made publicly available, so it includes other parameters in the set of parameters that could be exchangeable across stages.

The evaluation we present in this article aims to assess whether the overall operating characteristics of design improve by allowing robust integration of efficacy data across stages in scenarios under complete agreement, partial agreement, and complete disagreement between the stage I and stage II dose-efficacy profiles. In other words, we aim to evaluate how much we can benefit from sharing efficacy data across stages when the efficacy data are completely or partially consistent across stages based on the main effects set of parameters, but also to what extend we expect to penalize the design's operating characteristics when the efficacy data are inconsistent across stages. The assessment is done in the original setting with continuous dose combination levels under H_0 and H_1 following the case study described in Section 2.

In scenarios favoring the alternative hypothesis and visualized in Figure 2, we observe a generalized improvement of the operating characteristics by permitting sharing the efficacy data across stages (i.e., $\omega > 0$). The degree of improvement would depend, however, on the genuine extent of consistency between the stage-wise efficacy profiles and, of course, on the value of ω that we select. We note that with $\omega = 0.25$, there is already a considerable improvement of the designs operating characteristics in comparison with higher values of ω.

In scenarios under the null hypothesis, we observed a small inflation in the type-I error and a slight decrease in the probability of early stopping for futility. Under this hypothesis, having a high value of ω in situations of complete disagreement across the stage I and stage II efficacy profiles generally yields the worst performance. However, with $\omega = 0.25$, the differences are much smaller with respect to settings in which there is complete or partial agreement across the stage I and stage II efficacy profiles.

Overall, we believe that allowing for sharing of efficacy data across stages increases the probability of finding an appropriate dose combination for further phase III studies. However, this approach requires preliminary knowledge on the drug combination. We regard this as acceptable, because there is not a unique design configuration that will fit all applications. For example, in our proposal, we allow the main effects set of parameters to be exchangeable or nonexchangeable across stages, and thus, our definition of similarity or agreement across stages is based solely on the main effects parameters. Moreover, we have seen that two dose-efficacy profiles that are similar in terms of the two main effects set of parameters can have completely different intercepts, and can potentially induce either a power loss or a type-I error inflation. Therefore, a clear understanding of what is considered "similar" is key to decide how we want to synthesize the efficacy data across stages. A potential solution to this problem could be to include the intercept in the set of parameters for the assumption of exchangeability or nonexchangeability.

One potential extension of the methodology presented in this manuscript, which we plan to explore in the future, is to robustly combine toxicity data across stages in this particular setting. By doing so, we would eliminate the assumption that the dose-toxicity profiles are equivalent across different patient population and we would account for population-specific characteristics with respect to the MTD.

ACKNOWLEDGMENTS

We thank the associate editor, two referees, and the editor for the constructive comments that led to significant improvements in the article.

Dr. Jiménez is employed by Novartis Pharma AG, who provided support in the form of salary for the author but did not have any additional role in the preparation of the manuscript. Also, the views expressed in this publication are those of the authors and should not be attributed to any of the funding institutions or organizations to which the authors are affiliated.

FUNDING

Dr. Zheng's contribution to this manuscript was supported by Cancer Research UK (RCCPDF/100008).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The `R` and `JAGS` scripts needed to fully reproduce the results presented in this article are available at https://github.com/jjimenezm1989/Bayesian-phase-I-II-design-combining-efficacy-data

OPEN RESEARCH BADGES

This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available in the Supporting Information section.

This article has earned an open data badge "**Reproducible Research**" for making publicly available the code necessary to reproduce the reported results. The results reported in this article could fully be reproduced.

ORCID

José L. Jiménez https://orcid.org/0000-0002-8809-2717
Haiyan Zheng https://orcid.org/0000-0002-3385-2117

REFERENCES

Babb, J., Rogatko, A., & Zacks, S. (1998). Cancer phase I clinical trials: efficient dose escalation with overdose control. *Statistics in Medicine*, *17*(10), 1103–1120.

Cai, C., Yuan, Y., & Ji, Y. (2014). A Bayesian dose-finding design for oncology clinical trials of combinational biological agents. *Journal of the Royal Statistical Society. Series C, Applied Statistics*, *63*(1), 159.

Gasparini, M. (2013). General classes of multiple binary regression models in dose finding problems for combination therapies. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *62*(1), 115–133.

Ivanova, A. (2003). A new dose-finding design for bivariate outcomes. *Biometrics*, *59*(4), 1001–1007.

Ivanova, A., Liu, K., Snyder, E., & Snavely, D. (2009). An adaptive design for identifying the dose with the best efficacy/tolerability profile with application to a crossover dose-finding study. *Statistics in Medicine*, *28*(24), 2941–2951.

Jiang, L., Li, R., Yan, F., Yap, T. A., & Yuan, Y. (2021). Shotgun: A Bayesian seamless phase i-ii design to accelerate the development of targeted therapies and immunotherapy. *Contemporary Clinical Trials*, *104*, 106338.

Jiménez, J. L., Kim, S., & Tighiouart, M. (2020). A Bayesian seamless phase i–ii trial design with two stages for cancer clinical trials with drug combinations. *Biometrical Journal*, *62*(5), 1300–1314.

Jiménez, J. L., & Tighiouart, M. (2022). Combining cytotoxic agents with continuous dose levels in seamless phase i-ii clinical trials. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *71*(5), 1996–2013.

Jimenez, J. L., Tighiouart, M., & Gasparini, M. (2019). Cancer phase i trial design using drug combinations when a fraction of dose limiting toxicities is attributable to one or more agents. *Biometrical Journal*, *61*(2), 319–332.

Le Tourneau, C., Lee, J. J., & Siu, L. L. (2009). Dose escalation methods in phase i cancer clinical trials. *JNCI: Journal of the National Cancer Institute*, *101*(10), 708–720.

Lin, R., Thall, P. F., & Yuan, Y. (2020). An adaptive trial design to optimize dose-schedule regimes with delayed outcomes. *Biometrics*, *76*(1), 304–315.

Liu, S., Guo, B., & Yuan, Y. (2018). A Bayesian phase i/ii trial design for immunotherapy. *Journal of the American Statistical Association*, *113*(523), 1016–1027.

Lockhart, A. C., Sundaram, S., Sarantopoulos, J., Mita, M. M., Wang-Gillam, A., Moseley, J. L., Barber, S. L., Lane, A. R., Wack, C., & Kassalow, L. (2014). Phase i dose-escalation study of cabazitaxel administered in combination with cisplatin in patients with advanced solid tumors. *Investigational New Drugs*, *32*(6), 1236–1245.

Lyu, J., Ji, Y., Zhao, N., & Catenacci, D. V. (2019). Aaa: Triple adaptive Bayesian designs for the identification of optimal dose combinations in dual-agent dose finding trials. *Journal of the Royal Statistical Society. Series C, Applied Statistics*, *68*(2), 385.

Neuenschwander, B., Roychoudhury, S., & Schmidli, H. (2016). On the use of co-data in clinical trials. *Statistics in Biopharmaceutical Research*, *8*(3), 345–354.

Rogatko, A., Ghosh, P., Vidakovic, B., & Tighiouart, M. (2008). Patient-specific dose adjustment in the cancer clinical trial setting. *Pharmaceutical Medicine*, *22*(6), 345–350.

Roychoudhury, S., & Neuenschwander, B. (2020). Bayesian leveraging of historical control data for a clinical trial with time-to-event endpoint. *Statistics in Medicine*, *39*(7), 984–995.

Shi, Y., & Yin, G. (2013). Escalation with overdose control for phase i drug-combination trials. *Statistics in Medicine*, *32*(25), 4400–4412.

Thall, P. F., & Cook, J. D. (2004). Dose-finding based on efficacy-toxicity trade-offs. *Biometrics*, *60*(3), 684–693.

Tighiouart, M. (2019). Two-stage design for phase i–ii cancer clinical trials using continuous dose combinations of cytotoxic agents. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *68*(1), 235–250.

Tighiouart, M., Li, Q., & Rogatko, A. (2017). A Bayesian adaptive design for estimating the maximum tolerated dose curve using drug combinations in cancer phase i clinical trials. *Statistics in Medicine*, *36*(2), 280–290.

Tighiouart, M., & Rogatko, A. (2012). Number of patients per cohort and sample size considerations using dose escalation with overdose control. *Journal of Probability and Statistics*, *2012*.

Tighiouart, M., Rogatko, A., & Babb, J. S. (2005). Flexible Bayesian methods for cancer phase i clinical trials. Dose escalation with overdose control. *Statistics in Medicine*, *24*(14), 2183–2196.

Tighiouart, M., & Rogatko, A. (2010). Dose finding with escalation with overdose control (EWOC) in cancer clinical trials. *Statistical Science*, *25*(2), 217–226.

Wheeler, G. M., Sweeting, M. J., & Mander, A. P. (2017). Toxicity-dependent feasibility bounds for the escalation with overdose control approach in phase i cancer trials. *Statistics in Medicine*, *36*(16), 2499–2513.

Yuan, Y., & Yin, G. (2009). Bayesian dose finding by jointly modelling toxicity and efficacy as time-to-event outcomes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *58*(5), 719–736.

Zheng, H., Hampson, L. V., & Jaki, T. (2021). Bridging across patient subgroups in phase i oncology trials that incorporate animal data. *Statistical Methods in Medical Research*, *30*(4), 1057–1071.

Zheng, H., Hampson, L. V., & Wandel, S. (2020). A robust Bayesian meta-analytic approach to incorporate animal data into phase i oncology trials. *Statistical Methods in Medical Research*, *29*(1), 94–110.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Jiménez, J. L., & Zheng, H. (2023). A Bayesian adaptive design for dual-agent phase I–II oncology trials integrating efficacy data across stages. *Biometrical Journal*, 2200288. https://doi.org/10.1002/bimj.202200288