Computers in Biology and Medicine 188 (2025) 109848

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Optimal designs for efficacy-toxicity response in dose finding studies using the bivariate probit model

Belmiro P.M. Duarte [a,b,c,*], Anthony C. Atkinson [d]

[a] Polytechnic Institute of Coimbra, ISEC, Department of Chemical & Biological Engineering, Rua Pedro Nunes, 3030–199 Coimbra, Portugal
[b] University of Coimbra, INESC-Coimbra, Rua Sílvio Lima — Pólo II, 3030–790 Coimbra, Portugal
[c] University of Coimbra, CERES, Department of Chemical Engineering, Rua Sílvio Lima — Pólo II, 3030–790 Coimbra, Portugal
[d] Department of Statistics, London School of Economics, London WC2A 2AE, United Kingdom

ARTICLE INFO

Keywords:
Optimal design of experiments
Dose–response models
Semidefinite Programming
Efficacy-Toxicity
Probit model
K–optimality

ABSTRACT

Phase I clinical trials are the first-in-human studies that primarily focus on the safety profile of drugs. Traditionally, the primary aim of a phase I clinical trial is to establish the maximum tolerated dose and characterize the toxicity profile of the tested agents. As a secondary aim, some phase I studies also include studies to obtain preliminary efficacy information about the experimental agents. In our research, we consider the optimal design of experiments in extended phase I clinical trials where both efficacy and toxicity are measured and the maximum tolerated dose has been established. We represent the response of both outcomes using a bivariate probit model for correlated responses and propose systematic numerical approaches based on Semidefinite Programming to address the problem. We construct locally optimal experimental designs for the following situations: (i) responses with efficacy and toxicity strongly correlated versus non-correlated, by varying the correlation parameter; (ii) *a priori* known correlation versus unknown correlation; (iii) unconstrained versus constrained designs, where the constraints represent safety limits, budget constraints and probability bounds; (iv) single versus combined drugs. Additionally, we consider four distinct optimality criteria: D–, A–, E–, and K–optimality. Our methodologies are extensively tested, and we demonstrate the optimality of the designs using equivalence theorems. To enrich our analysis, an equivalence theorem for the K–optimality criterion is derived.

1. Motivation

The dose of a drug must be carefully calibrated to achieve efficacy without causing toxicity. Traditionally, safety and efficacy have been evaluated independently across different trial phases. However, efficacy and toxicity are often correlated at the individual level. This paper presents computational tools for designing experiments that simultaneously estimate the relationships between dose, efficacy, and toxicity. Our approach considers the bivariate probit model proposed by Fedorov and Wu [1] and Fedorov et al. [2].

The high attrition rate observed in phase III trials is widely attributed to inadequate dose selection [3]. Typically, phase I trials focus on assessing the safety of the test drug and determining the maximum tolerable dose (MTD) based solely on toxicity. In contrast, phase II trials evaluate the efficacy of the test drug [4]. However, under various circumstances, employing a joint model for toxicity and efficacy to select the optimal dose and analyze their correlation can be advantageous. Several joint modeling strategies have been explored in the literature, see Thall and Cook [5]. In the context of an HIV trial, O'Quigley et al. [6] proposed a model to describe the relationship between dose and success rate. Thall and Russell [7] integrated toxicity and efficacy into a single trinomial variable and employed a proportional odds (PO) regression model. Alternatively, Zhang et al. [8] used a continuation-ratio (CR) model instead of the PO model. Braun [9] introduced a bivariate distribution for correlated binary outcomes, including a parameter to represent the association between the two outcomes. Dragalin and Fedorov [10] modeled the distribution of bivariate binary endpoints using either Gumbel bivariate logistic regression or the Cox bivariate binary model. The bivariate probit model of Fedorov and Wu [1] and Fedorov et al. [2] has the advantage that it can model the correlation between the two outcomes. Dragalin et al. [11] applied this model to combinations of drugs.

These modeling methods are suitable for binary endpoints and effectively model the relationship between dose and the probability of

* Corresponding author at: Polytechnic Institute of Coimbra, ISEC, Department of Chemical & Biological Engineering, Rua Pedro Nunes, 3030–199 Coimbra, Portugal.
E-mail addresses: bduarte@isec.pt (B.P.M. Duarte), A.C.Atkinson@lse.ac.uk (A.C. Atkinson).

https://doi.org/10.1016/j.compbiomed.2025.109848
Received 30 November 2024; Received in revised form 9 February 2025; Accepted 10 February 2025
Available online 22 February 2025

efficacy and toxicity. Continuous outcome models have also been explored [12,13]. Recently, adaptive designs that integrate both efficacy and toxicity data have been investigated [14,15], along with Bayesian designs for mixed types of responses [16].

Notwithstanding the importance of dose finding, challenges persist. Despite advances in optimization algorithms and computational capabilities, systematic tools for designing experimental trials remain lacking. Recent literature provides a broad overview of optimization techniques and methodologies that can be used to construct optimal experimental designs. However, their application in clinical trials to simultaneously capture efficacy and toxicity outcomes has been largely overlooked. Integrating various constraints (such as safety and budgetary) into the design is particularly important when maximizing information within the limitations of the trial scenario. Our paper aims to address this gap. We study optimal experimental designs for clinical trials where both outcomes are measured as binary responses, and the goal is to minimize the confidence regions of the parameter estimates in the models of these outcomes. We assume that the maximum tolerated dose has already been established. We propose systematic tools based on Semidefinite Programming (SDP) to solve the optimization problems. We focus on finding locally optimal designs for normalized dose domains where the bivariate model of outcomes follows the probit class [17].

In our investigation, we focus on the most commonly used optimality criteria: D–, A–, E–, and K–optimality. These criteria are chosen for their simplicity, practical importance and ease of translation into tractable SDP formulations. The first three criteria are primarily used for prediction purposes, while the K–criterion aims to balance prediction with the minimization of ill-conditioning in the generalized least squares problem arising in parameter estimation. We employ equivalence theorems to verify the optimality of the designs found.

1.1. Novelty statement and organization

This paper offers several novel contributions:

(i) an automated numerical method for identifying locally optimal experimental designs for efficacy-toxicity responses, modeled using the bivariate probit framework;
(ii) application of the proposed computational approach to various scenarios, including: a. varying the correlation between the two outcomes; b. employing different optimality criteria; and c. adjusting the number of active agents combined in the trials;
(iii) incorporation of constraints related to safety limitations, such as toxicity thresholds or budgetary restrictions; and
(iv) derivation of an equivalence theorem for the K–optimality criterion, which, together with other equivalence theorems, provides a means of verifying the optimality of the identified designs. Similar to the E–optimality criterion, this equivalence theorem applies to Fisher Information Matrices where the eigenvalues have multiplicity 1 [18]. To the best of our knowledge, this theoretical result is novel.

The paper is organized as follows. Section 2 introduces the background and notation used to formulate the problem of finding optimal designs for bivariate binary outcomes that follow the probit model. Additionally, this Section covers the fundamentals of Semidefinite Programming used to handle these problems numerically. Section 3 presents the formulations used for determining optimal designs of experiments of both unconstrained and constrained classes. The application of these algorithms to a dose–response problem is detailed in Section 4. Section 5 offers a review of the formulations and provides a summary of the results obtained. The Appendix A contains the proof of the equivalence theorem for K–optimality criterion, used to check the optimality of the designs found numerically. The Appendix B includes the SDP formulations for D–, A–, and E–optimality criteria.

2. Notation and background

In our notation bold face lowercase letters represent column vectors, bold face capital letters stand for continuous domains, blackboard bold capital letters are used to denote discrete domains and capital letters are adopted for matrices. Finite sets containing ι elements are compactly represented by $[\![\iota]\!] \equiv \{1, \dots, \iota\}$. The transpose operation of a matrix or vector is represented by "$^\intercal$", and the trace of a matrix by $\mathrm{tr}(\bullet)$.

In Section 2.1, we present the efficacy-toxicity model. In Section 2.2 we introduce the basics of model-based optimal design of experiments. Following that, in Section 2.3, we provide the fundamental aspects of Semidefinite Programming.

2.1. Efficacy-toxicity model

Here, we introduce the probit model used to represent the bivariate efficacy-toxicity response, see Fedorov and Leonov [19, Chap. 6] and Fedorov et al. [2]

Let $\mathbf{x} \in \mathbf{X}$ represent the vector of doses administered in a drug response study involving a set of individuals. To simplify the notation, we consider the administration of a single drug. The extension to a drug combination is straightforward. Each element of $\mathbf{x}$, denoted by x_i, represents a single scaled dose, while the compact domain of possible doses is given by $\mathbf{X} \equiv [0, 1]$. Each patient receives a dose $x_i \in \mathbf{X}$, where $i \in [\![n_d]\!]$, and n_d represents the number of distinct doses used in the study. The same dose may be administered to multiple individuals; that is, the response to x_i is replicated n_i times, where n_i can vary across dose levels. If we extend the study to the administration of two combined drugs, the dose at the ith level is represented as a two-row vector $\mathbf{x}_i \in [0,\ 1]^2$ containing the doses of the two combined drugs. In this case, the full set of doses is given by $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_i, \dots, \mathbf{x}_{n_d})$.

Now, let us consider the binary variable Y (with possible values 0/1) indicating the response with respect to the efficacy of a given dose $x_i \in \mathbf{X}$. Here $Y = 0$ indicates that the dose has no efficacy on the individual and $Y = 1$ the opposite. Similarly, Z is a binary variable indicating the response to the toxicity of the same dose where $Z = 0$ indicates that the dose has no toxicity on the individual and $Z = 1$ the opposite. Thus, the combined response probability of having $Y = y$ and $Z = z$ with $y,\ z \in \{0,\ 1\}$ is

$$p_{y,z}(x_i, \boldsymbol{\theta}) = \mathbb{P}(Y = y, Z = z | x_i, \boldsymbol{\theta}), \tag{1}$$

where $\boldsymbol{\theta} \in \Theta$ is the vector of parameters included in the model representing both the efficacy and toxicity outcomes, $\Theta \in \mathbb{R}^{n_\theta}$ is the compact domain including all the possible values of parameters, n_θ stands for the number of parameters in both models and may include the correlation parameter relating them in case it is unknown *a priori*. $\mathbb{P}(\bullet)$ represents the probability of the combined response in terms of efficacy and toxicity of administering the dose x_i. In case we consider the case of two combined drugs x_i is replaced by $\mathbf{x}_i$

The parameter sets for the efficacy and toxicity models are denoted as $\boldsymbol{\theta}_1 \in \Theta_1$ and $\boldsymbol{\theta}_2 \in \Theta_2$, respectively. The full parameter set is given by $\boldsymbol{\theta} = \boldsymbol{\theta}_1 \oplus \boldsymbol{\theta}_2$, where $\oplus$ represents concatenation. The parameter space is defined as $\Theta = \Theta_1 \times \Theta_2 \subseteq \mathbb{R}^{n_{\theta_1}} \times \mathbb{R}^{n_{\theta_2}}$, where n_{θ_1} and n_{θ_2} are the dimensions of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, respectively.

Efficacy outcomes are represented by $y = 1$ (efficacious) or $y = 0$ (non-efficacious), and toxicity outcomes are $z = 1$ (non-toxic) or $z = 0$ (toxic). The probability of observing efficacy, toxicity, efficacy without toxicity, efficacy with toxicity, no efficacy without toxicity, and no efficacy with toxicity are denoted by $p_{y=1,z\in\{0,1\}}$, $p_{y\in\{0,1\},z=1}$, $p_{y=1,z=0}$, $p_{y=1,z=1}$, $p_{y=0,z=0}$, and $p_{y=0,z=1}$, respectively. The efficacy and toxicity response models for a dose x_i are given by

$$\eta_1(x_i, \boldsymbol{\theta}_1) = \boldsymbol{\theta}_1^\intercal \cdot \mathbf{f}_1(x_i) \tag{2a}$$

$$\eta_2(x_i, \boldsymbol{\theta}_2) = \boldsymbol{\theta}_2^\intercal \cdot \mathbf{f}_2(x_i). \tag{2b}$$

where $\mathbf{f}_1(\bullet)$ and $\mathbf{f}_2(\bullet)$ are vectors of polynomial functions of the covariates of interest, included in efficacy and toxicity response models, respectively.

These probabilities are modeled using the probit family [20,21] as:

$$p_{y=1,z\in\{0,1\}}(x_i,\boldsymbol{\theta}) = \Phi_1[\eta_1(x_i,\theta_1)], \tag{3a}$$

$$p_{y\in\{0,1\},z=1}(x_i,\boldsymbol{\theta}) = \Phi_1[\eta_2(x_i,\theta_2)], \tag{3b}$$

$$p_{y=1,z=1}(x_i,\boldsymbol{\theta}) = \Phi_2[\eta_1(x_i,\theta_1),\eta_2(x_i,\theta_2),\Sigma], \tag{3c}$$

$$p_{y=1,z=0}(x_i,\boldsymbol{\theta}) = \Phi_1[\eta_1(x_i,\theta_1)] - p_{y=1,z=1}(x_i,\boldsymbol{\theta}), \tag{3d}$$

$$p_{y=0,z=1}(x_i,\boldsymbol{\theta}) = \Phi_1[\eta_2(x_i,\theta_2)] - p_{y=1,z=1}(x_i,\boldsymbol{\theta}), \tag{3e}$$

$$p_{y=0,z=0}(x_i,\boldsymbol{\theta}) = 1 - \Phi_1[\eta_1(x_i,\theta_1)] - \Phi_1[\eta_2(x_i,\theta_2)] + p_{y=1,z=1}(x_i,\boldsymbol{\theta}). \tag{3f}$$

Here, $\Phi_1[a]$ is the standard univariate normal CDF evaluated at a, and $\Phi_2[b,c,\Sigma]$ is the bivariate normal CDF at (b,c) when both response variables are correlated by matrix Σ where

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

ρ being the correlation between the response variables.

The number of parameters of the two-outcome model is n_θ, each of the parameters being constrained to a compact domain, i.e. $\theta_i \in [\theta_i^L, \theta_i^U]$, where θ_i^L is the lower bound for parameter i, and θ_i^U the upper bound. Consequently, the Cartesian domain containing all the combinations of parameters is $\boldsymbol{\Theta} \equiv \otimes_{i=1}^{n_\theta}[\theta_i^L, \theta_i^U] \subset \mathbb{R}^{n_\theta}$. To distinguish between the generic vector $\boldsymbol{\theta}$ and a singleton realization vector of $\boldsymbol{\Theta}$, the latter is designated by $\mathbf{q}_0$.

The model, defined by Eqs. (2)–(3), can be compactly represented as

$$\mathbb{E}(\mathbf{y}) = \mathcal{G}(\mathbf{x},\boldsymbol{\theta}), \tag{4}$$

where $\mathbf{y}$ denotes the observed binary outcomes (0 or 1) for efficacy and toxicity, $\mathbf{x}$ represents the set of dose level(s) of each drug, and $\boldsymbol{\theta}$ is the complete parameter set governing the models for both outcomes.

To summarize, the primary goal of optimal designs is to estimate the parameters of the models for both outcomes by measuring the probability of toxicity and efficacy in a group of individuals tested with a set of doses. These doses are optimally chosen and may consist of a single active principle or a combination of active principles.

2.2. Optimal design of experiments

We revisit model (2)–(3) and consider a continuous design with ℓ support points located at $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_\ell$. In practice, ℓ corresponds to the number of distinct doses in the trial, denoted n_d. However, we use a different notation here because ℓ is treated as a variable to be determined in the optimal design of experiments, whereas n_d is assumed to be known for planning purposes in trials. Furthermore, we represent the support points as vectors, accommodating both scenarios: (i) single-drug trials, where $\mathbf{x}$ consists of a single element, and (ii) two-drug trials, where $\mathbf{x}$ includes two elements.

Continuous designs, also known as approximate designs, are employed to represent experimental setups as the number of observations, N, approaches infinity ($N \to +\infty$). In this setting, the weights are continuously distributed over the interval $[0,1]$, representing the proportion of total observations allocated to each support point. Continuous designs offer several advantages. They provide a unified framework for identifying optimal designs, especially in model-based optimal design of experiments, where the design criterion is often a concave or convex function over the set of approximate designs [22]. On the other hand, *exact designs* are used when N is finite, focusing on determining the exact number of experiments allocated to each support point. In this paper, we focus on continuous designs.

In the context of continuous designs, the weights at the support points are denoted as $w_1, w_2, \ldots, w_\ell$, where $\ell \geqslant n_\theta$. To implement the design for a total of N individuals, approximately $N \times w_k$ individuals are allocated to each support point $\mathbf{x}_k$ for every $k \in [\![\ell]\!]$. This allocation ensures that the total number of individuals is maintained, such that $N \times w_1 + \cdots + N \times w_\ell = N$, with each term being an integer.

For models with n_c control factors, where each factor corresponds to the dose of a different drug, the kth support point is $\mathbf{x}_k^\intercal = (x_{k,1}, \ldots, x_{k,n_c})$. In the single-drug case, this simplifies to x_k. For clarity, we present the single-drug case, noting that replacing x_k with $\mathbf{x}_k^\intercal$ extends the results to multiple drugs.

The continuous design space $\mathbf{X}$ is replaced by a discrete set, denoted as $\mathbb{X}^{[\![n_x]\!]}$, consisting of n_x candidate design points sampled from $\mathbf{X}$, often referred to as doses in this context. The continuous design ξ comprises ℓ columns, each represented as $(x_k^\intercal, w_k)^\intercal$ for $k \in [\![\ell]\!]$, subject to the constraint $\sum_{k=1}^{\ell} w_k = 1$. It is important to note that the number of support points, ℓ, is determined by the design algorithm. In practice, ℓ corresponds to the number of support points among the n_x candidate points (or doses) for which $w_k > 0$.

The theoretical foundations of the optimal design of experiments were established by Kiefer [23] and Kiefer and Wolfowitz [24], and we build on these principles to present the basic concepts. In the following sections, we define $\Xi \equiv \mathbb{X}^{[\![n_x]\!]} \times Y$ as the space of feasible ℓ-point designs over $\mathbf{X}$, where Y is the $n_x - 1$ simplex in the domain of weights, given by

$$Y = \left\{ w_k \mid w_k \geqslant 0, \forall k \in [\![n_x]\!], \sum_{k=1}^{n_x} w_k = 1 \right\}.$$

The information resulting from an experimental design is measured by its Fisher Information Matrix (FIM). For single drug case the elements of the normalized FIM are given by

$$\mathcal{M}(\xi,\boldsymbol{\theta}) = \int_{\xi\in\Xi} M(\mathbf{x},\boldsymbol{\theta})\,\mathrm{d}(\xi) = \sum_{k=1}^{\ell} w_k \; M(x_k,\boldsymbol{\theta}), \tag{5}$$

where $\mathcal{M}(\xi,\boldsymbol{\theta})$ is the *global* FIM from the design ξ and $M(x_k,\boldsymbol{\theta})$ is the *local* FIM from point x_k. For the model (4) let

$$M(x_k,\boldsymbol{\theta}) = \left.\frac{\partial \mathcal{G}(\mathbf{x},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\intercal}\right|_{x_k} \cdot \left.\frac{\partial \mathcal{G}(\mathbf{x},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{x_k} \tag{6}$$

where $\partial \mathcal{G}(\mathbf{x},\boldsymbol{\theta})/\partial \boldsymbol{\theta}^\intercal|_{x_k}$ is the vector of derivatives of $\mathcal{G}(\mathbf{x},\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ at x_k.

Consequently, the FIM for a single observation when ρ is known and $\boldsymbol{\theta} \in \mathbb{R}^{n_\theta}$ is

$$\begin{aligned} M(x_i,\boldsymbol{\theta}) =& C_1[\eta_1(x_i,\theta_1),\eta_2(x_i,\theta_2),x_i]\cdot \\ & \cdot C_2[\eta_1(x_i,\theta_1),\eta_2(x_i,\theta_2),x_i]\cdot \\ & \cdot (P - \mathbf{p}\cdot\mathbf{p}^\intercal)^{-1}\cdot C_2[\eta_1(x_i,\theta_1),\eta_2(x_i,\theta_2),x_i]^\intercal\cdot \\ & \cdot C_1[\eta_1(x_i,\theta_1),\eta_2(x_i,\theta_2),x_i]^\intercal \end{aligned} \tag{7a}$$

$$C_1[\eta_1(x_i,\theta_1),\eta_2(x_i,\theta_2),x_i] = \begin{pmatrix} \phi_1[\eta_1(x_i,\theta_1)]\cdot\mathbf{f}_1(x_i) & \mathbf{0}_{n_{\theta_1}} \\ \mathbf{0}_{n_{\theta_2}} & \phi_1[\eta_2(x_i,\theta_2)]\cdot\mathbf{f}_2(x_i) \end{pmatrix} \tag{7b}$$

$$C_2[\eta_1(x_i,\theta_1),\eta_2(x_i,\theta_2),x_i] = \begin{pmatrix} \Phi_1(u_1) & 1-\Phi_1(u_1) & -\Phi_1(u_1) \\ \Phi_1(u_2) & -\Phi_1(u_2) & 1-\Phi_1(u_2) \end{pmatrix} \tag{7c}$$

$$u_1 = \frac{\eta_2(x_i,\theta_2) - \rho\cdot\eta_1(x_i,\theta_1)}{\sqrt{1-\rho^2}} \tag{7d}$$

$$u_2 = \frac{\eta_1(x_i,\theta_1) - \rho\cdot\eta_2(x_i,\theta_2)}{\sqrt{1-\rho^2}} \tag{7e}$$

$$P = \begin{pmatrix} p_{y=1,z=1} & 0 & 0 \\ 0 & p_{y=1,z=0} & 0 \\ 0 & 0 & p_{y=0,z=1} \end{pmatrix} \tag{7f}$$

$$\mathbf{p} = (p_{y=1,z=1},\ p_{y=1,z=0},\ p_{y=0,z=1})^\intercal \tag{7g}$$

Here, $\phi_1[a]$ denotes the value of the univariate standard normal probability density function at a. P is a diagonal matrix containing the

probabilities defined in Eqs. (3c)–(3e), and $\mathbf{p}$ is the corresponding column vector. $C_1[\bullet]$ and $C_2[\bullet]$ are matrices of dimensions $n_\theta \times 2$ and 2×3, respectively.

For unknown values of ρ the FIM has an additional column and row, and is:

$$M(x_i,\boldsymbol{\theta}) = \begin{pmatrix} C_1[\eta_1(x_i,\theta_1),\eta_2(x_i,\theta_2),x] \cdot C_2[\eta_1(x_i,\theta_1),\eta_2(x_i,\theta_2),x_i] \\ \varphi(x_i,\boldsymbol{\theta},\Sigma) \qquad -\varphi(x_i,\boldsymbol{\theta},\Sigma) \qquad -\varphi(x_i,\boldsymbol{\theta},\Sigma) \end{pmatrix} \cdot (P - \mathbf{p}\cdot\mathbf{p}^\intercal)^{-1} \cdot \begin{pmatrix} C_1[\eta_1(x_i,\theta_1),\eta_2(x_i,\theta_2),x_i] \cdot C_2[\eta_1(x_i,\theta_1),\eta_2(x_i,\theta_2),x_i] \\ \varphi(x_i,\boldsymbol{\theta},\Sigma) \qquad -\varphi(x_i,\boldsymbol{\theta},\Sigma) \qquad -\varphi(x_i,\boldsymbol{\theta},\Sigma) \end{pmatrix}^\intercal, \tag{8a}$$

$$\varphi(x_i,\boldsymbol{\theta},\Sigma) = \phi_2[\eta_1(x_i,\theta_1),\eta_2(x_i,\theta_2),\Sigma]. \tag{8b}$$

$\phi_2(\eta_1(x_i,\theta_1),\eta_2(x_i,\theta_2),\Sigma)$ represents the bivariate normal probability distribution at $[(\eta_1(x_i,\theta_1),\ \eta_2(x_i,\theta_2))^\intercal]$ with correlation matrix Σ and $\varphi(x_i,\boldsymbol{\theta},\Sigma)$ an auxiliary variable used to represent it compactly. The remaining variables included have the same meaning as that in (7).

Since $\hat{\boldsymbol{\theta}}$ is asymptotically normally distributed, the volume of the asymptotic confidence region for $\boldsymbol{\theta}$ is inversely proportional to the square root of $\det[\mathcal{M}(\xi,\boldsymbol{\theta})]$. Therefore, maximizing the determinant of the FIM minimizes the volume of this confidence region. Various design criteria aim to optimize the FIM in different ways, typically formulated as convex functions of the FIM. When $\boldsymbol{\theta}$ is fixed, the locally D–, A–, E–, and K–optimal designs are defined as follows:

$$\xi_{\mathrm{D}} = \arg\max_{\xi\in\Xi} \{\det[\mathcal{M}(\xi,\boldsymbol{\theta})]\}^{1/n_\theta}, \tag{9a}$$

$$\xi_{\mathrm{A}} = \arg\min_{\xi\in\Xi} \left\{\mathrm{tr}[\mathcal{M}(\xi,\boldsymbol{\theta})^{-1}]\right\}, \tag{9b}$$

$$\xi_{\mathrm{E}} = \arg\max_{\xi\in\Xi} \left\{\lambda_{\min}[\mathcal{M}(\xi,\boldsymbol{\theta})]\right\}, \tag{9c}$$

$$\xi_{\mathrm{K}} = \arg\min_{\xi\in\Xi} \{\kappa[\mathcal{M}(\xi,\boldsymbol{\theta})]\}, \tag{9d}$$

where $\lambda_{\min}(\bullet)$ denotes the smallest eigenvalue of the FIM, and $\kappa(\bullet)$ is the condition number, defined as the ratio of the largest to the smallest eigenvalue. The optimal designs, D–, A–, E–, and K–optimal, are obtained by solving the respective optimization problems in (9).

To compare the Ψ-optimal efficiency, where $\Psi \in \{\mathrm{D},\mathrm{A},\mathrm{E},\mathrm{K}\}$, as an indicator of the information content extracted from two different designs, denoted ξ_Ψ and ξ_Ψ^{ref} (with the latter being the reference design), we use the following formulas:

$$\mathrm{Eff}_{\mathrm{D}} = \left\{\frac{\det[\mathcal{M}(\xi_D,\boldsymbol{\theta})]}{\det[\mathcal{M}(\xi_D^{\mathrm{ref}},\boldsymbol{\theta})]}\right\}^{1/n_\theta}, \tag{10a}$$

$$\mathrm{Eff}_{\mathrm{A}} = \frac{\mathrm{tr}[\mathcal{M}^{-1}(\xi_A^{\mathrm{ref}},\boldsymbol{\theta})]}{\mathrm{tr}[\mathcal{M}^{-1}(\xi_A,\boldsymbol{\theta})]}, \tag{10b}$$

$$\mathrm{Eff}_{\mathrm{E}} = \frac{\lambda_{\min}[\mathcal{M}(\xi_E,\boldsymbol{\theta})]}{\lambda_{\min}[\mathcal{M}(\xi_E^{\mathrm{ref}},\boldsymbol{\theta})]}, \tag{10c}$$

$$\mathrm{Eff}_{\mathrm{K}} = \frac{\kappa[\mathcal{M}(\xi_K^{\mathrm{ref}},\boldsymbol{\theta})]}{\kappa[\mathcal{M}(\xi_K,\boldsymbol{\theta})]}. \tag{10d}$$

Since these criteria Ψ are convex or concave over the space of information matrices, the global optimality of any design $\xi \in \Xi$ can be verified using equivalence theorems (see, for instance, Whittle [25], Kiefer [26] and Fedorov [22]). The convexity of the K–optimality criterion is discussed in Yue et al. [27]. These theorems are derived from considerations of directional derivatives and share a general structure, with each convex criterion exhibiting its specific form. A design $\xi \in \Xi$ is considered Ψ-optimal if the corresponding function $\upsilon_\Psi(\mathbf{x},\boldsymbol{\theta})$, for $\Psi \in \{\mathrm{D},\mathrm{A},\mathrm{E},\mathrm{K}\}$—commonly referred to as the dispersion function—is bounded above by zero and achieves this bound at the support points of ξ. The dispersion functions are as follows:

$$\upsilon_{\mathrm{D}}(\mathbf{x},\boldsymbol{\theta}) = \mathrm{tr}\left[M(\mathbf{x},\boldsymbol{\theta})\cdot\mathcal{M}(\xi_{\mathrm{D}},\boldsymbol{\theta})^{-1}\right] - n_\theta \leqslant 0, \quad \forall \mathbf{x}\in\mathbf{X}, \tag{11a}$$

$$\upsilon_{\mathrm{A}}(\mathbf{x},\boldsymbol{\theta}) = \mathrm{tr}\left[M(\mathbf{x},\boldsymbol{\theta})\cdot\mathcal{M}(\xi_{\mathrm{A}},\boldsymbol{\theta})^{-2}\right] - \mathrm{tr}[\mathcal{M}(\xi_{\mathrm{A}},\boldsymbol{\theta})^{-1}] \leqslant 0, \quad \forall \mathbf{x}\in\mathbf{X}, \tag{11b}$$

$$\upsilon_{\mathrm{E}}(\mathbf{x},\boldsymbol{\theta}) = \mathrm{tr}\left[M(\mathbf{x},\boldsymbol{\theta})\cdot Q^{\min}(\boldsymbol{\delta}_x,\boldsymbol{\theta})\right] - \lambda_{\min}[\mathcal{M}(\xi_{\mathrm{E}},\boldsymbol{\theta})] \leqslant 0, \quad \forall \mathbf{x}\in\mathbf{X}, \tag{11c}$$

$$\upsilon_{\mathrm{K}}(\mathbf{x},\boldsymbol{\theta}) = \kappa[\mathcal{M}(\xi_{\mathrm{K}},\boldsymbol{\theta})] - \frac{\mathrm{tr}\left[M(\mathbf{x},\boldsymbol{\theta})\cdot Q^{\max}(\boldsymbol{\delta}_x,\boldsymbol{\theta})\right]}{\mathrm{tr}\left[M(\mathbf{x},\boldsymbol{\theta})\cdot Q^{\min}(\boldsymbol{\delta}_x,\boldsymbol{\theta})\right]} \leqslant 0, \quad \forall \mathbf{x}\in\mathbf{X}, \tag{11d}$$

where $\mathbf{x}$ is a generic point in $\mathbb{R}^{n_c}$, and $Q^{\min}(\mathbf{x},\boldsymbol{\theta}) = \mathbf{v}_{\min}(\mathbf{x},\boldsymbol{\theta})\cdot\mathbf{v}_{\min}^\intercal(\mathbf{x},\boldsymbol{\theta})$ and $Q^{\max}(\mathbf{x},\boldsymbol{\theta}) = \mathbf{v}_{\max}(\mathbf{x},\boldsymbol{\theta})\cdot\mathbf{v}_{\max}^\intercal(\mathbf{x},\boldsymbol{\theta})$ are $n_\theta\times n_\theta$ matrices. Here, $\mathbf{v}_{\min}(\mathbf{x},\boldsymbol{\theta})$ and $\mathbf{v}_{\max}(\mathbf{x},\boldsymbol{\theta})$ are the eigenvectors of the FIM associated with the minimum and maximum eigenvalues, $\lambda_{\min}$ and $\lambda_{\max}$, respectively, ξ_{D}, ξ_{A}, ξ_{E} and ξ_{K} are the optimal designs obtained using D–, A–, E– and K–optimality criteria. Since the result for the K–optimality criterion is novel, it is demonstrated in Appendix A.

2.3. Semidefinite Programming

In this Section, we introduce the fundamentals of this class of convex optimization methods, which are used to solve optimal design of experiments problems. Specifically, we focus on the case where the discretized design domain $\mathbb{X}^{[\![n_x]\!]}$ consists of n_x candidate experimental points.

Let $\mathbb{S}_+^{n_\theta}$ be the space of $n_\theta \times n_\theta$ symmetric positive semidefinite matrices, and $\mathbb{S}^{n_\theta}$ the space of $n_\theta \times n_\theta$ symmetric matrices. A convex set $\mathbf{S} \in \mathbb{R}^{n_\theta}$ is semidefinite representable (SDr) if $\mathrm{proj}_{\mathbf{S}^{\mathrm{exp}}}(\boldsymbol{\zeta})$, $\forall\boldsymbol{\zeta}\in\mathbf{S}$, interpreted as the projection $\boldsymbol{\zeta}$ on to a higher dimensional set $\mathbf{S}^{\mathrm{exp}}$, can be described by Linear Matrix Inequalities (LMIs).

In turn, a convex (or concave) function $\varphi : \mathbb{R}^{m_1} \mapsto \mathbb{R}$ is SDr if and only if the epigraph of φ, $\{(t,\boldsymbol{\zeta}) : \varphi(\boldsymbol{\zeta}) \leqslant t\}$, or the hypograph, $\{(t,\boldsymbol{\zeta}) : \varphi(\boldsymbol{\zeta}) \geqslant t\}$, respectively, are SDr and can be represented by LMIs [28,29]. The optimal values, $\boldsymbol{\zeta}$, of SDr functions are then formulated as *semidefinite programs* of the form

$$\max_{\boldsymbol{\zeta}} \left\{ \mathbf{d}^\intercal\,\boldsymbol{\zeta}, \sum_{i=1}^{m_1} \zeta_i\, M_i - M_0 \succeq 0 \right\}. \tag{12}$$

In our design context, $\mathbf{d}$ is a vector of known constants that depends on the specific design problem. The matrices M_i, for $i \in \{0,\ldots,m_1\}$, represent local Fisher Information Matrices and other matrices derived from reformulating the functions $\varphi(\boldsymbol{\zeta})$ into LMIs. The notation $M_0 \succeq 0$ indicates that the matrix M_0 must be semidefinite positive for the solution to be feasible.

The decision variables, contained in the vector $\boldsymbol{\zeta}$, include the weights w_i for $i \in [\![n_x]\!]$ that define the optimal design, as well as other auxiliary variables needed for the formulation. The design problem is to calculate the optimal design for a pre-specified set of candidate experiments $\mathbb{X}^{[\![n_x]\!]}$, consisting of points $\mathbf{x}_i$ for $i \in [\![n_x]\!]$. This is solved using the formulation in (12), subject to the linear constraints on $\mathbf{w}$: (i) $\mathbf{w} \geqslant 0$, and (ii) $\mathbf{1}_{n_x}^\intercal\,\mathbf{w} = 1$, where $\mathbf{1}_{n_x}$ is the unit column vector of length n_x. The problem in (12) is a classic SDP problem, incorporating LMIs that represent conic constraints.

Ben-Tal and Nemirovski [28] provide a list of SDr functions useful for solving continuous optimal design problems with SDP formulations, see Boyd and Vandenberghe [29, §7.3]. Recently, Sagnol [30] showed that each criterion in the Kiefer class of optimality criteria is SDr for all rational values of $\omega \in (-\infty, 0]$ and general Semidefinite Programming formulations exist for them. Here, ω is the coefficient in the Kiefer general class of criteria Ψ_ω [26]. Notice that A–optimality corresponds to $\omega = -1$, E–optimality to $\omega \to -\infty$ and D–optimality to $\omega \to 0$. Practically, the problem of finding optimal approximate experimental designs for the most common convex (or concave) criteria can be formulated as a Semidefinite Programming problem falling into the general representation (12) complemented with the constraints on $\mathbf{w}$, see Vandenberghe and Boyd [31] and Duarte and Wong [32] among others.

3. Formulations for finding optimal designs of experiments

In this Section, we introduce the proposed formulations for finding the alphabetic optimal experimental designs used to allocate doses to individuals. First, in Section 3.1, we discuss unconstrained problems. Then, in Section 3.2, we generalize the formulation to constrained scenarios which are important in the application to the design of clinical trials.

We consider the regression model (4). A uniformly spaced grid is used for discretization purposes, with step size Δx. Consequently, the continuous design space $\mathbf{X}$ of the regressors is approximated by a finite discrete set of candidate points, $\mathbb{X}^{[\![n_x]\!]}$. Here, $n_x = 1+(x^U-x^L)/\Delta x$, where x^U and x^L are the upper and lower bounds of $\mathbf{X}$, respectively. After discretizing the design space, the *local* Fisher Information Matrices at each candidate point are constructed. For nonlinear models, the singleton vector of parameters used to construct the FIMs is $\mathbf{q}_0$.

The same strategy is applied to two-dimensional covariates in drug combination setups. In this case, the candidate points result from the discretization of the dimensions representing the doses of both drugs, corresponding to the nodes of the grid.

3.1. Finding alphabetic unconstrained locally optimal experimental designs

This Section presents the formulations for finding locally optimal alphabetic designs via Semidefinite Programming. While SDP ensures the global optimum for a grid of discrete candidate points, it can be computationally challenging if the number of candidate experiments is large. The SDP formulations for all criteria follow the general form (12). The formulations for D–, A–, and E–optimality criteria are state–of-the-art and can be found in Vandenberghe and Boyd [31] and Boyd and Vandenberghe [29]. Detailed formulations for the D–, A– and E–optimality criteria, which are currently considered state–of-the-art, are provided in Appendix B. The formulation for K–optimality is less familiar and is therefore analyzed here in more detail. The K–optimality criterion minimizes the condition number of the Fisher Information Matrix by choosing the optimal experimental design points, see the representation in (9d).

The concept of K-optimal designs was introduced by Ye and Zhou [33], who demonstrated that the condition number of the FIM—a naturally positive semidefinite matrix—is a smooth function for polynomial regression models with design intervals constrained to $[-1,+1]$. This insight enables the formulation of K–optimal designs as a semidefinite programming (SDP) problem within the space of regressors.

The K-optimality criterion is defined as the ratio $\lambda_{\max}[\mathcal{M}(\xi,\theta)]/\lambda_{\min}[\mathcal{M}(\xi,\theta)]$, where $\lambda_{\max}$ and $\lambda_{\min}$ are the largest and smallest eigenvalues of the FIM, respectively. While $\lambda_{\max}$ is non-convex and $\lambda_{\min}$ is concave, the criterion is a locally Lipschitz function that is Clarke regular (exhibiting predictable behavior even at nonsmooth points) and strongly pseudoconvex, ensuring convergence to a unique global minimizer [34]. Consequently, the global solution can be approximated by solving a nonsmooth convex program.

Lu and Pong [35] reformulated the condition number optimization in a convex framework, solvable via SDP, while Boyd et al. [36, Chap. 3] showed that such problems can be represented as Linear Matrix Inequalities (LMIs), further facilitating their solution through SDP.

Although the condition number is quasi-convex – characterized by having convex sublevel sets – it can still be efficiently addressed using convex optimization techniques such as semidefinite programming (SDP), for the following reasons:

1. The condition number, defined via eigenvalues, can be reformulated as Linear Matrix Inequalities (LMIs), making it naturally compatible with SDP frameworks.
2. The feasible set, consisting of positive semidefinite matrices, is inherently convex. This convexity enables robust convex optimization methods to effectively manage quasi-convex objectives, ensuring computational tractability.

Our representation of the dispersion function for the K–optimality criterion (see Fig. 1(d)) further strengthens the quasi-convex nature of the problem.

We now adopt a reformulation similar to that of Ye and Zhou [33] to establish the K–optimality criterion. The SDP formulation for K–optimal designs is as follows:

$$\text{Opt} \equiv \min_{s,t,\mathbf{z}} \quad s \tag{13a}$$

$$\text{s.t.} \quad s \cdot I_{n_\theta} - \mathcal{M}(\xi,\mathbf{q}_0) \succeq 0 \tag{13b}$$

$$\mathcal{M}(\xi,\mathbf{q}_0) - I_{n_\theta} \succeq 0 \tag{13c}$$

$$\mathcal{M}(\xi,\mathbf{q}_0) = \sum_{i=1}^{n_x} z_i M(x_i,\mathbf{q}_0) \tag{13d}$$

$$\sum_{i=1}^{n_x} z_i = t \tag{13e}$$

$$z_i \geqslant 0, \quad i \in [\![n_x]\!], \quad t > 0 \tag{13f}$$

Here, Eq. (13b) establishes the upper bound for the eigenvalues of the Fisher Information Matrix (FIM), while Eq. (13c) ensures the positive definiteness of the FIM. Eq. (13d) constructs the global FIM by aggregating the local FIMs at the candidate points x_i, denoted as $M(x_i,\mathbf{q}_0)$, where $\mathbf{q}_0$ captures the dependence on parameters in the nonlinear models. Eq. (13e) ensures that the sum of the values of $\mathbf{z}$ equals t. The variable s quantifies the condition number, and t represents the sum of the weights of the candidate points in a non-unitary domain.

After solving this optimization problem, the optimal design weights are computed by normalizing the vector $\mathbf{z}$ as follows:

$$w_i = \frac{z_i}{t}, \quad i \in [\![n_x]\!]. \tag{14}$$

3.2. Finding alphabetic constrained locally optimal experimental designs

Constrained designs arise when the decision variables $\mathbf{w}$ in ξ are limited due to external factors such as a cost function. Systematically handling these constraints involves incorporating them, represented as inequalities or equalities, into optimal design problems, such as that resulting from K–optimality (model (13)).

We consider three types of constraints. The first type involves restricting the administered dose in tests, referred to as *dose-constrained designs*. These designs penalize certain configurations by limiting doses to a therapeutic range [19], expressed as:

$$w_i = 0, \quad i \notin \mathbb{I} \equiv \{i \mid i \in [\![n_x]\!], x_{\text{lim}}^{LO} \leqslant x_i \leqslant x_{\text{lim}}^{UP}\}, \tag{15}$$

where x_{lim}^{LO} and x_{lim}^{UP} are the lower and upper dose limits, imposed to mitigate toxicity concerns, and are set to 0 and 1, respectively, and $\mathbb{I}$ the set of infeasible candidate doses. These designs are comparable to those proposed by Haines et al. [37], which assign a probability of 1 to scenarios where $p_{y=1,z\in\{0,1\}} \geqslant q_E$ and $p_{y\in\{0,1\},z=1} \leqslant q_T$, while assigning a value of $+\infty$ to all other cases.

We now consider *budget-constrained designs*, where the allocation vector $\mathbf{w}$ is restricted by a specified budget, B. The costs associated with administering each dose comprise two components: (i) a fixed cost, denoted by α; and (ii) a variable cost that increases linearly with dose size, represented by a per-unit increment β. The cost of administering dose x_i is thus expressed as $c_i = \alpha + \beta \cdot x_i$, and the budget constraint is formulated as:

$$\sum_{i=1}^{n_x} w_i c_i \leqslant B. \tag{16}$$

This formulation aligns with one of the problems explored in Harman et al. [38]. Often, the natural metric for "dose" is the logarithm of the administered amount, as interest may lie in the effects of changes, such as doubling the dose. The cost function should be adjusted accordingly, though this modification does not affect the numerical method used to compute optimal designs. Then, the constraint is linear in $\mathbf{w}$, making it straightforward to incorporate into a semidefinite programming problem.

Finally, we address *probability-constrained designs*, which allow for the exclusion of doses that result in, for example, lower values of efficacy without toxicity, denoted as $p_{y=1,z=0}$. Naturally, other probabilities listed in (3) can also be constrained. To establish this setup, let us consider that the probability of efficacy without toxicity should be at least τ for a dose to be included in the experiment. This constraint can be formulated as:

$$w_i = 0, \quad i \in \mathbb{I} \equiv \{i \mid i \in [\![n_x]\!], p_{y=1,z=0}(x_i, \boldsymbol{\theta}) \leqslant \tau\}. \tag{17}$$

In this Equation, $\mathbb{I}$ represents the set of indices corresponding to doses where the probability of efficacy without toxicity is less than or equal to the threshold τ. In practice, *dose-constrained designs* are a specific subset of *probability-constrained designs*, derived using Eq. (2). The distinction lies in their implementation: *dose-constrained designs* directly restrict the candidate dose levels, while *probability-constrained designs* impose constraints on a functional transformation of these levels.

In our work, we solved the Semidefinite Programming (SDP) problems using the CVX environment (version 2.2) in combination with the `Mosek` solver (version 10), which employs an efficient Interior Point algorithm [39]. The relative and absolute tolerances for solving the SDP were set to 1×10^{-5}. All computations in Section 4 were performed on a machine equipped with an AMD Ryzen 7 5800X 8-core processor, running a 64-bit Windows 10 operating system with a clock speed of 3.80 GHz and 32 GB of RAM.

4. Application examples

This Section presents the results of the application of the formulations introduced in Section 3. First, in Section 4.1, we consider unconstrained designs. Then, in Section 4.2, we specifically address constrained designs of the classes listed in Section 3.2.

The reference setup considers a single drug with doses expressed in a unitary domain, i.e. $x \in [0, 1]$. The functions $\eta_1(x, \boldsymbol{\theta}_1)$ and $\eta_2(x, \boldsymbol{\theta}_2)$ are linear, defined as follows:

$$\eta_1(x, \boldsymbol{\theta}_1) = \boldsymbol{\theta}_1^\intercal \cdot \mathbf{f}_1(x) = (\theta_{1,0},\ \theta_{1,1}) \cdot (1,\ x)^\intercal = \theta_{1,0} + \theta_{1,1} \cdot x \tag{18a}$$

$$\eta_2(x, \boldsymbol{\theta}_2) = \boldsymbol{\theta}_2^\intercal \cdot \mathbf{f}_2(x) = (\theta_{2,0},\ \theta_{2,1}) \cdot (1,\ x)^\intercal = \theta_{2,0} + \theta_{2,1} \cdot x. \tag{18b}$$

In scenarios where ρ is known, the complete model includes four parameters $\boldsymbol{\theta} = (\theta_{1,0},\ \theta_{1,1},\ \theta_{2,0},\ \theta_{2,1})^\intercal$, which are to be determined from experiments. In scenarios where ρ is unknown, $\boldsymbol{\theta}$ includes additionally ρ, and n_θ becomes 5. The locally optimal designs obtained for single drug scenarios consider the vector of postulated parameter values $\mathbf{q}_0 = (-0.9, 7.0, -1.2, 1.6)^\intercal$ and $\rho = 0.5$. This vector of parameters was proposed in Fedorov and Leonov [19, Chap. 6] for a similar model, and we have also used it for reference in our formulations. The design domain is discretized with $\Delta x = 0.0025$, which leads to $n_x = 401$. Naturally, other values of $\mathbf{q}_0$, ρ and Δx can be used.

4.1. Results for unconstrained designs

In this Section, we present results for unconstrained designs, focusing on the impact of: (i) the correlation parameter ρ in setups where it is known, (ii) whether ρ is known *a priori* or estimated, and (iii) the use of a combination of two drugs compared to a single drug, on the optimal designs.

In the matrices representing the "optimal designs" in the result tables of the following sections, the bottom row corresponds to the weights, while the other rows represent the dosing levels for the first drug and, where applicable, the second drug.

Table 1
Optimal designs for various values of ρ, computed using the D–, A–, E–, and K–optimality criteria. The setup parameters are $\mathbf{q}_0 = (-0.9, 7.0, -1.2, 1.6)^\intercal$, $\mathbf{X} = [0, 1]$, and $\Delta x = 0.0025$.

Criterion	Optimal design	Optimum
$\rho = 0.0$		
D–	$\begin{pmatrix} 0.0000 & 0.3000 & 1.0000 \\ 0.4132 & 0.3542 & 0.2326 \end{pmatrix}$	9.5840×10^{-2}
A–	$\begin{pmatrix} 0.0000 & 0.3575 & 1.0000 \\ 0.4073 & 0.4886 & 0.1041 \end{pmatrix}$	1.4046×10^{2}
E–	$\begin{pmatrix} 0.0000 & 0.3625 \\ 0.4223 & 0.5777 \end{pmatrix}$	1.0149×10^{-2}
K–	$\begin{pmatrix} 0.0000 & 0.3400 \\ 0.5082 & 0.4918 \end{pmatrix}$	4.8118×10^{1}
$\rho = 0.2$		
D–	$\begin{pmatrix} 0.0000 & 0.3000 & 1.0000 \\ 0.4134 & 0.3541 & 0.2325 \end{pmatrix}$	9.6280×10^{-2}
A–	$\begin{pmatrix} 0.0000 & 0.3550 & 1.0000 \\ 0.4085 & 0.4864 & 0.1051 \end{pmatrix}$	1.3980×10^{2}
E–	$\begin{pmatrix} 0.0000 & 0.3650 & 1.0000 \\ 0.4224 & 0.5723 & 0.0053 \end{pmatrix}$	1.0008×10^{-2}
K–	$\begin{pmatrix} 0.0000 & 0.3500 \\ 0.4981 & 0.5019 \end{pmatrix}$	5.0263×10^{1}
$\rho = 0.5$		
D–	$\begin{pmatrix} 0.0000 & 0.2950 & 1.0000 \\ 0.4128 & 0.3539 & 0.2333 \end{pmatrix}$	9.8790×10^{-2}
A–	$\begin{pmatrix} 0.0000 & 0.3525 & 1.0000 \\ 0.4047 & 0.4855 & 0.1098 \end{pmatrix}$	1.3677×10^{2}
E–	$\begin{pmatrix} 0.0000 & 0.3625 & 1.0000 \\ 0.4153 & 0.5559 & 0.0288 \end{pmatrix}$	9.8600×10^{-3}
K–	$\begin{pmatrix} 0.0000 & 0.3600 & 1.0000 \\ 0.4522 & 0.5263 & 0.0215 \end{pmatrix}$	5.6437×10^{1}
$\rho = 0.6$		
D–	$\begin{pmatrix} 0.0000 & 0.2925 & 1.0000 \\ 0.4124 & 0.3541 & 0.2335 \end{pmatrix}$	1.0027×10^{-1}
A–	$\begin{pmatrix} 0.0000 & 0.3500 & 1.0000 \\ 0.4021 & 0.4855 & 0.1124 \end{pmatrix}$	1.3540×10^{2}
E–	$\begin{pmatrix} 0.0000 & 0.3600 & 1.0000 \\ 0.4128 & 0.5497 & 0.0375 \end{pmatrix}$	9.8433×10^{-3}
K–	$\begin{pmatrix} 0.0000 & 0.3625 & 1.0000 \\ 0.4323 & 0.5380 & 0.0297 \end{pmatrix}$	5.9123×10^{1}

First, we analyze the impact of the correlation parameter. To do so, we varied ρ discretely in $\{0.0, 0.2, 0.5, 0.6\}$, where $\rho = 0.0$ indicates no correlation between efficacy and toxicity, and $\rho = 0.6$ indicates moderate correlation. The optimal designs obtained are displayed in Table 1. All the designs, except the K–optimal and E–optimal designs for small values of ρ, are based on three support points. The exceptions require only two support points. As we have four parameters and observe two response variables per experiment, saturated designs (designs with the minimum number of support points) would require only two. Thus, most of the designs have one additional support point. The last column of Table 1 reports the optimum at convergence.

The results suggest that the value of ρ has a minimal impact on the optimal designs, with only slight variations observed. Specifically, the weights at $x = 1$ for the E- and K–optimality criteria remain very small, with a maximum value of 0.0373. The observed optima follow the anticipated trends: (i) the optima increase with increasing ρ for D- and K–optimality; and (ii) the optima decrease with increasing ρ for A- and E–optimality. Lastly, the collapse of the support point at $x = 1$ for the E- and K–optimality criteria is attributed to $\lambda_{\min}[\mathcal{M}(\xi, \boldsymbol{\theta})]$ becoming too small.

To evaluate the optimality of the designs presented in Table 1, we examine the scenario where $\rho = 0.5$, corresponding to the third tableau in the table. We plot the dispersion functions given by Eq. (11) for each optimality criterion. As shown in Fig. 1, the equivalence theorems are confirmed: the dispersion functions are below zero, with their maxima

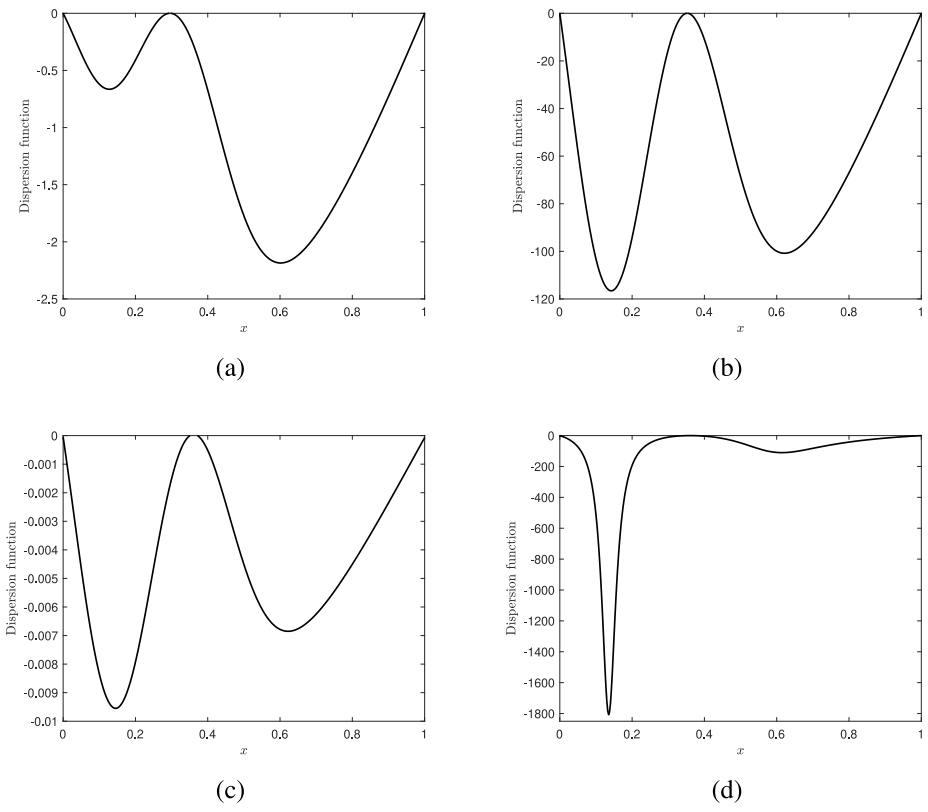

Fig. 1. Dispersion functions for optimal designs found for $\mathbf{q}_0 = (-0.9, 7.0, -1.2, 1.6)^\intercal$, $\rho = 0.5$, $\mathbf{X} = [0, 1]$, $\Delta x = 0.0025$: (a) D–optimality criterion; (b) A–optimality criterion; (c) E–optimality criterion; (d) K–optimality criterion.

occurring at the support points. This pattern is also confirmed for the K–optimality criterion.

The Supplementary Material presents a sensitivity analysis of the design efficiency, examining the effects of variation of model parameters. The efficiency for all optimality criteria was computed using the formulas in Eq. (10). The reference designs were those obtained for the singleton $\mathbf{q}_0 = (-0.9, 7.0, -1.2, 1.6)^\intercal$ and $\rho = 0.5$. In all simulations, we considered $\mathbf{X} = [0, 1]$ and $\Delta x = 0.0025$ (i.e., $n_x = 401$). The analysis of the results reveals a relative loss in efficiency as the parameters vary, consistent with theoretical expectations.

Next, we consider the case where ρ is an additional parameter to be estimated from the data. Thus, the parameter vector now includes an extra element: $\boldsymbol{\theta} = (\theta_{1,0}, \theta_{1,1}, \theta_{2,0}, \theta_{2,1}, \rho)^\intercal$. The first four elements remain the same as in the previous setup, and the fifth element corresponds to ρ. The optimal designs are determined for the vector $\mathbf{q}_0 = (-0.9, 7.0, -1.2, 1.6, 0.5)^\intercal$, where $\rho = 0.5$. All other parameters used in the algorithm remain constant. For this scenario, we use Eq. (8) instead of Eq. (7a) to compute the local Fisher Information Matrices.

The results are presented in Table 2. They are similar to those obtained when ρ is known and set to 0.5 (refer to the third tableau in Table 1). All variations are minor, but are greatest for the D–optimality criterion.

Finally, we consider the combination of two drugs, where the dose of drug 1 is represented by x_1 and the dose of drug 2 by x_2. The design space is defined as $\mathbf{X} = [0, 1]^2$. To discretize this space, we use an equispaced grid with interval $\Delta x = 0.01$ in both domains resulting in a total of 10 201 candidate points $\mathbf{x} = (x_1,\ x_2)^\intercal$.

The linear response models account for both first-order and interaction terms to capture the potential combined effects of the drugs. Specifically, the models are defined as follows:

$$\begin{aligned}\eta_1(\mathbf{x}, \boldsymbol{\theta}_1) &= \boldsymbol{\theta}_1^\intercal \cdot \mathbf{f}_1(\mathbf{x}) = (\theta_{1,0},\ \theta_{1,1},\ \theta_{1,2},\ \theta_{1,3}) \cdot (1,\ x_1,\ x_2,\ x_1 \cdot x_2)^\intercal \\ &= \theta_{1,0} + \theta_{1,1} \cdot x_1 + \theta_{1,2} \cdot x_2 + \theta_{1,3} \cdot x_1 \cdot x_2 \end{aligned} \tag{19a}$$

Table 2
Optimal designs for various values of ρ, computed using the D–, A–, E–, and K–optimality criteria. The setup parameters are $\mathbf{q}_0 = (-0.9, 7.0, -1.2, 1.6)^\intercal$, $\mathbf{X} = [0, 1]$, and $\Delta x = 0.0025$; ρ is an additional parameter to be estimated.

Criterion	Optimal design			Optimum
D–	0.0000	0.2580	1.0000	1.1524×10^{-1}
	0.4499	0.3628	0.1873	
A–	0.0000	0.3500	1.0000	1.4280×10^{2}
	0.4142	0.4789	0.1069	
E–	0.0000	0.3625	1.0000	9.8146×10^{-3}
	0.4157	0.5552	0.0291	
K–	0.0000	0.3605	1.0000	5.6799×10^{1}
	0.4522	0.5261	0.0217	

$$\begin{aligned}\eta_2(\mathbf{x}, \boldsymbol{\theta}_2) &= \boldsymbol{\theta}_2^\intercal \cdot \mathbf{f}_2(\mathbf{x}) = (\theta_{2,0},\ \theta_{2,1},\ \theta_{2,2},\ \theta_{2,3}) \cdot (1,\ x_1,\ x_2,\ x_1 \cdot x_2)^\intercal \\ &= \theta_{1,0} + \theta_{2,1} \cdot x_1 + \theta_{2,2} \cdot x_2 + \theta_{2,3} \cdot x_1 \cdot x_2. \end{aligned} \tag{19b}$$

Here, $\boldsymbol{\theta}$ includes 8 parameters, with ρ known and set to 0.5. Table 3 presents the locally optimal designs obtained for the singleton $\mathbf{q}_0 = (-0.9, 7.0, -2.0, 0.8, -1.2, 1.6, -0.5, 0.5)^\intercal$. The D–, A–, and E–optimal designs have 6 support points, with the doses of drug 2 typically at either its minimum ($x_2 = 0$) or maximum ($x_2 = 1$). The K–optimal design requires only 5 support points. Notably, the saturated designs have 4 support points. Once again, the weight at points of maximum dose for drug 1 ($x_1 = 1$) is low, similar to that observed for the single drug case. The designs are far from the product designs often found for linear models.

4.2. Results for constrained designs

In this Section, we determine constrained optimal designs. We consider a single drug with a linear response represented by model (18).

Table 3
Optimal designs for various values of ρ, computed using the D–, A–, E–, and K–optimality criteria. The setup parameters are $\mathbf{q}_0 = (-0.9, 7.0, -2.0, 0.8, -1.2, 1.6, -0.5, 0.5)^\intercal$, $\rho = 0.5$, $\mathbf{X} = [0,1] \times [0,1]$, $\Delta\mathbf{x} = (0.01,\ 0.01)^\intercal$.

Criterion	Optimal design						Optimum
D–	0.0000	0.2155	0.2970	0.5100	1.0000	1.0000	4.5174×10^{-2}
	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	
	0.2065	0.2134	0.1772	0.1768	0.1164	0.1097	
A–	0.0000	0.1600	0.3500	0.5700	1.0000	1.0000	8.6440×10^{2}
	0.0000	1.0000	0.0200	1.0000	0.0000	1.0000	
	0.2290	0.2272	0.2715	0.1685	0.0617	0.0421	
E–	0.0000	0.1600	0.3600	0.5800	1.0000	1.0000	1.8877×10^{-3}
	0.0000	0.9300	0.0700	1.0000	0.0000	1.0000	
	0.2523	0.2097	0.3093	0.2038	0.0191	0.0058	
K–	0.0000	0.0900	0.3600	0.5600	1.0000		3.1970×10^{2}
	0.0000	0.8900	0.0000	1.0000	0.0000		
	0.2901	0.3022	0.2940	0.1020	0.0117		

Table 4
Dose constrained optimal designs for $\rho = 0.5$ computed using the D–, A–, E–, and K–optimality criteria. The setup parameters are $\mathbf{q}_0 = (-0.9, 7.0, -1.2, 1.6)^\intercal$, $\mathbf{X} = [0, 0.65]$, $\Delta x = 0.0025$, $x_{\text{lim}}^{UP} = 0.65$.

$\rho = 0.5$

Criterion	Optimal design			Optimum
D–	0.0000	0.3050	0.6500	8.0131×10^{-2}
	0.4536	0.3397	0.2067	
A–	0.0000	0.3575	0.6500	1.6000×10^{2}
	0.4283	0.4538	0.1179	
E–	0.0000	0.3700	0.6500	9.5140×10^{-3}
	0.4294	0.5473	0.0233	
K–	0.0000	0.3725	0.6500	5.7986×10^{1}
	0.4655	0.5265	0.0080	

Table 5
Budget constrained optimal designs for $\rho = 0.5$ computed using the D–, A–, E–, and K–optimality criteria. The setup parameters are $\mathbf{q}_0 = (-0.9, 7.0, -1.2, 1.6)^\intercal$, $\mathbf{X} = [0, 0.65]$, $\Delta x = 0.0025$, $\alpha = 0.3$, $\beta = 0.7$, $B = 0.4$.

$\rho = 0.5$

Criterion	Optimal design			Optimum
D-	0.0000	0.2775	1.0000	8.3630×10^{-2}
	0.6718	0.2563	0.0719	
A-	0.0000	0.3150	1.0000	1.6594×10^{2}
	0.6237	0.3408	0.0355	
E-	0.0000	0.3275	1.0000	8.8427×10^{-3}
	0.5844	0.4055	0.0101	
K-	0.0000	0.3400	1.0000	6.2299×10^{1}
	0.5937	0.3992	0.0071	

The discretization interval is $\Delta x = 0.0025$ and $\rho = 0.5$.

First, we consider dose-constrained designs described by Eq. (15), with x_{lim}^{UP} set to 0.65 and x_{lim}^{LO} to 0.0. The optimal designs obtained are shown in Table 4. Comparing these results with those in the third tableau of Table 1, we observe that the number of support points remains the same, but the weight of $x = 0.0$ increases regardless of the optimality criterion. As expected, one of the support points becomes coincident with the maximum dose limit allowed, x_{lim}^{UP}, with weight close to that when x_{lim}^{UP} is 1.

Now, we consider budget-constrained designs. The constraint imposed is represented by Eq. (16), and for simulation, we set $\alpha = 0.3$, $\beta = 0.7$, and $B = 0.4$. The results are presented in Table 5. The designs still utilize three support points, but the lower dose points now have larger weights at the expense of those with higher values of x. This outcome results from penalizing the cost of tests with higher doses to meet a limited budget.

Finally, we illustrate the process of finding probability-constrained designs. The constraint is defined by Eq. (17), and we set τ to 0.35. The

Table 6
Probability constrained optimal designs for $\rho = 0.5$. Setup: $\mathbf{q}_0 = (-0.9, 7.0, -1.2, 1.6)^\intercal$, $\mathbf{X} = [0, 0.65]$, $\Delta x = 0.0025$, $\tau = 0.35$.

$\rho = 0.5$

Criterion	Optimal design			Optimum
D-	0.1200	0.3425	0.9900	8.1762×10^{-2}
	0.4129	0.3502	0.2369	
A-	0.1200	0.3900	0.9900	2.4855×10^{2}
	0.3350	0.5675	0.0975	
E-	0.1200	0.3950	0.9900	5.1054×10^{-3}
	0.3443	0.6345	0.0212	
K-	0.1200	0.3900	0.9900	1.2107×10^{2}
	0.3643	0.6193	0.0164	

Table 7
Efficiency and CPU time required to compute constrained designs (reference designs shown in the third tableau of Table 1). The values in parentheses indicate CPU times, expressed in s.

Efficiency	Dose-constrained design	Budget-constrained design	Probability-constrained design
Eff_D	0.8129 (1.25)	0.8465 (1.76)	0.8276 (1.50)
Eff_A	0.8542 (1.10)	0.8236 (0.85)	0.5499 (0.98)
Eff_E	0.9649 (1.14)	0.8968 (0.78)	0.5178 (1.11)
Eff_K	0.9733 (1.44)	0.9059 (1.21)	0.4662 (1.51)

optimal design is presented in Table 6. Comparing with the reference design reveals that the lowest dose now tested is $x = 0.12$, as lower doses fail to meet the constraint (the probability of efficacy is below 0.35). The same reason holds for the choice of the highest dose tested – $x = 0.99$, not $x = 1.0$.

The Ψ-optimality efficiencies of the constrained designs are presented in Table 7. These efficiencies were computed using Eq. (10), with the reference designs being the unconstrained designs determined for $\rho = 0.5$, as listed in the third tableau of Table 1. As expected, all constrained designs exhibit a loss of efficiency. However, within the feasibility regions defined by the constraints, the designs found are optimal. Table 7 also reports the CPU times for computing constrained designs (see the values in parentheses), which are modest and comparable to those for unconstrained designs.

The strategy described above also highlights the simplicity of solving constrained optimal design problems. For linear constraints on weights, these are directly included in the design problem as additional constraints. Consequently, the basic formulation remains unchanged, but is supplemented with additional equality or inequality constraints, which are converted to active structures in a way similar to those defining the original design problem.

Note that the constraints apply only to the probabilities in (3). For example, $p_{y=1,z=0}(x_i, \boldsymbol{\theta}) \leqslant \tau$ restricts the probability of efficacy without toxicity, rather than directly limiting toxicity. These constraints can only be derived for the probabilities in (3). Specifically, stating $p_{y=1,z=0}(x_i, \boldsymbol{\theta}) \leqslant \tau$ imposes a limit on the probability of efficacy without toxicity, not on toxicity itself.

This approach extends to models with both efficacy and toxicity constraints. In this case, the set $\mathbb{I}$ in Eq. (17) is reduced. For instance, if the probability of efficacy without toxicity is bounded below by τ and the probability of efficacy with toxicity is bounded above by ϑ, the candidate dose set for optimal design becomes

$$w_i = 0, \quad i \in \mathbb{I} \equiv \{i \mid i \in [\![n_x]\!], p_{y=1,z=0}(x_i, \boldsymbol{\theta}) \leqslant \tau \cap p_{y=1,z=1}(x_i, \boldsymbol{\theta}) \geqslant \vartheta\}. \tag{20}$$

5. Conclusions

We have addressed the challenge of systematically finding approximate D–, A–, E–, and K–optimal designs for phase I and II clinical trials aimed at establishing the efficacy and toxicity profiles of agents. We consider a model with a bivariate response, where the outcomes are toxicity and efficacy, and the control factor is the scaled dose level. The response variables are binary, and the model relating the probability of toxicity and efficacy to the dose is of the probit type, with correlated outcomes.

To find optimal designs on a previously discretized grid of candidate (dose) points, we utilize Semidefinite Programming. Our approach accommodates both unconstrained and constrained designs by appending additional constraints on weights without altering the basic formulation. In our study of unconstrained designs, we explored: (i) the significance of the correlation between outcomes; (ii) the impact of knowing the correlation coefficient a priori versus estimating it from experiments; and (iii) the extension from single-drug design spaces to combined-drug problems.

For constrained designs, we examined: (i) dose-constrained designs; (ii) budget-constrained designs; and (iii) probability-constrained designs.

We demonstrate the optimality of the unconstrained designs using equivalence theorems. For the D–, A–, and E–optimality criteria, we applied well-established results from the literature. Additionally, we derived an equivalence theorem for the K–optimality criterion, which, to the best of our knowledge, represents a novel contribution. Our formulations efficiently solved all design problems, encompassing single- and double-drug setups as well as constrained designs. Single-drug problems required approximately 1 s of CPU time across all cases. Constrained designs exhibited similar computational times, while double-drug problems were solved in less than 10 s.

This work also opens several research avenues for future exploration. One particular topic worth further investigation is the design of optimal experiments for drug combinations. By exploring the combined effects of drugs, researchers can aim to improve efficacy while limiting toxicity, as described by Mihich and Grindey [40] in the context of chemotherapy. Another promising area involves maximizing the information extracted from a study while considering various constraints. These constraints often include toxicity limitations, as discussed by Lee et al. [41] and Boston and Gaffney [42]. Finally, in the absence of specific knowledge required for designing optimal allocation schemes, adaptive approaches may be crucial. Methods suggested by Dragalin and Fedorov [43] and Ji and Wang [44] could ensure that the knowledge gained is effectively used to determine the most appropriate dose for subsequent individuals.

Numerical results for the adaptive procedure of Dragalin and Fedorov [43] for one and two-component treatments are given in Fedorov et al. [2, §8.3] who use D–optimality and update using the results from treating one patient at a time. Often, patients will be allocated treatments in cohorts, which will reduce the efficiency demonstrated in these simulations of adaptive designs. The combination of adaptive designs and constraints on toxicity is, theoretically, an attractive safeguard against a poor initial determination of the maximum tolerated dose. However it is important that, although the effect of treatment may be almost immediate, toxicities may be delayed. See for example Liu and Ning [45], who use a logistic model for the relationship of efficacy and toxicity. The papers in Sverdlov [46] discuss many aspects of adaptive design in clinical trials, but with an emphasis on Phase III.

CRediT authorship contribution statement

Belmiro P.M. Duarte: Writing – review & editing, Writing – original draft, Software, Methodology, Formal analysis, Conceptualization. **Anthony C. Atkinson:** Writing – review & editing, Validation, Methodology, Formal analysis.

Ethical statement for solid state ionics

Hereby, Belmiro Duarte consciously assure that for the manuscript "Optimal designs for efficacy-toxicity response in dose finding studies using the bivariate probit model" the following is fulfilled:

(1) This material is the authors' own original work, which has not been previously published elsewhere.

(2) The paper is not currently being considered for publication elsewhere.

(3) The paper reflects the authors' own research and analysis in a truthful and complete manner.

(4) The paper properly credits the meaningful contributions of co-authors and co-researchers.

(5) The results are appropriately placed in the context of prior and existing research.

(6) All sources used are properly disclosed (correct citation). Literally copying of text must be indicated as such by using quotation marks and giving proper reference.

(7) All authors have been personally and actively involved in substantial work leading to the paper, and will take public responsibility for its content.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Equivalence for K–optimality criterion

Theorem 1. *Let ξ_K^* be such that both the largest and the smallest eigenvalue of $\mathcal{M}(\xi_K^*)$ have multiplicity one. Then ξ_K^* is K–optimal if and only if the following condition is satisfied:*

The inequality

$$\kappa[\mathcal{M}(\xi_K^*)] - \frac{\mathrm{tr}[M(\mathbf{x})Q^{\max}]}{\mathrm{tr}[M(\mathbf{x})Q^{\min}]} \leqslant 0 \tag{A.1}$$

holds for all $\mathbf{x} \in \mathbf{X}$, with strict equality at the support points of ξ_K^*.

Here, $\mathcal{M}(\xi_K^*)$ denotes the FIM corresponding to the design ξ_K^*, $\kappa[\mathcal{M}(\xi_K^*)]$ is the condition number of $\mathcal{M}(\xi_K^*)$, $M(\mathbf{x})$ is the local Fisher information matrix at the design point $\mathbf{x}$, $Q^{\max} = \mathbf{v}_{\max}\mathbf{v}_{\max}^{\intercal}$ and $Q^{\min} = \mathbf{v}_{\min}\mathbf{v}_{\min}^{\intercal}$, $\mathbf{v}_{\max}$ and $\mathbf{v}_{\min}$ are standardized eigenvectors of length 1 of $\mathcal{M}(\xi_K^*)$ associated with $\lambda_{\max}^*$ and $\lambda_{\min}^*$, respectively.

Proof. Let us consider an experimental design

$$\xi = \begin{pmatrix} \mathbf{x}_i \\ w_i \end{pmatrix}_{i \in [\![I]\!]}$$

with I support points $\mathbf{x}_i$ each one with weight w_i. Let the model representing the response being linear, i.e.

$$\mathbb{E}(y) = \boldsymbol{\theta}^{\intercal} \cdot \mathbf{f}(\mathbf{x}) \tag{A.2}$$

$\mathbf{f}(\mathbf{x})$ is a vector of polynomial terms and $\mathbf{x}$ a set of covariates.

The global FIM of ξ_K^* is

$$\mathcal{M}(\xi_K^*) = \sum_{i=1}^{I} w_i^* \cdot \mathbf{f}(\mathbf{x}_i^*) \cdot \mathbf{f}(\mathbf{x}_i^*)^{\intercal}, \tag{A.3}$$

where w_i^* designate the weights in ξ^* and $\mathbf{x}_i^*$ the support points.

The K–optimality criterion is such that $\xi_K^* = \arg\min_{\xi} \kappa[\mathcal{M}(\xi)]$. Now we use the perturbation method to construct the global FIM of a design infinitely close to ξ_K^*, say $\xi_K^* + \varepsilon\, \delta_{\mathbf{x}}$ where ε is a positive small value and

$\boldsymbol{\delta}_{\mathbf{x}}$ is a Dirac measure in the domain of the covariates. The FIM becomes

$$\mathcal{M}(\xi_K^* + \varepsilon\ \boldsymbol{\delta}_{\mathbf{x}}) = \mathcal{M}(\xi_K^*) + \varepsilon\ \mathbf{f}(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x})^{\intercal}, \tag{A.4}$$

where $\mathbf{x}^*$ is point of the design space.

The variation in the eigenvalues of the FIM is

$$\Delta\lambda_i \approx \mathbf{v}_i^{\intercal} \left[\varepsilon\ \mathbf{f}(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x})^{\intercal}\right] \mathbf{v}_i, \quad i \in [\![I]\!] \tag{A.5}$$

where $\Delta\lambda_i$ is the variation of the ith eigenvalue and $\mathbf{v}_i$ the corresponding eigenvector. The application of (A.5) to maximum and minimum eigenvalues of the FIM for the K–optimal design produces

$$\Delta\lambda_{\max} \approx \mathbf{v}_{\max}^{\intercal} \left[\varepsilon\ \mathbf{f}(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x})^{\intercal}\right] \mathbf{v}_{\max} \tag{A.6a}$$

$$\Delta\lambda_{\min} \approx \mathbf{v}_{\min}^{\intercal} \left[\varepsilon\ \mathbf{f}(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x})^{\intercal}\right] \mathbf{v}_{\min} \tag{A.6b}$$

Let $\lambda'_{\max} = \lambda^*_{\max} + \Delta\lambda_{\max}$ and $\lambda'_{\min} = \lambda^*_{\min} + \Delta\lambda_{\min}$. Then, the perturbation in condition number is

$$\begin{aligned} \kappa[\mathcal{M}(\xi_K^* + \varepsilon\ \boldsymbol{\delta}_{\mathbf{x}})] &= \frac{\lambda'_{\max}}{\lambda'_{\min}} = \frac{\lambda^*_{\max} + \mathbf{v}_{\max}^{\intercal} \left[\varepsilon\ \mathbf{f}(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x})^{\intercal}\right] \mathbf{v}_{\max}}{\lambda^*_{\min} + \mathbf{v}_{\min}^{\intercal} \left[\varepsilon\ \mathbf{f}(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x})^{\intercal}\right] \mathbf{v}_{\min}} \\ &= \frac{\lambda^*_{\max}}{\lambda^*_{\min}} \cdot \frac{\left(1 + \varepsilon\ \mathbf{v}_{\max}^{\intercal} \left[\mathbf{f}(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x})^{\intercal}\right] \mathbf{v}_{\max} / \lambda^*_{\max}\right)}{\left(1 + \varepsilon\ \mathbf{v}_{\min}^{\intercal} \left[\mathbf{f}(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x})^{\intercal}\right] \mathbf{v}_{\min} / \lambda^*_{\min}\right)} \end{aligned} \tag{A.7}$$

If ξ_K^* is optimal, then

$$\frac{\lambda^*_{\max}}{\lambda^*_{\min}} \cdot \frac{\left(1 + \varepsilon\ \mathbf{v}_{\max}^{\intercal} \left[\mathbf{f}(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x})^{\intercal}\right] \mathbf{v}_{\max} / \lambda^*_{\max}\right)}{\left(1 + \varepsilon\ \mathbf{v}_{\min}^{\intercal} \left[\mathbf{f}(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x})^{\intercal}\right] \mathbf{v}_{\min} / \lambda^*_{\min}\right)} \geqslant \frac{\lambda^*_{\max}}{\lambda^*_{\min}} \tag{A.8}$$

which after algebraic manipulation yields

$$\frac{\mathbf{v}_{\max}^{\intercal} \left[\mathbf{f}(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x})^{\intercal}\right] \mathbf{v}_{\max}}{\mathbf{v}_{\min}^{\intercal} \left[\mathbf{f}(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x})^{\intercal}\right] \mathbf{v}_{\min}} \geqslant \frac{\lambda^*_{\max}}{\lambda^*_{\min}}. \tag{A.9}$$

Let $Q^{\max} = \mathbf{v}_{\max}^{\intercal} \cdot \mathbf{v}_{\max}$, $Q^{\min} = \mathbf{v}_{\min}^{\intercal} \cdot \mathbf{v}_{\min}$, $\mathbf{x} = \mathbf{x}^* + \boldsymbol{\delta}_{\mathbf{x}}$ be a generic point of the design space, and $M(\mathbf{x}) = \mathbf{f}(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x})^{\intercal}$ be the local FIM at $\mathbf{x}$. Then, (A.9) leads to

$$\frac{\mathrm{tr}[M(\mathbf{x}) \cdot Q^{\max}]}{\mathrm{tr}[M(\mathbf{x}) \cdot Q^{\min}]} \geqslant \kappa[\mathcal{M}(\xi_K^*)] \tag{A.10}$$

Finally, we obtain the equivalence theorem

$$\kappa[\mathcal{M}(\xi_K^*)] - \frac{\mathrm{tr}[M(\mathbf{x}) \cdot Q^{\max}]}{\mathrm{tr}[M(\mathbf{x}) \cdot Q^{\min}]} \leqslant 0 \quad \mathbf{x} \in \mathbf{X}. \quad \square \tag{A.11}$$

Appendix B. Formulations to determine the optimal allocation via Semidefinite Programming

Here, we list the SDP formulations for the D–, A– and E–optimality criteria. The first three were introduced in Vandenberghe and Boyd [31, 47] and Ben-Tal and Nemirovski [28]. We start with the formulation for D–optimal designs:

$$\mathrm{Opt} \equiv \max_{\mathbf{w},\mathcal{B},t} \quad t \tag{B.1a}$$

$$\text{s.t.} \quad \begin{pmatrix} \mathcal{M}(\xi) & \mathcal{B}^{\intercal} \\ \mathcal{B} & \mathrm{diag}(\mathcal{B}) \end{pmatrix} \succeq 0 \tag{B.1b}$$

$$t \leqslant \prod_{i=1}^{n_\theta} \mathcal{B}_{i,i}^{1/n_\theta} \tag{B.1c}$$

$$\sum_{i=1}^{k} w_i = 1 \tag{B.1d}$$

$$0 \leqslant w_i \leqslant 1, \quad i \in \{1, \ldots, k\}. \tag{B.1e}$$

The formulation for computing A–optimal designs is:

$$\mathrm{Opt} \equiv \min_{\mathbf{w},\mathcal{B},t} \quad t \tag{B.2a}$$

$$\text{s.t.} \quad \begin{pmatrix} \mathcal{M}(\xi) & I_{n_\theta} \\ I_{n_\theta} & \mathcal{B} \end{pmatrix} \succeq 0 \tag{B.2b}$$

$$t \geqslant \sum_{i=1}^{n_\theta} \mathcal{B}_{i,i} \tag{B.2c}$$

$$\sum_{i=1}^{k} w_i = 1 \tag{B.2d}$$

$$0 \leqslant w_i \leqslant 1, \quad i \in \{1, \ldots, k\}, \tag{B.2e}$$

Finally, for E–optimal designs, we have:

$$\mathrm{Opt} \equiv \max_{\mathbf{w},t} \quad t \tag{B.3a}$$

$$\text{s.t.} \quad \mathcal{M}(\xi) - t\ I_{n_\theta} \succeq 0 \tag{B.3b}$$

$$\sum_{i=1}^{k} w_i = 1 \tag{B.3c}$$

$$0 \leqslant w_i \leqslant 1, \quad i \in \{1, \ldots, k\}. \tag{B.3d}$$

Appendix C. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.compbiomed.2025.109848.

References

[1] V.V. Fedorov, Y. Wu, Dose finding designs for continuous responses and binary utility, J. Biopharm. Statist. 17 (6) (2007) 1085–1096.

[2] V. Fedorov, Y. Wu, R. Zhang, Optimal dose-finding designs with correlated continuous and discrete responses, Stat. Med. 31 (3) (2012) 217–234.

[3] B. Bornkamp, F. Bretz, A. Dmitrienko, G. Enas, B. Gaydos, C.-H. Hsu, F. König, M. Krams, Q. Liu, B. Neuenschwander, et al., Innovative approaches for designing and analyzing adaptive dose-ranging trials, J. Biopharm. Statist. 17 (6) (2007) 965–995.

[4] S. Chevret, Statistical methods for dose-finding experiments, Statistics in Practice, Wiley, 2006.

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

[6] J. O'Quigley, M.D. Hughes, T. Fenton, Dose-finding designs for HIV studies, Biometrics 57 (4) (2001) 1018–1029.

[7] P.F. Thall, K.E. Russell, A strategy for dose-finding and safety monitoring based on efficacy and adverse outcomes in phase I/II clinical trials, Biometrics 54 (1) (1998) 251–264.

[8] W. Zhang, D.J. Sargent, S. Mandrekar, An adaptive dose-finding design incorporating both toxicity and efficacy, Stat. Med. 25 (14) (2006) 2365–2383, http://dx.doi.org/10.1002/sim.2325.

[9] T.M. Braun, The bivariate continual reassessment method: extending the CRM to phase I trials of two competing outcomes, Controlled Clin. Trials 23 (3) (2002) 240–256.

[10] V. Dragalin, V. Fedorov, Adaptive designs for dose-finding based on efficacy–toxicity response, J. Statist. Plann. Inference 136 (6) (2006) 1800–1823.

[11] V. Dragalin, V. Fedorov, Y. Wu, Adaptive designs for selecting drug combinations based on efficacy-toxicity response, J. Statist. Plann. Inference 138 (2) (2008) 352–373, Special Issue: Statistical Design and Analysis in the Health Sciences.

[12] A. Tao, A Dose Finding Method in Joint Modeling of Efficacy and Safety Endpoints in Phase II Studies – An Extension of the MCP-MOD Method (Ph.D. thesis), Rutgers, The State University of New Jersey, New Brunswick, NJ, USA, 2010.

[13] Y. Tao, J. Liu, Z. Li, J. Lin, T. Lu, F. Yan, Dose-finding based on bivariate efficacy-toxicity outcome using archimedean copula, PLoS One 8 (11) (2013) e78805.

[14] Y. Du, J. Yin, D.J. Sargent, S.J. Mandrekar, An adaptive multi-stage phase I dose-finding design incorporating continuous efficacy and toxicity data from multiple treatment cycles, J. Biopharm. Statist. 29 (2) (2019) 271–286.

[15] J. Yin, R. Qin, M. Ezzalfani, D.J. Sargent, S.J. Mandrekar, A Bayesian dose-finding design incorporating toxicity data from multiple treatment cycles, Stat. Med. 36 (1) (2017) 67–80.

[16] D. Zhang, J. Xu, A Bayesian design for finding optimal biological dose with mixed types of responses of toxicity and efficacy, Contemp. Clin. Trials 127 (2023) 107113.

[17] C.I. Bliss, The method of probits, Sci. 79 (2037) (1934) 38–39.

[18] F. Pukelsheim, Optimal Design of Experiments, Wiley, New York, USA, 1993.

[19] V.V. Fedorov, S.L. Leonov, Optimal Design for Nonlinear Response Models, Chapman and Hall/CRC Press, Boca Raton, 2014.

[20] N. Mantel, Part IV. The concept of threshold in carcinogenesis, Clin. Pharmacol. Ther. 4 (1) (1963) 104–109, http://dx.doi.org/10.1002/cpt196341104.

[21] R. Hartung, Dose—Response Relationships, Springer US, New York, NY, 1987, pp. 29–46, http://dx.doi.org/10.1007/978-1-4684-5290-7_3.

[22] V.V. Fedorov, Convex design theory, Math. Oper. Stat. Ser. Stat. 11 (1980) 403–413.
[23] J.C. Kiefer, Optimum experimental designs, J. R. Stat. Soc. Ser. B 21 (1959) 272–319.
[24] J. Kiefer, J. Wolfowitz, The equivalence of two extremum problem, Canad. J. Math. 12 (1960) 363–366.
[25] P. Whittle, Some general points in the theory of optimal experimental design, J. R. Stat. Soc. Ser. B 35 (1973) 123–130.
[26] J. Kiefer, General equivalence theory for optimum design (approximate theory), Ann. Stat. 2 (1974) 849–879.
[27] Z. Yue, X. Zhang, P. Driessche, J. Zhou, Constructing K-optimal designs for regression models, Statist. Papers 64 (2023) 205–226.
[28] A. Ben-Tal, A.S. Nemirovski, Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications, Society for Industrial and Applied Mathematics, Philadelphia, 2001.
[29] S. Boyd, L. Vandenberghe, Convex Optimization, University Press, Cambridge, 2004.
[30] G. Sagnol, On the semidefinite representation of real functions applied to symmetric matrices, Linear Algebra Appl. 439 (10) (2013) 2829–2843.
[31] L. Vandenberghe, S. Boyd, Applications of semidefinite programming, Appl. Numer. Math. 29 (1999) 283–299.
[32] B.P.M. Duarte, W.K. Wong, Finding Bayesian optimal designs for nonlinear models: A semidefinite programming-based approach, Int. Stat. Rev. 83 (2) (2015) 239–262.
[33] J.J. Ye, J. Zhou, Minimizing the condition number to construct design points for polynomial regression models, SIAM J. Optim. 23 (1) (2013) 666–686.
[34] P. Maréchal, J.J. Ye, Optimizing condition numbers, SIAM J. Optim. 20 (2) (2009) 935–947, http://dx.doi.org/10.1137/080740544.
[35] Z. Lu, T.K. Pong, Minimizing condition number via convex programming, SIAM J. Matrix Anal. Appl. 32 (4) (2011) 1193–1211, http://dx.doi.org/10.1137/100795097.
[36] S. Boyd, L.E. Ghaoui, E. Feron, V. Balakrishnan, Linear matrix inequalities in system and control theory, Studies in Applied and Numerical Mathematics, SIAM, Philadelphia, 1994, URL https://epubs.siam.org/doi/book/10.1137/1.9781611970777.
[37] L.M. Haines, I. Perevozskaya, W.F. Rosenberger, Bayesian optimal designs for phase I clinical trials, Biometrics 59 (3) (2003) 591–600.
[38] R. Harman, A. Bachratá, L. Filová, Construction of efficient experimental designs under multiple resource constraints, Appl. Stoch. Models Bus. Ind. 32 (1) (2016) 3–17.
[39] Y. Ye, Interior Point Algorithms: Theory and Analysis, John Wiley & Sons, New York, 1997.
[40] E. Mihich, G.B. Grindey, Multiple basis of combination chemotherapy, Cancer 40 (S1) (1977) 534–543, http://dx.doi.org/10.1002/1097-0142(197707)40:1+<534::AID-CNCR2820400720>3.0.CO;2-3.
[41] S.M. Lee, M. Ursino, Y.K. Cheung, S. Zohar, Dose-finding designs for cumulative toxicities using multiple constraints, Biostatistics 20 (1) (2017) 17–29, http://dx.doi.org/10.1093/biostatistics/kxx059.
[42] E.A. Boston, E.A. Gaffney, The influence of toxicity constraints in models of chemotherapeutic protocol escalation, Math. Med. Biol.: A J. IMA 28 (4) (2011) 357–384.
[43] V. Dragalin, V. Fedorov, Adaptive designs for dose-finding based on efficacy-toxicity response, J. Statist. Plann. Inference 136 (6) (2006) 1800–1823, Adaptive Designs in Clinical Trials.
[44] Y. Ji, S.-J. Wang, Modified toxicity probability interval design: a safer and more reliable method than the 3+ 3 design for practical phase I trials, J. Clin. Oncol. 31 (14) (2013) 1785–1791.
[45] S. Liu, J. Ning, A Bayesian dose-finding design for drug combination trials with delayed toxicities, Bayesian Anal. 8 (2013) 703–722.
[46] O. Sverdlov (Ed.), Modern Adaptive Randomized Clinical Trials: Statistical and Practical Aspects, Chapman and Hall/ CRC Press, Boca Raton, 2015.
[47] L. Vandenberghe, S. Boyd, Semidefinite programming, SIAM Rev. 8 (1996) 49–95.