Statistics and Computing (2025) 35:169
https://doi.org/10.1007/s11222-025-10690-y

ORIGINAL PAPER

Information-Theoretic Criteria for Optimizing Designs of Individually Randomized Stepped-Wedge Clinical Trials

Belmiro P. M. Duarte[1,2,3] · **Anthony C. Atkinson**[4] · **Mirjam Moerbeek**[5]

Received: 19 March 2025 / Accepted: 22 July 2025

Abstract
Clinical trials are essential for advancing medical knowledge and improving health care, with Randomized Clinical Trials (RCTs) considered the gold standard for minimizing bias and generating reliable evidence on treatment efficacy and safety. Stepped-wedge individual RCTs, which randomize participants into sequences transitioning from control to intervention at staggered time points, are increasingly adopted. To improve their design, we propose an information-theoretic framework based on D– and A–optimality criteria for participant allocation to sequences. Our approach leverages semidefinite programming for automated computation and is applicable across a range of settings, varying in: (i) number of sequences, (ii) attrition rates, (iii) optimality criteria, (iv) error correlation structures, and (v) multi-objective designs using the ϵ-constraint method.

Keywords Optimal design of experiments · Clinical trials · Randomized stepped-wedge · Information-theoretic criteria · Correlation structure

Mathematics Subject Classification 62K05 · 90C47

1 Motivation

We address the challenge of determining optimal allocation schemes in stepped-wedge individual randomized clinical trials (RCTs). Building on recent contributions (Moerbeek 2023b; Wilson et al. 2023), we present a general framework grounded in information-theoretic criteria to guide the optimal design of such trials. Specifically, we exploit the connection between minimizing the covariance matrix of parameter estimates and maximizing information-based criteria, such as the D- and A-optimality criteria derived from the Fisher Information Matrix. Our approach casts the problem within the classical theory of model-based optimal design of experiments, supported by well-established theoretical foundations and computational tools. We also investigate the impact of varying modeling assumptions on the resulting designs, illustrating the flexibility and robustness of the proposed methodology in addressing a broad class of design problems in RCTs.

Clinical trials aim to systematically assess the safety, efficacy, and overall effectiveness of medical interventions—such as drugs, devices, procedures, or behavioral therapies—in humans. These trials are designed to answer critical questions about intervention performance under controlled conditions (Knifed et al. 2008). Optimizing clinical trial allocation is essential for enhancing both the ethical and scientific dimensions of clinical research. Key aspects include (Kadane 2011): (i) ethically allocating participants to minimize exposure to unsafe or ineffective treatments; (ii) ensuring scientific rigor by maximizing the trial's power to

✉ Belmiro P. M. Duarte
bduarte@isec.pt

Anthony C. Atkinson
A.C.Atkinson@lse.ac.uk

Mirjam Moerbeek
m.moerbeek@uu.nl

1 Polytechnic Institute of Coimbra, ISEC, Rua Pedro Nunes, 3030–199 Coimbra, Portugal

2 University of Coimbra, INESC-Coimbra, Rua Sílvio Lima — Pólo II, 3030–790 Coimbra, Portugal

3 Department of Chemical Engineering, University of Coimbra, CERES, Rua Sílvio Lima — Pólo II, 3030–790 Coimbra, Portugal

4 Department of Statistics, London School of Economics, London WC2A 2AE, United Kingdom

5 Department of Methodology and Statistics, Utrecht University, PO Box 80140, 3508, TC Utrecht, The Netherlands

Published online: 11 August 2025

detect treatment effects; and (iii) adhering to regulatory standards and ethical guidelines.

Randomization plays a crucial role in clinical trial allocation as it provides several key advantages: (i) Minimizing bias; (ii) Controlling for confounding variables; (iii) Ensuring statistical validity; (iv) Addressing fairness and ethical concerns; and (v) Enhancing replicability and transparency (Royall 1991). Among the most common types of clinical trials are those that randomize individual participants to treatment or control groups, such as cross-over designs, and those that randomize entire groups to treatment or control, such as cluster randomized trials (Piantadosi 2024). This approach enables a direct comparison of treatments by assessing how each participant responds to different interventions at multiple time points, rather than comparing separate groups of participants. Additionally, this methodology is frequently employed in pragmatic trials, particularly when there is prior evidence suggesting that an intervention will outperform the control, and the trial's objective is to confirm this advantage.

Stepped-wedge (SW) designs involve all groups of individuals eventually receiving the intervention, but the order in which they do so is randomized. This design is particularly valuable when it is unethical or impractical to withhold an intervention from some groups. It ensures that all participants will ultimately benefit from the intervention while maintaining randomization throughout the study. Although stepped-wedge (SW) designs are gaining popularity, they are often implemented by allocating participants uniformly to control-intervention sequences (Thompson et al. 2017; Lawrie et al. 2015; Zhan et al. 2018; Li et al. 2018). However, Moerbeek (2023a, b) highlight two key concerns: (i) this approach may not always yield the optimal solution, and (ii) it must account for participant attrition, which can be modeled as an attrition rate. Their studies emphasize the significance of these issues and suggest that there is considerable room for further investigation, particularly in the development of systematic numerical tools to facilitate the broader adoption of SW designs.

This paper seeks to bridge the knowledge gap identified by these studies by providing a theoretical foundation for the development of numerical tools designed to optimize participant allocation in SW randomized clinical trials. We approach the problem as a challenge in model-based optimal design of experiments and draw upon the extensive existing literature to support potential solutions. Computationally, we employ well-established deterministic optimization techniques, such as Semidefinite Programming (SDP), to efficiently solve the problem.

1.1 Novelty Statement and Organization

This paper presents several novel contributions:

(i) A formulation for the optimal allocation of participants in SW designs based on theoretical information criteria;
(ii) A numerical tool using SDP to automate the computational solution;
(iii) The application of the proposed computational approach to various scenarios, including: a. Varying numbers of sequences; b. Different attrition rates; c. Various optimality criteria; and d. Different models for the intercorrelation of observational errors.

The paper is structured as follows: Section 2 introduces the background and notation used to formulate the problem, covering the fundamentals of model-based optimal designs and Semidefinite Programming techniques used to solve these problems numerically. Section 3 presents the formulation for determining the optimal allocation of participants. The application of these algorithms is demonstrated in Section 4. Finally, Section 5 summarizes the formulations, reviews the results obtained and indicates topics for further exploration.

2 Notation and background

In our notation, boldface lowercase letters represent column vectors, while boldface uppercase letters denote continuous domains. Blackboard bold uppercase letters are used for discrete domains, and capital letters are reserved for matrices. Finite sets containing ι elements are compactly represented by $[\![\iota]\!] \equiv \{1, \ldots, \iota\}$. The transpose of a matrix or vector is denoted by "$^\intercal$," and the trace of a matrix is represented as $\mathrm{tr}(\bullet)$.

In §2.1, we introduce the model for representing a stepped-wedge randomized clinical trial. In §2.2, we present the fundamentals of model-based optimal design of experiments. Finally, in §2.3, we discuss the key concepts of Semidefinite Programming.

2.1 Individually Stepped-Wedge Randomized Clinical Trial

In this Section, we introduce the model used to represent the randomized stepped-wedge clinical trial, as described by Hussey and Hughes (2007). Specifically, we focus on the variant that incorporates an attrition rate due to participant drop-out, as proposed by Moerbeek (2023b).

The stepped-wedge design is a distinctive variant of the cross-over design, characterized by a unidirectional transition in which participants move exclusively from the control condition to the intervention condition. Figure 1 provides a

symbolic representation of a stepped-wedge design comprising 7 time instants (corresponding to 6 time intervals, which may vary in duration) and 5 sequences of participants. In this design, all sequences begin in the control condition. Transitions to the intervention condition occur sequentially at the end of each time interval. Specifically, the first sequence transitions at the end of the initial time interval, while subsequent sequences transition at the conclusion of their respective time intervals, each staggered by one slot relative to the preceding sequence.

Let $t_{\max}$ denote the total number of time intervals (including the initial time) in the trial, $t_{\max}+1$ the total of time instants including the initial time, and $s_{\max}$ represent the number of participant sequences, where $s_{\max} = t_{\max} - 1$. The set of time instants is defined as $\mathcal{T} = \{0, \ldots, t_{\max}\}$, and the set of sequences is denoted by $\mathcal{S} = [\![s_{\max}]\!]$. Each participant is measured at every time instant, with the response (i.e., the measurement) represented as $y_{i,s,t}$. Here, the subscript $i \in [\![n_s]\!]$ identifies the participant, $s \in \mathcal{S}$ specifies the sequence, $t \in \mathcal{T}$ refers to the time instant, and n_s the number of participants in s^{th} sequence.

The response model is expressed as:

$$y_{i,s,t} = \alpha + \beta_t + \gamma x_{s,t} + \epsilon_{i,s,t}, \tag{1}$$

where the model parameters are: α (baseline score), β_t (time effect, with $\beta_1 = 0$ as reference), γ (treatment effect), and $x_{s,t}$ (binary indicator: 0 for control, 1 for intervention). The observational error $\epsilon_{i,s,t}$ is normally distributed with mean 0 and variance σ^2_ϵ. Correlations between errors at time instants t and t' follow an exponential decay or a Toeplitz structure, $\text{corr}(\epsilon_{i,s,t}, \epsilon_{i,s,t'}) = \rho^{|t-t'|}$, where $\rho \in [0,1]$ is the correlation between consecutive errors (Kasza et al. 2019). Other correlation structures, such as compound symmetry or autoregressive, are also applicable (Morgan and Case 2013).

Each participant $i \in [\![n_s]\!]$ in the trial provides a set of $t_{\max}$ observations, collected at time instants $t \in [\![t_{\max}]\!]$, which can be represented in matrix form:

$$\mathbf{y}_{i,s} = X_s\,\boldsymbol{\theta} + \boldsymbol{\epsilon}_{i,s}, \quad i \in [\![n_s]\!],\ s \in \mathcal{S} \tag{2}$$

where $\mathbf{y}_{i,s} = \left(y_{i,s,1}\ y_{i,s,2} \cdots y_{i,s,t_{\max}}\right)^\intercal$ is a column vector with $t_{\max}$ rows,

$$X_s = \begin{pmatrix} 1\,0\,0\cdots 0 & x_{s,1} \\ 1\,1\,0\cdots 0 & x_{s,2} \\ 1\,0\,1\cdots 0 & x_{s,3} \\ \vdots\ \vdots\ \vdots\ \cdots\ \vdots & \vdots \\ 1\,0\,0\cdots 1 & x_{s,t_{\max}} \end{pmatrix}$$

a $t_{\max} \times (t_{\max}+1)$ matrix, $\boldsymbol{\theta} = \left(\alpha\ \beta_2 \cdots \beta_{t_{\max}}\ \gamma\right)^\intercal$ is the vector of $t_{\max}+1$ parameters to be estimated from the trial, including α, β_t for $t \in \{2, \cdots, t_{\max}\}$, and γ. Finally, $\boldsymbol{\epsilon}_{i,s} = \left(\epsilon_{i,s,1}\ \epsilon_{i,s,2} \cdots \epsilon_{i,s,t_{\max}}\right)^\intercal$ is the vector of observational errors. The variance-covariance matrix of this vector is given by:

$$V = \sigma^2_\epsilon \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{t_{\max}-1} & \rho^{t_{\max}} \\ \rho & 1 & \rho & \cdots & \rho^{t_{\max}-2} & \rho^{t_{\max}-1} \\ \rho^2 & \rho & 1 & \cdots & \rho^{t_{\max}-3} & \rho^{t_{\max}-2} \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ \rho^{t_{\max}} & \rho^{t_{\max}-1} & \rho^{t_{\max}-2} & \cdots & \rho & 1 \end{pmatrix}.$$

Since both the correlation ρ and the error variance σ^2_ϵ are assumed constant across individuals, the resulting variance-covariance matrix V is also constant across participants. The corresponding variance-covariance matrix for the estimate of $\boldsymbol{\theta}$, denoted $\hat{\boldsymbol{\theta}}$, is given by:

$$\text{cov}(\hat{\boldsymbol{\theta}}) = \left(\sum_{s=1}^{s_{\max}} n_s\ X_s^\intercal\ \hat{V}^{-1}\ X_s\right)^{-1}. \tag{3}$$

where $\hat{V}$ is the estimated variance-covariance matrix V.

We now incorporate the effects of attrition into the model by accounting for participants who drop out during the course of the trial. Specifically, we assume a constant attrition rate, denoted by r, across all time intervals. That is, at the end of each time interval $t \in [\![t_{\max}]\!]$, a proportion r of the participants enrolled in each sequence $s \in [\![s_{\max}]\!]$ drop out of the study. The number of participants in sequence s who drop out at time interval t is denoted by $n_{s,t}$, and the total number of participants dropping out in sequence s is given by $n_s = \sum_{t=1}^{t_{\max}} n_{s,t}$. This yields:

$$n_{s,t} = n_s\left[(1-r)^{t-1} - (1-r)^t\right] = n_s(1-r)^{t-1}r, \tag{4}$$

where the total number of participants across all sequences satisfies $\sum_{s=1}^{s_{\max}} n_s = N$, with N denoting the total number of participants in the trial.

We note that r represents the *expected* attrition rate, which may be elicited from historical or pilot data. While attrition is inherently a stochastic process and the actual number of dropouts at each time point is random, our model approximates this randomness by assuming that participant dropout follows its expected pattern over the course of the study. Galbraith and Marschner (2002) compared this approximation and the random process for parallel group longitudinal studies and found the two methods gave very similar designs.

As some participants drop out during the trial, the regressor matrices vary with time interval t. We denote the individual regressor matrices as $\mathcal{X}_{s,t}$, which include the first t rows of X_s. Similarly, the estimated variance-covariance matrix $\hat{\mathcal{V}}_t$ is a $t \times t$ matrix consisting of the first $t_{\max}$ rows and columns of V. The variance–covariance matrix of the

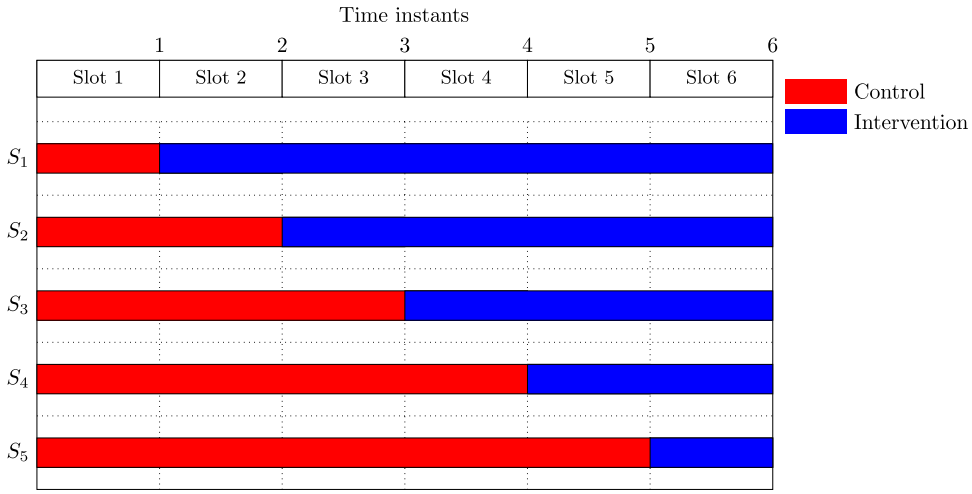

Fig. 1 Stepped-wedge trial schedule with 6 time intervals and 5 participant sequences

estimated regression coefficients is given by:

$$\text{cov}(\hat{\boldsymbol{\theta}}) = \left(\sum_{s=1}^{s_{\max}} \sum_{t=1}^{t_{\max}} n_{s,t}\, \mathcal{X}_{s,t}^{\mathsf{T}}\, \hat{\mathcal{V}}_t\, \mathcal{X}_{s,t} \right)^{-1} = \\ = N \cdot \left(\sum_{s=1}^{s_{\max}} \sum_{t=1}^{t_{\max}} w_{s,t}\, \mathcal{X}_{s,t}^{\mathsf{T}}\, \hat{\mathcal{V}}_t\, \mathcal{X}_{s,t} \right)^{-1}, \tag{5}$$

where $w_{s,t}$ represents the proportion of participants in sequence s at the t^{th} time interval. These weights satisfy $\sum_{t=1}^{t_{\max}} w_{s,t} = w_s$ and $\sum_{s=1}^{s_{\max}} w_s = 1$.

The inverse of the Fisher Information Matrix bounds the variance-covariance matrix of parameter estimates (Rao 1945). The global FIM, $\mathcal{M}(\hat{\boldsymbol{\theta}})$, is

$$\mathcal{M}(\hat{\boldsymbol{\theta}}) = N^{-1} \sum_{s=1}^{s_{\max}} \sum_{t=1}^{t_{\max}} w_{s,t} \mathcal{X}_{s,t}^{\mathsf{T}} \hat{\mathcal{V}}_t \mathcal{X}_{s,t}, \tag{6}$$

where $w_{s,t}$ is the proportion of participants in sequence s at time $t_{\max}$. The local FIMs for participants in sequence s and time interval $t_{\max}$ are

$$M_{s,t} = \mathcal{X}_{s,t}^{\mathsf{T}} \hat{\mathcal{V}}_t \mathcal{X}_{s,t}.$$

The equivalence between our formulation, which minimizes a convex function of the global FIM, and that of Moerbeek (2023b), which minimizes the $(t_{\max}+1, t_{\max}+1)$-th element of the variance-covariance matrix that is the variance of $\hat{\gamma}$ (obtained by inverting the FIM), is shown in Sections 1-2 of the Supplementary Material. This result allows us to express the optimal allocation problem as the maximization of a convex function of the FIM, which is a state-of-the-art approach in Model-based Optimal Design of Experiments. Such problems can be reformulated as Semidefinite Programming (SDP) or Second Order Cone Programming (SOCP) problems, ensuring global optimality and polynomial-time computational complexity. In contrast, minimizing the $(t_{\max}+1, t_{\max}+1)$-th element of the variance-covariance matrix requires Nonlinear Programming (NLP) (see Boer and Hendrix (2000); Duarte et al. (2016); Gribik and Kortanek (1977)), which does not guarantee global optimality unless global optimization solvers are used, and these solvers are NP-hard.

2.2 Optimal design of experiments

In this Section, we present the methodology to optimize participant allocation in stepped-wedge randomized clinical trials with attrition, as introduced in §2.1. The goal is to determine the optimal allocation proportions (weights) w_s for each sequence and $w_{s,t}$ for each time interval, subject to the constraint

$$w_{s,t} = w_s (1-r)^{t-1} r, \quad s \in [\![s_{\max}]\!],\ t \in [\![t_{\max}]\!], \tag{7}$$

obtained by scaling Eq. (4).

In stepped-wedge designs, the optimization criterion is often the variance of the treatment effect estimator. Here, we connect this criterion to information-theoretic approaches that minimize the parametric confidence region of the estimators. Specifically, we propose using alphabetic information-based criteria derived from the FIM as alternatives. These criteria address multiple objectives, including efficient parameter estimation, robust model predictions, and computational scalability. Moreover, for example, D– and A–optimality often enhance the precision of the treatment effect estimator, maintaining relevance to clinical and statistical decision-making. For cases where the treatment effect is the primary focus, constraints can be incorporated to emphasize its variance in the optimization.

Continuous designs, also referred to as approximate designs, represent experimental setups in the theoretical limit

where the number of observations N approaches infinity ($N \to \infty$). In this framework, weights are distributed over $[0, 1]$, representing the proportion of total observations allocated to each sequence $s \in [\![s_{\max}]\!]\}$. The primary motivations for studying continuous designs for stepped-wedge trials include: (i) the optimization problem for finding approximate designs is convex (or can be reformulated as such) when the design criterion is a convex function of the Fisher Information Matrix (FIM). This ensures global optimality, as guaranteed by equivalence theorems (Kiefer 1974; Pukelsheim 1993), and enables the use of tailored optimization algorithms (Vandenberghe and Boyd 1999); (ii) continuous designs offer valuable theoretical insights into the structure of optimal designs and serve as benchmarks for evaluating exact designs; (iii) they are more scalable for problems involving a large number of design variables or sequences; (iv) they naturally extend to sequential or adaptive design frameworks by incorporating incoming information; and (v) they can often be discretized by constructing exact designs from the continuous solution through weight rounding. For more details on discretization methods, see Pukelsheim and Rieder (1992); Duarte et al. (2020).

To systematically find continuous optimal designs of experiments, recent advancements have introduced several convex optimization algorithms (Pronzato and Pázman 2013; Harman et al. 2020, Chap. 9). In this study, we leverage the finite number of variables $w_{s,t}$ and the independence of the Fisher Information Matrix (FIM) from the model parameters, which results from the linearity of the response model. This enables us to reformulate the problem as a SDP problem, where the optimization variables are the weights (proportions) of individuals allocated to each sequence. Specifically, we consider approximate designs represented by $s_{\max}$-point tuples as follows:

$$\xi = \begin{pmatrix} 1 & 2 & \cdots & s_{\max}-1 & s_{\max} \\ w_1 & w_2 & \cdots & w_{s_{\max}-1} & w_{s_{\max}} \end{pmatrix}.$$

The information provided by an experimental design is captured by its Fisher Information Matrix. Since the FIM depends on the design ξ, we rescale and reformulate Eq. (6) to explicitly depend on ξ:

$$\mathcal{M}(\xi) = \sum_{s=1}^{s_{\max}} \sum_{t=1}^{t_{\max}} w_{s,t} \mathcal{X}_{s,t}^{\mathsf{T}} \hat{\mathcal{V}}_t \mathcal{X}_{s,t}. \tag{8}$$

This matrix is $(t_{\max}+1) \times (t_{\max}+1)$ and is semidefinite positive by construction similarly as the local FIMs $M_{s,t}$.

Since $\hat{\boldsymbol{\theta}}$ is asymptotically multivariate normally distributed, the volume of the asymptotic confidence region for $\boldsymbol{\theta}$ is inversely proportional to the square root of $\det[\mathcal{M}(\xi)]$. For simplicity, we assume that $\mathcal{M}(\xi)$ is independent of $\hat{\boldsymbol{\theta}}$, as the model in Eq. (1) is linear with respect to the parameters, and we do not consider r and ρ as additional parameters to estimate from experimental data. Consequently, maximizing the determinant of the FIM minimizes the volume of this confidence region. Various design criteria seek to optimize the FIM in different ways, often formulated as convex (or reformulable as convex) functions of the FIM.

To formalize the problem in our context, we consider the general optimal design criterion within the Kiefer (1975) framework:

$$\xi_\ell = \arg\max_{\xi \in \Xi} \Phi_\ell[\mathcal{M}(\xi)] = \left\{ \frac{1}{p} \operatorname{tr}\left[\mathcal{M}^{-\ell}(\xi) \right] \right\}^{1/\ell}, \tag{9}$$

where $\ell < 0$ is a parameter that determines the specific form of the criterion by controlling the power to which the inverse FIM is raised. For example, setting $\ell = -1$ yields A–optimality (minimizing the average variance of parameter estimates), the limit $\ell \to 0$ corresponds to D–optimality (maximizing the determinant of $\mathcal{M}(\xi)$), and the limit $\ell \to -\infty$ gives E–optimality (maximizing the smallest eigenvalue of $\mathcal{M}(\xi)$). Here, p denotes the dimension of the FIM, and Φ_ℓ represents the associated ℓ-optimality criterion.

From the general formulation in (9), the D–optimal and A–optimal designs are special cases that can be derived by selecting appropriate values for ℓ. These designs are defined as follows:

$$\xi_{\mathrm{D}} = \arg\max_{\xi \in \Xi} \{\log\det[\mathcal{M}(\xi)]\}, \tag{10a}$$

$$\xi_{\mathrm{A}} = \arg\min_{\xi \in \Xi} \left\{ \operatorname{tr}[\mathcal{M}(\xi)^{-1}] \right\}, \tag{10b}$$

where n_θ denotes the number of parameters, corresponding to p in (9). The D–optimal and A–optimal designs are obtained by solving the respective optimization problems in (10), subject to the constraints outlined in (7). Further, it is important to note that the optimization is performed over the design measure ξ, which is characterized solely by the weight vector $\mathbf{w} \in \Xi$ (see the definition of ξ above). The feasible design space is given by:

$$\Xi = \left\{ \mathbf{w} \mid \mathbf{1}_{s_{\max}}^{\mathsf{T}} \mathbf{w} = 1,\ \mathbf{w} \in [0,1]^{s_{\max}} \right\},$$

where $\mathbf{1}_{s_{\max}}$ represents a column vector of ones of size $s_{\max}$.

Following Eq. (10), the efficiency of a given design ξ relative to a reference design ξ^*, characterized by the Fisher Information Matrix $\mathcal{M}^*(\xi^*)$ on a per-observation basis, is defined as:

$$\eta_{\mathrm{D}} = \left\{ \frac{\det[\mathcal{M}(\xi)]}{\det[\mathcal{M}^*(\xi^*)]} \right\}^{1/n_\theta}, \tag{11a}$$

$$\eta_{\mathrm{A}} = \left\{ \frac{\operatorname{tr}[\mathcal{M}^*(\xi^*)^{-1}]}{\operatorname{tr}[\mathcal{M}(\xi)^{-1}]} \right\}^{1/n_\theta}. \tag{11b}$$

Here, η denotes efficiency, and the subscript indicates the optimality criterion under consideration. The reference design ξ^* is the unconstrained optimal design for the respective criterion (e.g., D- or A-optimality). Efficiency thus provides a relative measure of the potential information loss per observation when the design ξ is constrained—such as when satisfying multiple criteria simultaneously. Although uniform designs (which assign equal weights to each sequence) are included in our numerical studies as intuitive and widely used baselines, they are not used as reference designs in the efficiency definitions given in Eq. (11).

Cook and Wong (1994) and Clyde and Chaloner (1996) introduced compound optimal designs, formulating them as constrained optimization problems to balance multiple design objectives. Common approaches include: (i) min-max formulations, which can be challenging when combining criteria of different classes (e.g., convex and concave); (ii) weighted averaging of criteria; and (iii) multiple-objective reformulations, such as the ϵ-constraint method (Miettinen 1999).

In this study, we adopt the ϵ-constraint method to handle multiple objectives. Specifically, we maximize one design criterion, Φ_ℓ, while requiring that another criterion, Φ_ν, satisfies a minimum efficiency threshold, τ, as in Cook and Wong (1994); Huang and Wong (1998). This leads to the following constrained optimization problem:

$$\xi_\ell = \arg\max_{\xi \in \Xi} \Phi_\ell[\mathcal{M}(\xi)], \tag{12a}$$

$$\text{s.t.} \quad \Phi_\nu[\mathcal{M}(\xi)] \geq \tau, \quad \nu \neq \ell, \tag{12b}$$

where τ specifies the minimum acceptable efficiency for the secondary criterion. The choice of τ influences the trade-off between objectives: a higher value ensures stronger performance on the secondary criterion, possibly at the expense of the primary one. In this work, we adopt a stringent threshold of $\tau = 0.98$ to guarantee that the resulting compound designs are both feasible and practically relevant, effectively balancing the competing objectives.

2.3 Semidefinite Programming

In this Section, we present the key principles of convex optimization methods applied to optimal design of experiments, with a particular focus on scenarios where the discretized design domain consists of $n_c = s_{\max} \times t_{\max}$ candidate experimental points.

Let $\mathbb{S}_+^{n_\theta}$ denote 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 said to be semidefinite representable (SDr) if for any $\boldsymbol{\zeta} \in \mathbf{S}$, the projection of $\boldsymbol{\zeta}$ onto a higher-dimensional set $\mathbf{S}^{\text{exp}}$, denoted by $\text{proj}_{\mathbf{S}^{\text{exp}}}(\boldsymbol{\zeta})$, can be described by Linear Matrix Inequalities (LMIs).

A convex (or concave) function $\varphi : \mathbb{R}^{m_1} \to \mathbb{R}$ is SDr if and only if its epigraph $\{(t, \boldsymbol{\zeta}) : \varphi(\boldsymbol{\zeta}) \leq t\}$ or hypograph $\{(t, \boldsymbol{\zeta}) : \varphi(\boldsymbol{\zeta}) \geq t\}$ can be represented by LMIs (Ben-Tal and Nemirovski 2001; Boyd and Vandenberghe 2004). The optimal values of SDr functions can be formulated as semidefinite programs (SDPs) of the form:

$$\max_{\boldsymbol{\zeta}} \left\{ \mathbf{d}^\intercal \boldsymbol{\zeta}, \sum_{i=1}^{m_1} \zeta_i M_i - M_0 \succeq 0 \right\}. \tag{13}$$

where $\mathbf{d}$ is a vector of constants specific to the design problem, and M_i are local Fisher Information Matrices, with the constraint $M_0 \succeq 0$ ensuring the solution is feasible.

In this formulation, the decision variables in the vector $\boldsymbol{\zeta}$ include the weights w_i for the design points, as well as auxiliary variables. The optimization problem seeks to find the optimal design for a given set of candidate sequences, subject to the following constraints on the weights: (i) $\mathbf{w} \geq 0$, and (ii) $\mathbf{1}_{n_c}^\intercal \mathbf{w} = 1$, where $\mathbf{1}_{n_c}$ is the unit column vector of length n_c. The problem, as represented in equation (13), is a standard SDP that incorporates LMIs as conic constraints. Ben-Tal and Nemirovski (2001) provide a list of SDr functions that are useful for solving continuous optimal design problems using SDP formulations (see Boyd and Vandenberghe (2004, §7.3)). Sagnol (2013) demonstrated that every criterion in the Kiefer class of optimality criteria is SDr for all rational values of $\omega \in (-\infty, 0]$. General SDP formulations for these criteria exist, with ω being the coefficient in the general Kiefer class Ψ_ω (Kiefer 1974). In particular, A–optimality corresponds to $\omega = -1$, E–optimality to $\omega \to -\infty$, and D–optimality to $\omega \to 0$. The problem of finding optimal approximate experimental designs for these criteria can be formulated as an SDP problem, as discussed in Vandenberghe and Boyd (1999) and Duarte and Wong (2015), among others.

3 Formulations for finding optimal designs of experiments

In this Section, we present the proposed formulations for identifying alphabetic optimal experimental designs to allocate participants across sequences.

We base our approach on the regression model (1). The design space comprises $n_c = s_{\max} \times t_{\max}$ instances, distributed across $s_{\max}$ sequences. The *local* Fisher Information Matrices (FIMs) at each instance are defined as

$$M_{s,t} = \mathcal{X}_{s,t}^\intercal \hat{\mathcal{V}}_t \mathcal{X}_{s,t}, \tag{14}$$

while the *global* FIM is given by Eq. (8). The weights $w_{s,t}$ are subject to the constraints outlined in Eqs. (6-7).

We consider the optimal design problems formulated in Eq. (10). The D–optimality criterion is given by:

$$\text{Opt} \equiv \max_{\mathbf{w}} \quad \log\det[\mathcal{M}(\xi)] \tag{15a}$$

$$\text{s.t.} \quad \mathcal{M}(\xi) = \sum_{s=1}^{s_{\max}} \sum_{t=1}^{t_{\max}} w_{s,t} M_{s,t} \succeq 0 \tag{15b}$$

$$M_{s,t} = \mathcal{X}_{s,t}^{\intercal} \hat{\mathcal{V}}_t \mathcal{X}_{s,t}, \quad s \in [\![s_{\max}]\!],\ t \in [\![t_{\max}]\!] \tag{15c}$$

$$w_{s,t} = w_s (1-r)^{t-1} r, \quad s \in [\![s_{\max}]\!],\ t \in [\![t_{\max}]\!] \tag{15d}$$

$$\sum_{s=1}^{s_{\max}} w_s = 1 \tag{15e}$$

$$0 \leq w_{s,t}, w_s \leq 1 \quad s \in [\![s_{\max}]\!],\ t \in [\![t_{\max}]\!]. \tag{15f}$$

Here, Opt represents the optimum value of the problem, and Eq. (15a) defines the objective function. The log-determinant function is a semidefinite representable (SDr) function (see MOSEK (2024a)) and fits within the general reformulation in Eq. (13). The constraints are as follows: Eq. (15b) ensures the global Fisher Information Matrix is positive semidefinite, Eq. (15c) specifies the construction of local FIMs as in §2.1, Eq. (15d) enforces equality constraints within the same sequence, Eq. (15e) ensures the weights sum to one, and Eq. (15f) restricts all weights to [0, 1].

The formulation for computing A–optimal designs follows from Eq. (10b) and is given by:

$$\text{Opt} \equiv \min_{\mathbf{w}} \quad \text{tr}[\mathcal{M}(\xi)^{-1}] \tag{16a}$$

$$\text{s.t.} \quad \text{Eqs. (15b)-(15f).} \tag{16b}$$

We note that the D–optimality problem involves minimizing $\log\det[\mathcal{M}(\xi)^{-1}]$ (as described in the general problem in Eq. (9)). This can be reformulated as maximizing $\log\det[\mathcal{M}(\xi)]$ by exploiting the concavity of the $\log\det(\bullet)$ function for positive semidefinite matrices. On the other hand, the trace of the inverse of a semidefinite positive matrix is a convex function, which explains why the A–optimal design problem is solved as a minimization problem (MOSEK 2024a).

For systematization, multiple-objective designs are denoted by two joint letters: the first represents the optimality criterion in the optimization problem, and the second corresponds to the constrained criterion. The formulations for multiple-objective designs are derived from the general framework outlined in Eq. (12), using the ϵ-constraint method.

For example, in the case of DA-optimal designs, where the A-optimal efficiency is constrained, the optimization problem can be expressed as:

$$\text{Opt} \equiv \max_{\xi} \quad \log\det[\mathcal{M}(\xi)], \tag{17a}$$

$$\text{s.t.} \quad \tau_{\text{A}} \cdot \text{tr}[\mathcal{M}^{-1}(\xi)] \leq \text{tr}[\mathcal{M}^{*}(\xi^{*})^{-1}], \tag{17b}$$

$$\text{Eqs. (15b)-(15f),} \tag{17c}$$

where the objective function is defined in Eq. (17a), and the constraint in Eq. (17b) ensures that the A–optimal efficiency meets the target $\tau_{\text{A}} \in [0, 1]$, as derived from Eq. (11b). T he additional constraints in Eqs. (15b)–(15f) impose feasibility conditions for the design.

Similarly, for AD–optimal designs, where the D–optimal efficiency is constrained, the optimization problem becomes:

$$\text{Opt} \equiv \min_{\xi} \quad \text{tr}[\mathcal{M}(\xi)^{-1}], \tag{18a}$$

$$\text{s.t.} \quad \log\det[\mathcal{M}(\xi)] \geq \log\det[\mathcal{M}^{*}(\xi^{*})] + \log(\tau_{\text{D}}), \tag{18b}$$

$$\text{Eqs. (15b)-(15f),} \tag{18c}$$

where the objective function in Eq. (18a) maximizes the trace-based criterion, and the constraint in Eq. (18b) ensures that the D–optimal efficiency surpasses the target $\tau_{\text{D}} \in [0, 1]$.

The numerical solution of multiple-objective designs within a $\epsilon-$constraint method design framework was first introduced by Zhu and Wong (2000). More recently, Wong and Zhou (2023); Duarte et al. (2024) systematically tackled this problem using semidefinite programming (SDP).

We solved the semidefinite programming (SDP) problems using the `CVXPY` environment (Diamond and Boyd 2016) with the `Mosek` solver (MOSEK 2024b), which employs an efficient Interior Point algorithm (Karmarkar 1984; Ye 1997). The relative and absolute tolerances were set to 1×10^{-5}. All computations in §4 were performed on an Intel Core i7 machine with a 3.80 GHz processor running a 64-bit Windows 10 operating system.

4 Results

This Section presents results for the stepped-wedge randomized allocation problem modeled in §3. We analyze how key factors influence optimal designs by solving problems (15)-(16). Specifically, §4.1 examines the impact of the number of sequences (S), §4.2 explores the effect of the attrition rate, §4.3 compares D– and A–optimality criteria, §4.4 evaluates the influence of different correlation models for observational error, and §4.5 considers multiple objective designs.

The baseline analysis considers $t_{\max} = 7$ time periods, $s_{\max} = 6$ sequences, and an attrition rate of $r = 0.05$, assuming the D–optimality criterion and an exponential decay correlation structure for observational error. Variations in

these parameters are explored, with correlation values $\rho \in \boldsymbol{\varrho}$, where

$$\boldsymbol{\varrho} = \begin{pmatrix} 0.0 & 0.1 & 0.2 & 0.3 & 0.4 & 0.5 & 0.6 & 0.7 & 0.8 & 0.9 & 0.99 \end{pmatrix}^{\mathsf{T}},$$

ranging from no correlation ($\rho = 0.0$) to near-perfect correlation ($\rho = 0.99$). Results for other $t_{\max}$ values (and corresponding $s_{\max} = t_{\max} - 1$) are in the Supplementary Material (SM).

All results presented in the following sections were obtained with modest computational effort, with each run for a specific $t_{\max}$ and ρ requiring approximately 0.6 s of CPU time.

4.1 Impact of the number of sequences

The impact of varying the number of sequences of participants enrolled in the clinical trial is illustrated in Figure 2. The figures display the optimal values of w_s, $s \in [\![s_{\max}]\!]$ for each value of $\rho \in \boldsymbol{\varrho}$. Specifically, Figure 2(a) shows the results for $s_{\max} = 6$, while Figure 2(b) presents the results for $s_{\max} = 5$.

The results in Figure 2 for $r = 0.05$ reveal several key structures:

(i) When the observational error is uncorrelated ($\rho = 0.0$), only two sequences receive nonzero weights. These are the extreme sequences (S_1 and S_6) in the left-hand panel $S = 6$ and (S_1 and S_5) in the right-hand panel $S = 5$.
(ii) For highly correlated observational error ($\rho = 0.99$), the allocation becomes more uniform, with all sequences receiving nearly equal weights. This pattern aligns with the D–optimality criterion (Pukelsheim 1993) for first-order models.
(iii) For intermediate values of ρ, central sequences tend to be assigned lower weights compared to those at the extremes. In the left-hand panel for intermediate ρ the highest weights are, in order, for (S_1 and S_6), the next highest are for (S_2 and S_5) with the lowest for (S_3 and S_4). For high values of ρ the trajectories cross. In the right-hand panel the lowest weight is initially for the central sequence S_3.

These results are tabulated for more values of S in Supplementary Material (Table 1).

Table 1 presents the D-optimal efficiencies for $s_{\max} = 6$, $\rho \in \boldsymbol{\varrho}$, and $r \in (0.00, 0.05, 0.20)$, computed using Eq. (11a). The reference designs correspond to those obtained using our proposed formulation, while the designs ξ in Eq. (11a) represent uniform allocations, as assumed, for example, by Thompson et al. (2017). Our results indicate that the proposed designs are more efficient than uniform allocations. In fact, the efficiency of uniform designs is smaller than that of the optimal designs, highlighting the importance of using optimal design strategies especially for higher attrition rates.

4.2 Impact of the attrition rate

The results in Figure 2 assumed a low value of the attrition rate. Figure 3 compares the results of D–optimal designs for $r = 0.00$ and $r = 0.20$, obtained with $s_{\max} = 6$. In Figure 3(a) $r = 0.00$, while in Figure 3(b) $r = 0.20$. These results are also comparable to Figure 2(a), which shows the case for $r = 0.05$. The plot for $r = 0.00$ is symmetrical in the sequences: S_1 and S_6, S_2 and S_5 and S_3 and S_4. For $\rho = 0$ the weights for S_1 and S_6 are both 0.5. In the absence of attrition and correlation, only the two most extreme sequences are included in the design. On the other hand, when $r = 0.20$, increasing ρ requires designs with six distinct weights. At the right-hand side of Figure 3(b) the weights decrease from 0.3171 for S_1 to zero for S_6. The high attrition rate causes the later interventions to provide relatively uninformative readings as the sample sizes become small. Careful inspection of the left-hand panel of Figure 2(a) when $\rho = 1$, shows that, even for $r = 0.00$, the sequence weights are ordered from w_1 to w_6, just as they are in Figure 3(b). The attrition rate has a strong effect on the optimal designs as a function of ρ. Numerical details of the optimal designs are given in Tables 2 and 3 of the Supplementary Material.

4.3 Impact of the optimality criterion

Figure 4 presents the results for the A–optimality criterion, considering $s_{\max} = 6$, with $r = 0.05$ (see Figure 4(a)) and $r = 0.20$ (see Figure 4(b)). Several key findings can be highlighted:

(i) As expected, the designs obtained using the A-optimality criterion differ significantly from those based on the D–optimality criterion. This distinction is evident when comparing Figures 2(a) and 4(a) (both for $r = 0.05$) as well as Figures 3(b) and 4(b) (both for $r = 0.20$).
(ii) When the observational error is uncorrelated, participants are assigned to sequence S_6 with weight 0.6516 and to S_1 with weight 0.3484. As ρ increases, S_6 continues to have the highest weight; for $\rho = 0.99$, $w_6 = 0.4160$. The weights on the other sequences decrease in order, with $w_1 = 0.0449$. The structure of Figure 4(b) with $r = 0.20$ is similar. For zero correlation the weights are the same as those in Figure 4(b). When $\rho = 0.99$, w_6 is slightly reduced to 0.3802 and $w_1 = 0.1651$, close to the value for w_5. In order, the remaining weights are for w_2, w_4 and w_3. The main effect of higher drop-out is to increase the weight for S_1 for higher values of ρ.
There is a clear difference in structure between the D–

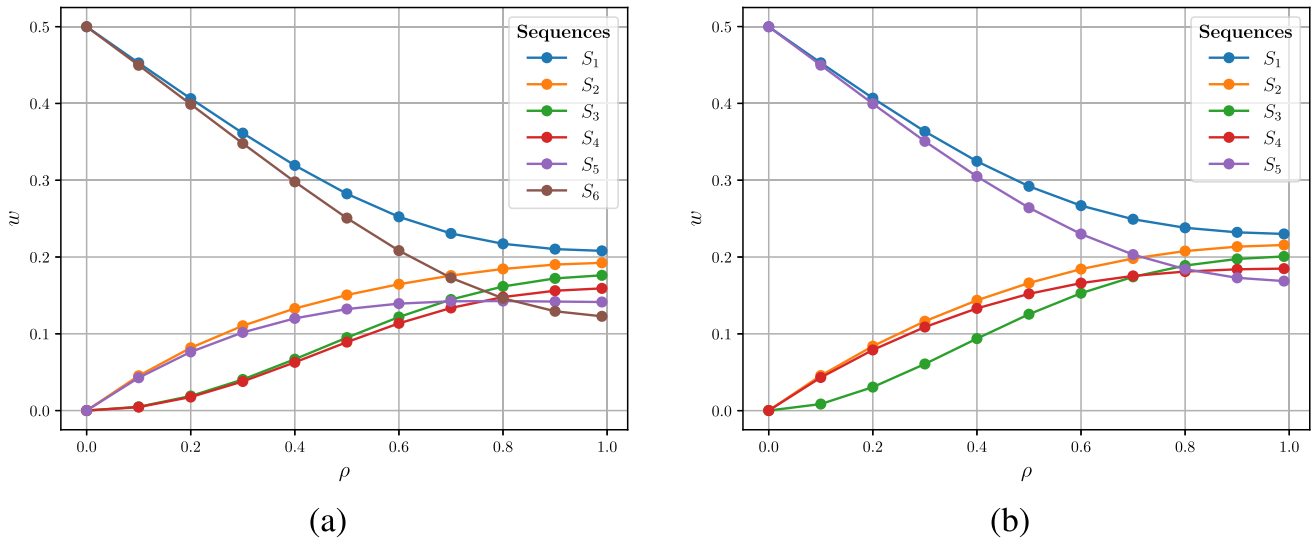

Fig. 2 Comparison of optimal allocation for $\boldsymbol{\varrho}$ considering: (a) $s_{\max} = 6$ sequences; (b) $s_{\max} = 5$ sequences. Setup: D–optimality criterion, $r = 0.05$, and exponential decay correlation structure

Table 1 D–optimal efficiency of designs obtained for $s_{\max} = 6$, $\rho \in \boldsymbol{\varrho}$, and $r \in (0.00, 0.05, 0.20)$. The reference designs, ξ^*, are those obtained using our proposed formulation, while ξ corresponds to uniform designs

r	ρ										
	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.99
0.00	0.7778	0.8158	0.8575	0.8999	0.9383	0.9680	0.9868	0.9961	0.9993	1.0000	1.0000
0.05	0.7774	0.8142	0.8547	0.8958	0.9333	0.9628	0.9817	0.9910	0.994	0.9947	0.9938
0.20	0.7702	0.8024	0.8366	0.8706	0.9007	0.9227	0.9335	0.9332	0.9246	0.9131	0.9058

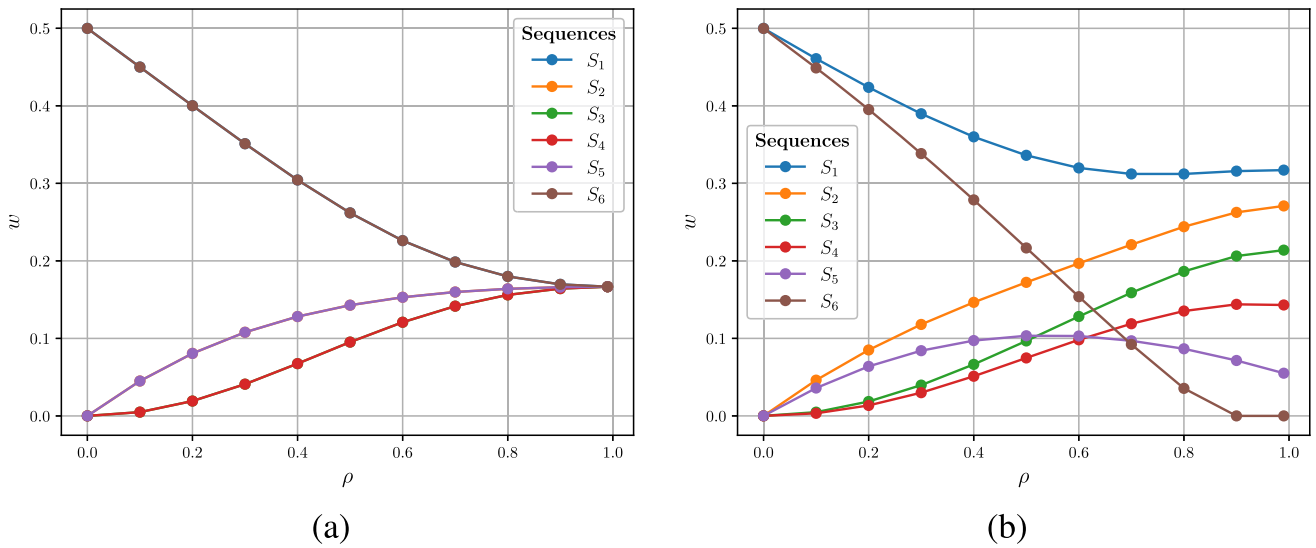

Fig. 3 Comparison of optimal allocation for $\boldsymbol{\varrho}$ considering: (a) $r = 0.00$; (b) $r = 0.20$. Setup: D–optimality criterion, $s_{\max} = 6$, and exponential decay correlation structure. The design for $r = 0.00$ is symmetrical in the sequences: S_1 and S_6, S_2 and S_5 and S_3 and S_4

optimal designs of Figure 3 and the A–optimal designs of Figure 4. For all values of ρ, the A-optimal designs put highest weight on S_6, in order to reduce the variances of the parameter estimates associated with the later stages of the design. The D–optimal criterion is more focused on the overall properties of the parameter estimates.

Further results are presented in Tables 4, 5, and 6 of the Supplementary Material. We note that Moerbeek (2023b) do not address the construction of A–optimal designs, and that $\text{tr}[\mathcal{M}(\xi)]$ does not correspond to the $(T+1, T+1)$ element of the variance-covariance matrix. Thus, to the best of our knowledge, these designs are novel.

Table 2 presents the A-optimal efficiency values computed using Eq. (11b) for $s_{\max} = 6$, $\rho \in \boldsymbol{\varrho}$, and $r \in (0.00, 0.05, 0.20)$. Similar to the D-optimal case, the designs obtained with our formulation outperform uniform designs in terms of efficiency. However, in this case, we observe a wider range of variation in efficiency across different parameter settings. Additionally, inefficiency appears to increase with the attrition rate r and decrease with ρ. This decrease, however, is bounded by a threshold greater than $\rho = 0.7$, the exact value depending on r.

4.4 Impact of the correlation structure between the observational error

Now, we modify the observational error correlation structure to compare it with the exponential decay function. For illustration, we adopt the Matérn correlation:

$$V_{i,j} = \begin{cases} 1.0, & \text{for } i = j, \\ C_{\nu,\phi}(|i-j|), & \text{for } |i-j| \geq 1, \end{cases} \tag{19a}$$

$$C_{\nu,\phi}(|i-j|) = \frac{1}{\Gamma(2\nu)2^{\nu-1}} \left(\frac{\sqrt{2\nu}|i-j|}{\phi} \right)^{\nu} K_{\nu} \left(\frac{\sqrt{2\nu}|i-j|}{\phi} \right), \tag{19b}$$

where $C_{\nu,\phi}(|i-j|)$ denotes the Matérn correlation function with smoothness parameter $\nu > 0$ and length-scale $\phi > 0$, $\Gamma(\bullet)$ is the Gamma function, and $K_{\nu}(\bullet)$ is the modified Bessel function of the second kind. In this context, i and j represent the row and column indices of the correlation matrix V, respectively. We define $\nu = a\ (\rho + b)$ and $\phi = c\ (\rho + b)$, where a, b, and c are constants, with $b = 0.01$, $c = 4.0$, and $a \in \{0.5, 1.5\}$; and $\rho \in \boldsymbol{\varrho}$ as described in the previous sections.

Figure 5 illustrates the results obtained using the Matérn correlation function. Specifically, Figure 5(a) corresponds to $a = 0.5$, while Figure 5(b) corresponds to $a = 1.5$. The allocation weight profiles differ from those obtained with the exponential decay function shown in Figure 2(a), yet they show similar weight profiles, particularly for $a = 0.5$. If a serious comparison of the effect of various ρ functions is required, the values of the parameters in the Matérn function should be chosen to take the two correlation functions as close together as possible at the values of ρ that are of interest.

Additional results can be found in Tables 7 and 8 of the Supplementary Material.

4.5 Multiple-objective designs

Finally, we consider multi-objective designs, specifically DA-optimal and AD–optimal designs, whose SDP formulations were introduced in §3 (see problems (17) and (18), respectively). For both designs, we impose efficiency targets of at least 97% for the constrained criterion, setting τ_{A} and τ_{D} to 0.97. In other words, the ϵ in the ϵ-constraint method is set to 0.97.

Figure 6 compares the results for DA–optimal designs (Figure 6(a)) and AD–optimal designs (Figure 6(b)). Figure 2(a) shows the D–optimal design for the same value of r. The general structure of this figure is similar to that of Figure 6(a). Likewise, the A-optimal design, Figure 4(a) has a similar structure of weights to the compound design of Figure 6(b). The implication is that the requirement of 97% efficiency for the fixed part of the design has surprisingly little effect on the design.

Additional results from the calculation of compound designs are provided in Tables 9 and 10 of the Supplementary Material.

These numerical results demonstrate that the proposed formulation can handle a wide range of problems where the following parameters are modifiable: (i) the number of sequences; (ii) the attrition rate; (iii) the optimality criterion; (iv) the function describing the observational error correlation; and (v) multiple-objective designs.

5 Conclusions

We investigated the problem of determining the optimal allocation of individuals in stepped-wedge clinical trials. Our work extends the framework established by Moerbeek (2023a, b), which incorporates individual drop-out modeled through a constant attrition rate. The proposed approach connects the most commonly used optimization criterion for this problem—the minimization of the variance of the treatment effect estimator—with information-theoretic criteria, specifically those minimizing a convex function of the inverse of the Fisher Information Matrix.

Building on this relationship, we reformulate the original problem using standard alphabetic-based criteria for contin-

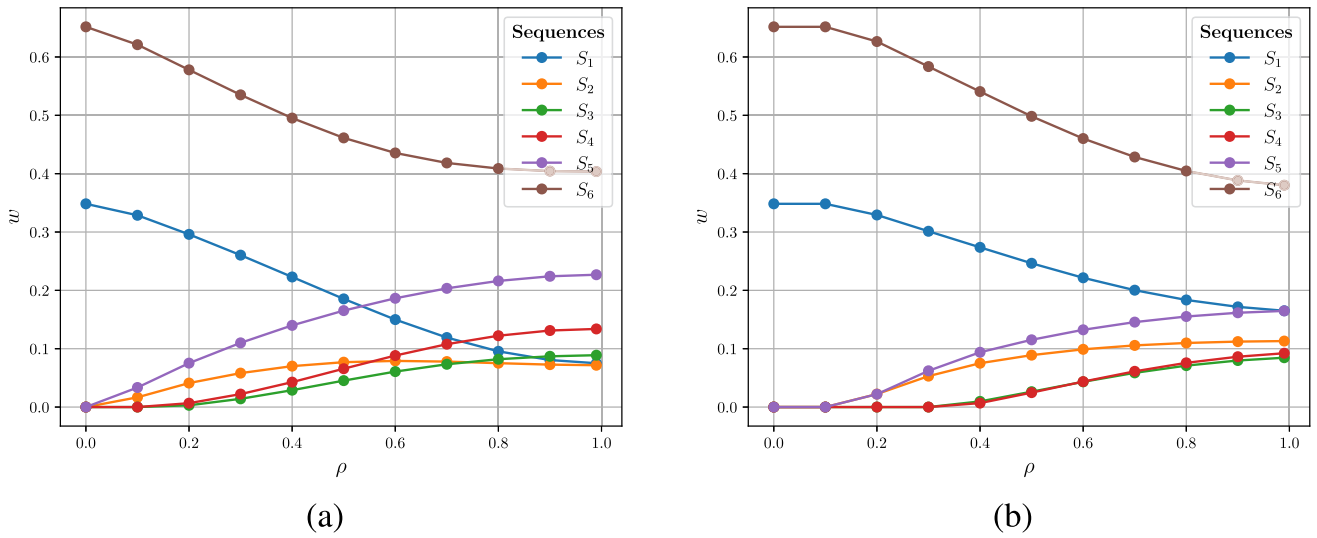

Fig. 4 Comparison of optimal allocation for $\boldsymbol{\varrho}$ considering: (a) A–optimality criterion and $r = 0.05$; (b) A–optimality criterion and $r = 0.20$. Setup: $s_{\max} = 6$, and exponential decay correlation structure

Table 2 A-optimal efficiency of designs obtained for $s_{\max} = 6$, $\rho \in \boldsymbol{\varrho}$, and $r \in (0.00, 0.05, 0.20)$. The reference designs, ξ^*, are those obtained using our proposed formulation, while ξ corresponds to uniform designs

r	ρ										
	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.99
0	0.7149	0.7063	0.712	0.735	0.7811	0.8615	0.9985	0.8039	0.5704	0.304	0.03556
0.05	0.6053	0.5981	0.6031	0.6225	0.6614	0.7289	0.8439	0.9525	0.6769	0.361	0.04194
0.2	0.3453	0.3416	0.3448	0.3562	0.3785	0.4169	0.4818	0.5979	0.8388	0.6368	0.07228

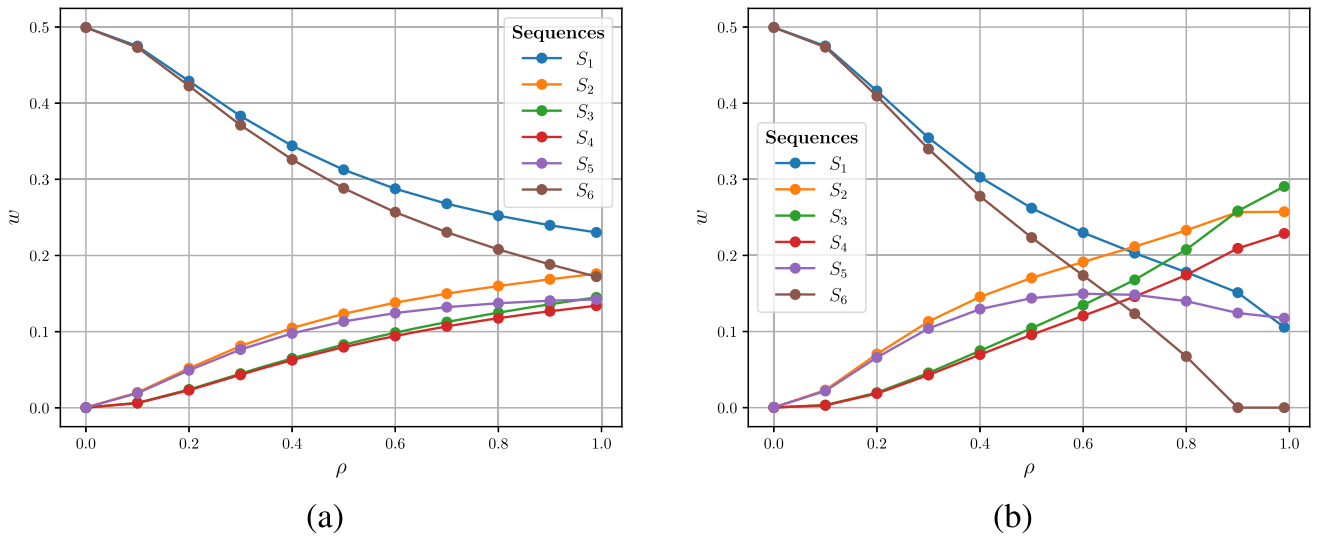

Fig. 5 Comparison of optimal allocation for $\boldsymbol{\varrho}$ considering: (a) Matérn correlation function with $a = 0.5$; (b) Matérn correlation function with $a = 1.5$. Setup: D–optimality criterion, $s_{\max} = 6$, $r = 0.05$, $\nu = a\ (\rho + b)$, $\phi = c\ (\rho + b)$, $b = 0.01$, and $c = 4.0$

uous experimental designs. Specifically, we propose D- and A-optimal design formulations to address the challenge. The numerical tools we introduce are grounded in Semidefinite Programming, enabling the systematic and efficient resolution of optimal design problems.

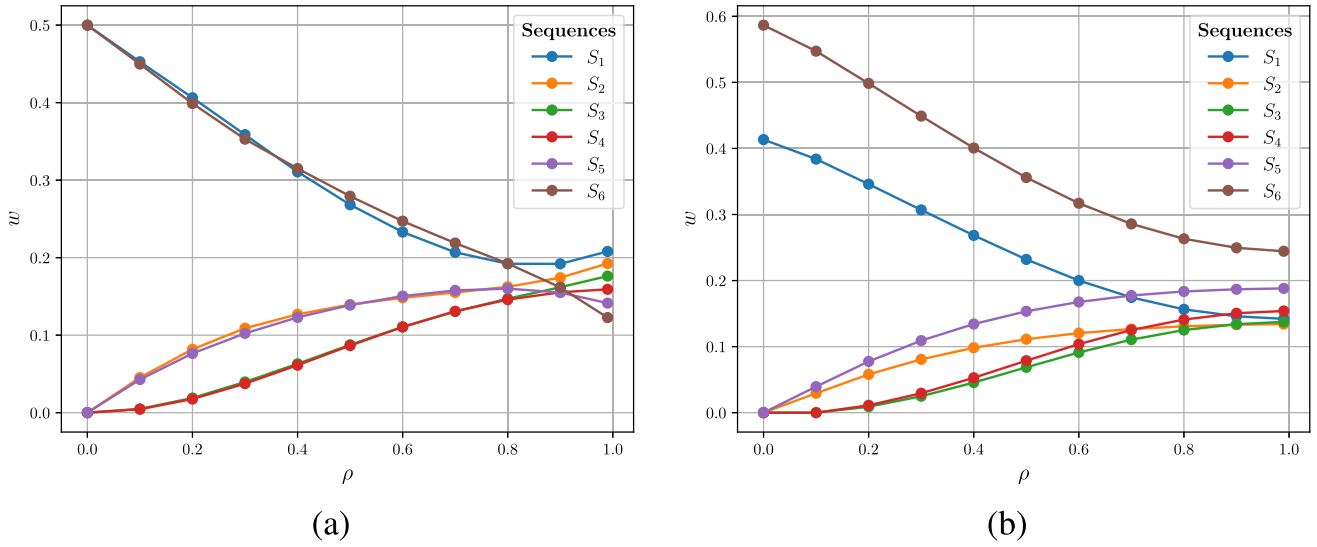

Fig. 6 Comparison of optimal allocation for $\boldsymbol{\rho}$ considering multiple-objective designs: (a) DA-optimal designs ($\tau_A = 0.97$); (b) AD–optimal designs ($\tau_D = 0.97$). Setup: $s_{max} = 6$, $r = 0.05$

We addressed the problem across a broad range of input factors to evaluate the influence of key elements on optimal allocation. Specifically, we considered: (i) the number of sequences; (ii) attrition rates; (iii) optimality criteria; (iv) error correlation structures; and (v) multi-objective designs using the ϵ-constraint method. Our analysis revealed distinct patterns for D–optimal and A-optimal designs. Notably, attrition rates do not impact the allocation of individuals across sequences but affect the allocation within each sequence. D–optimal designs tend to favor symmetric allocation across the sequence order, whereas A–optimal designs result in non-symmetric weights. The proposed formulations and tools proved robust and flexible, effectively accommodating a diverse array of factors.

One interesting extension to the results presented here is to finding exact designs. Although the problem becomes quite challenging, requiring sophisticated MINLP solvers, we would like to obtain some results in an endeavour to determine whether such designs are significant improvements on rounded approximate designs. A second extension is to analyse the robustness of designs to variations in ρ. However, this problem cannot be solved using extensions of the SDP method explored in this paper.

A more general point is to consider using different attrition rates for different groups. For example, those who have been treated may have a lower attrition rate than those who are still waiting for treatment. More generally, designs with dynamic allocation should also be explored. A basic idea of SW designs was to overcome the effect of limited resources or geographical constraints that make it impossible to apply the treatment to a large proportion of the patients. The stepped wedge design allows the researcher to implement the intervention for smaller groups of patients at each time point. However, with attrition, the number of patients presenting at successive time points becomes smaller. A dynamic allocation would switch some patients from control to intervention, so making full use of the available resources. Exploration of such ideas might require the use of cost optimal designs. Examples of such designs for problems in clinical trials are in Fedorov and Leonov (2014, Chapter 7).

Another valuable extension is to multi-arm stepped-wedge cluster randomized trials, where multiple interventions are rolled out across clusters over time. These designs significantly expand the design space and introduce added complexity in allocation and optimality criteria, including treatment ordering, interaction effects, and within- and between-arm correlations. Methodological work in this area is ongoing (e.g., Arnup et al. 2019), and adapting our framework to this setting could support principled optimization of these increasingly relevant designs.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11222-025-10690-y.

Funding Open access funding provided by FCT|FCCN (b-on). Open access funding provided by FCT|FCCN (b-on).

Declarations

Software Availability The software used in this study is available from the corresponding author upon reasonable request.

References

Arnup, S.J., Forbes, A.B., Kahan, B.C., Morgan, K.E., McKenzie, J.E.: Appropriate statistical methods were infrequently used in stepped-wedge cluster randomised trials: a systematic review. J. Clin. Epidemiol. **107**, 77–88 (2019). https://doi.org/10.1016/j.jclinepi.2018.11.006

Ben-Tal, A., Nemirovski, A.S.: Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications. Society for Industrial and Applied Mathematics, Philadelphia (2001)

Boer, E., Hendrix, E.: Global optimization problems in optimal design of experiments in regression models. J. Global Optim. **18**, 385–398 (2000)

Boyd, S., Vandenberghe, L.: Convex Optim. University Press, Cambridge (2004)

Clyde, M., Chaloner, K.: The equivalence of constrained and weighted designs in multiple objective design problems. J. Am. Stat. Assoc. **91**(435), 1236–1244 (1996)

Cook, R.D., Wong, W.K.: On the equivalence of constrained and compound optimal designs. J. Am. Stat. Assoc. **89**(426), 687–692 (1994)

Diamond, S., Boyd, S.: CVXPY: A Python-embedded modeling language for convex optimization. J. Mach. Learn. Res. **17**(83), 1–5 (2016)

Duarte, B.P.M., Atkinson, A.C., Oliveira, N.M.C.: Compound optimal design of experiments – semidefinite programming formulations. Optimization and Engineering (2024). Submitted

Duarte, B.P.M., Granjo, J.F.O., Wong, W.K.: Optimal exact designs of experiments via Mixed Integer Nonlinear Programming. Stat. Comput. **30**, 93–112 (2020)

Duarte, B.P.M., Wong, W.K.: Finding Bayesian optimal designs for nonlinear models: A semidefinite programming-based approach. Int. Stat. Rev. **83**(2), 239–262 (2015)

Duarte, B.P.M., Wong, W.K., Oliveira, N.M.C.: Model-based optimal design of experiments - Semidefinite and nonlinear programming formulations. Chemom. Intell. Lab. Syst. **151**, 153–163 (2016)

Fedorov, V.V., Leonov, S.L.: Optimal Design for Nonlinear Response Models. Chapman and Hall/CRC Press, Boca Raton (2014)

Galbraith, S., Marschner, I.C.: Guidelines for the design of clinical trials with longitudinal outcomes. Control. Clin. Trials **23**(3), 257–273 (2002). https://doi.org/10.1016/s0197-2456(02)00205-2

Gribik, P.R., Kortanek, K.O.: Equivalence theorems and cutting plane algorithms for a class of experimental design problems. SIAM J. Appl. Mathematics **32**, 232–259 (1977)

Harman, R., Filová, L., Richtárik, P.: A randomized exchange algorithm for computing optimal approximate designs of experiments. J. Am. Stat. Assoc. **115**(529), 348–361 (2020). https://doi.org/10.1080/01621459.2018.1546588

Huang, Y.C., Wong, W.K.: Multiple-objective optimal designs. J. Biopharm. Stat. **8**(4), 635–643 (1998). https://doi.org/10.1080/10543409808835265. (**PMID: 9855039**)

Hussey, M.A., Hughes, J.P.: Design and analysis of stepped wedge cluster randomized trials. Contemp. Clin. Trials **28**(2), 182–191 (2007). https://doi.org/10.1016/j.cct.2006.05.007. (**Epub 2006 Jul 7**)

Kadane, J.B. (ed.): Bayesian Methods and Ethics in a Clinical Trial Design. John Wiley & Sons, Hoboken, NJ (2011)

Karmarkar, N.: A new polynomial-time algorithm for linear programming. Combinatorica **4**(4), 373–395 (1984)

Kasza, J., Hemming, K., Hooper, R., Matthews, J., Forbes, A.: Impact of non-uniform correlation structure on sample size and power in multiple-period cluster randomised trials. Stat. Methods Med. Res. **28**(3), 703–716 (2019)

Kiefer, J.: General equivalence theory for optimum design (approximate theory). Ann. Stat. **2**, 849–879 (1974)

Kiefer, J.: Optimal design: Variation in structure and performance under change of criterion. Biometrika **62**(2), 277–288 (1975)

Knifed, E., Lipsman, N., Mason, W., Bernstein, M.: Patients' perception of the informed consent process for neurooncology clinical trials. Neuro Oncol. **10**(3), 348–354 (2008)

Lawrie, J., Carlin, J.B., Forbes, A.B.: Optimal stepped wedge designs. Stat. & Probab. Lett. **99**, 210–214 (2015)

Li, F., Turner, E.L., Preisser, J.S.: Optimal allocation of clusters in cohort stepped wedge designs. Stat. & Probab. Lett. **137**, 257–263 (2018)

Miettinen, K.: Nonlinear Multiobjective Optimization. Kluwer Academic Publishers, Boston (1999)

Moerbeek, M.: Optimal allocation of clusters in stepped wedge designs with a decaying correlation structure. Plos One **18**(8), e0289,275 (2023a)

Moerbeek, M.: Optimal allocation to treatment sequences in individually randomized stepped-wedge designs with attrition. Clin. Trials **20**(3), 242–251 (2023b). https://doi.org/10.1177/17407745231154260. (**Epub 2023 Feb 24**)

Morgan, T.M., Case, L.D.: Conservative sample size determination for repeated measures analysis of covariance. Ann. Biostat. and Biom. **1**(1), 1002 (2013)

MOSEK ApS: MOSEK Modeling Cookbook (2024a) Release 3.3.1. Retrieved from https://docs.mosek.com/modeling-cookbook/index.html

MOSEK ApS: Mosek optimization suite (2024b). https://www.mosek.com. Version 11, accessed: 2024-10-12

Piantadosi, S.: Clinical Trials: A Methodologic Perspective. Willey Series in Probability and Statistics. Wiley (2024)

Pronzato, L., Pázman, A.: Design of Experiments in Nonlinear Models : Asymptotic Normality, Optimality Criteria and Small-Sample Properties, Lecture Notes in Statistics, vol. 212. 2013 edn. Springer New York, New York, NY (2013)

Pukelsheim, F.: Optimal Design of Experiments. SIAM, Philadelphia (1993)

Pukelsheim, F., Rieder, S.: Efficient rounding of approximate designs. Biom. **79**, 763–770 (1992)

Rao, C.R.: Information and the accuracy attainable in the estimation of statistical parameters. Reson. J. Sci. Educ. **20**, 78–90 (1945)

Royall, R.M.: Ethics and statistics in randomized clinical trials. Stat. Sci. **6**(1), 52–62 (1991)

Sagnol, G.: On the semidefinite representation of real functions applied to symmetric matrices. Linear Algebra Appl. **439**(10), 2829–2843 (2013)

Thompson, J.A., Fielding, K., Hargreaves, J., Copas, A.: The optimal design of stepped wedge trials with equal allocation to sequences and a comparison to other trial designs. Clin. Trials **14**(6), 639–647 (2017)

Vandenberghe, L., Boyd, S.: Applications of semidefinite programming. Appl. Numer. Math. **29**, 283–299 (1999)

Wilson, P.M., Ramar, P., Philpot, L.M., Soleimani, J., Ebbert, J.O., Storlie, C.B., Morgan, A.A., Schaeferle, G.M., Asai, S.W., Herasevich, V., et al.: Effect of an artificial intelligence decision support tool on palliative care referral in hospitalized patients: a randomized clinical trial. J. Pain Symptom Manage. **66**(1), 24–32 (2023)

Wong, W.K., Zhou, J.: Using cvx to construct optimal designs for biomedical studies with multiple objectives. J. Comput. Graph.

Stat. **32**(2), 744–753 (2023). https://doi.org/10.1080/10618600.2022.2104858
Ye, Y.: Interior Point Algorithms: Theory and Analysis. John Wiley & Sons, New York (1997)
Zhan, Z., de Bock, G.H., van den Heuvel, E.R.: Optimal unidirectional switch designs. Stat. Med. **37**(25), 3573–3588 (2018). https://doi.org/10.1002/sim.7853
Zhu, W., Wong, W.K.: Multiple-objective designs in a dose-response experiment. J. Biopharm. Stat. **10**(1), 1–14 (2000)