UCLA
UCLA Previously Published Works

Title
On optimal designs for clinical trials: An updated review

Permalink
https://escholarship.org/uc/item/3gp9b5fp

Authors
Sverdlov, Oleksandr
Ryeznik, Yevgen
Wong, Weng Kee

Publication Date
2019-12-09

DOI
10.1007/s42519-019-0073-4

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Journal of Statistical Theory and Practice (2020) 14:10
https://doi.org/10.1007/s42519-019-0073-4

ORIGINAL ARTICLE

On Optimal Designs for Clinical Trials: An Updated Review

Oleksandr Sverdlov[1] · **Yevgen Ryeznik**[2,3] · **Weng Kee Wong**[4]

© Grace Scientific Publishing 2019

Abstract
Optimization of clinical trial designs can help investigators achieve higher quality results for the given resource constraints. The present paper gives an overview of optimal designs for various important problems that arise in different stages of clinical drug development, including phase I dose–toxicity studies; phase I/II studies that consider early efficacy and toxicity outcomes simultaneously; phase II dose–response studies driven by multiple comparisons (MCP), modeling techniques (Mod), or their combination (MCP–Mod); phase III randomized controlled multi-arm multi-objective clinical trials to test difference among several treatment groups; and population pharmacokinetics–pharmacodynamics experiments. We find that modern literature is very rich with optimal design methodologies that can be utilized by clinical researchers to improve efficiency of drug development.

Keywords Estimation efficiency · Dose-finding · Multiple comparisons and modeling · Optimal response-adaptive randomization · Phase I/II studies · Population PK/PD studies · Power

1 Introduction

Recent years have seen significant advances in biotechnology, genomics, and medicine. The number of investigational compounds that hold promise of becoming treatments for highly unmet medical needs has been steadily increasing, and

Part of special issue guest edited by Pritam Ranjan and Min Yang--- Algorithms, Analysis and Advanced Methodologies in the Design of Experiments.

✉ Weng Kee Wong
wkwong@ucla.edu

[1] Early Development Biostatistics, Novartis Pharmaceuticals, East Hanover, USA

[2] Department of Mathematics, Uppsala University, Uppsala, Sweden

[3] Department of Pharmaceutical Biosciences, Uppsala University, Uppsala, Sweden

[4] Department of Biostatistics, Fielding School of Public Health, University of California, Los Angeles, USA

Published online: 09 December 2019

so has the need for efficient research methodologies to evaluate these compounds in clinic. Clinical trial designs are becoming increasingly more elaborate as investigators aim at evaluating the effects of multiple therapies in different subgroups of patients with respect to multiple clinical endpoints within a single trial infrastructure. Such multi-arm and multi-objective clinical trials can potentially increase efficiency of clinical research and development [123].

Optimization of a clinical trial design can help an investigator achieve higher quality results (e.g., higher statistical power or more accurate estimates of the treatment effects) for the given resource constraints. In the context of clinical research, efficient achievement of the study objectives using frequently minimum sample size is particularly important because study subjects are humans, suffering from a severe disease. Medical ethics (e.g., the Declaration of Helsinki) prescribes that every trial participant's welfare is always prime, and therefore optimal clinical trial designs that extract maximum information from the trial while minimizing exposure of study subjects to suboptimal (inefficacious or toxic) treatment regimens warrant careful consideration in practice.

Sverdlov and Rosenberger [109] gave an overview of optimal allocation designs in clinical trials. That review primarily concerned the methods for parallel-group comparative studies where the design points (treatment arms) are prespecified, and an optimal design problem is to determine optimal allocation proportions across the treatment arms. The current paper extends the aforementioned work in the following important ways:

1. We broaden the scope of the review by also considering clinical trials with dose-finding objectives, i.e., where the experimental goals may include estimation of the dose–response relationship and other important parameters, such as quantiles of the dose–response curve.
2. We give an update of important optimal design methodologies for multi-arm comparative clinical trials that have been developed since the publication of the earlier review [109], including some novel optimal allocation targets and response-adaptive randomization methods for implementing these optimal targets in practice.
3. We provide an overview of optimal designs for some population pharmacokinetics–pharmacodynamics experiments using nonlinear mixed-effects models.

In the present paper, we focus on optimal designs only in the context of clinical research. Some broader and more mathematical expositions on modern optimal designs with non-clinical applications can be found in Cook and Fedorov [25], Wong [120], Fedorov [39], Yang et al. [125], among others. There are excellent monographs on design, see for example, Atkinson et al. [3] and Fedorov and Leonov [40].

To facilitate a discussion, we consider a clinical trial with a univariate response variable Y with density function $\pi(y|x, \boldsymbol{\theta})$, where the variable x is subject to control by an experimenter and belongs to some compact set $\boldsymbol{\mathfrak{X}}$, i.e., $x \in \boldsymbol{\mathfrak{X}}$, and $\boldsymbol{\theta}$ is a vector of model parameters. The set $\boldsymbol{\mathfrak{X}}$ may be a closed interval, say,

$\mathfrak{X} = [0, 1]$, which corresponds to a continuum of dose levels; or a finite set, say, $\mathfrak{X} = \{x_1, \ldots, x_K\}$, which may be the case when x_i represents the ith treatment group.

If Y is a normally distributed variable, a common choice is a linear model

$$Y = \boldsymbol{f}'(x)\boldsymbol{\theta} + \varepsilon, \tag{1}$$

where $\boldsymbol{f}'(x) = \big(f_1(x), \ldots, f_p(x)\big)$ and its components are given linearly independent regression functions. The vector $\boldsymbol{\theta}' = \big(\theta_1, \ldots, \theta_p\big)$ contains the unknown regression coefficients, and ε is a random error term that follows a normal distribution with $E(\varepsilon) = 0$ and $\mathrm{var}(\varepsilon) = \sigma^2(x)$. Throughout, we assume all observations are independent and we have resources to take a sample of n subjects.

Suppose $\mathfrak{X} = \{x_1, \ldots, x_K\}$. For a trial of size n, let $n_i \geq 1$ be the number of subjects whose response is to be observed at x_i. With the observed data $\{y_{ij}, i = 1, \ldots, K,\ j = 1, \ldots, n_i\}$, the likelihood function is $\mathcal{L}_n(\boldsymbol{\theta}) = \prod_{i=1}^{K}\prod_{j=1}^{n_i} \pi\big(y_{ij}|x_i, \boldsymbol{\theta}\big)$. The maximum likelihood estimator, $\hat{\boldsymbol{\theta}}_{MLE}$, maximizes $\mathcal{L}_n(\boldsymbol{\theta})$; it can be found by solving the system of p score equations: $\frac{\partial}{\partial\boldsymbol{\theta}} \log \mathcal{L}_n(\boldsymbol{\theta}) = 0$. Under certain regularity conditions on $\pi(y|x, \boldsymbol{\theta})$, $\sqrt{n}\big(\hat{\boldsymbol{\theta}}_{MLE} - \boldsymbol{\theta}\big)$ has asymptotically normal distribution with zero mean and variance–covariance matrix $\mathbf{V}(\xi, \boldsymbol{\theta}) = \mathbf{M}^{-1}(\xi, \boldsymbol{\theta})$, where $\mathbf{M}(\xi, \boldsymbol{\theta})$ is the Fisher information matrix (FIM) for $\boldsymbol{\theta}$ given design $\xi = \big\{\big(x_i, n_i\big), i = 1, \ldots, K\big\}$.

The FIM is the key object in the optimal design theory. Its inverse provides an asymptotic lower bound for the variance of an efficient estimator of $\boldsymbol{\theta}$. With a K-point design $\xi = \big\{\big(x_i, n_i\big), i = 1, \ldots, K\big\}$, the $p \times p$ FIM can be written as

$$\mathbf{M}(\xi, \boldsymbol{\theta}) = \sum_{i=1}^{K} n_i \boldsymbol{\mu}\big(x_i, \boldsymbol{\theta}\big), \tag{2}$$

where $\boldsymbol{\mu}\big(x_i, \boldsymbol{\theta}\big) = -E\Big\{\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} \log \pi(y|x_i, \boldsymbol{\theta})\Big\}$ is the information matrix of a single observation at $x_i \in \mathfrak{X}$, $i = 1, \ldots, K$. For instance, in the normal linear model case, we have $\boldsymbol{\mu}\big(x_i, \boldsymbol{\theta}\big) = \sigma^{-2}\big(x_i\big)\boldsymbol{f}\big(x_i\big)\boldsymbol{f}'\big(x_i\big)$.

To make further progress, we focus on approximate designs, which are probability measures on the design space. We denote such a design by $\xi = \big\{\big(x_i, \rho_i\big), i = 1, \ldots, K\big\}$, where $\rho_i \in (0, 1)$ is the allocation proportion at x_i and $\sum_{i=1}^{K} \rho_i = 1$. For a trial of size n, the number of subjects assigned by the design ξ to x_i is $n_i \approx n\rho_i$, after rounding each $n\rho_i$ to a positive integer, subject to $n_1 + n_2 + \cdots + n_K = n$. Given an objective function carefully selected to reflect the study objective, we first formulate it as a convex function $\Psi(\cdot)$ of $\mathbf{M}(\xi, \boldsymbol{\theta})$ over the space of all designs on the (compact) design space $\mathfrak{X}$. The goal is to find an approximate design ξ^* such that $\xi^* = \arg\min_{\xi} \Psi\big(\mathbf{M}^{-1}(\xi, \boldsymbol{\theta})\big)$. For nonlinear models, the optimal designs depend on $\boldsymbol{\theta}$ and so they are *locally* optimal. This means that such designs can be implemented only when a best guess of the value of $\boldsymbol{\theta}$ is available either from prior or similar studies. For convenience, we refer locally optimal designs simply as optimal designs.

A common choice for $\Psi(\cdot)$ is D-optimality, which seeks a design ξ_D^* that minimizes the function $\log\left|\mathbf{M}^{-1}(\xi,\boldsymbol{\theta})\right| = -\log|\mathbf{M}(\xi,\boldsymbol{\theta})|$over all designs on $\boldsymbol{\mathfrak{X}}$. Such a design minimizes the generalized variance of the estimates of $\boldsymbol{\theta}$ and so estimates the model parameters most accurately. The D-optimal design ξ_D^* is typically used as an important benchmark to facilitate comparison among designs for estimating $\boldsymbol{\theta}$. The D-efficiency of a design ξ relative to ξ_D^* is defined as

$$D_{\text{eff}}(\xi,\boldsymbol{\theta}) = \left\{\frac{\left|\mathbf{M}^{-1}\left(\xi_D^*,\boldsymbol{\theta}\right)\right|}{\left|\mathbf{M}^{-1}(\xi,\boldsymbol{\theta})\right|}\right\}^{1/p}. \tag{3}$$

For instance, if $D_{\text{eff}}(\xi,\boldsymbol{\theta}) = 0.90$, this means that the design ξ is 90% as efficient as ξ_D^*, and the sample size for a study with the design ξ must be increased by 10% to achieve the same level of estimation efficiency as with the D-optimal design ξ_D^*. Note that the D-optimal design addresses a single objective, i.e., to estimate all parameters in the mean function of the model as accurately as possible. Other objectives can carry equal, or even greater importance to an experimenter, and finding designs that provide optimal trade-off among the selected objectives are often warranted.

In this paper, we discuss various approaches to optimization of clinical trial designs. In Sect. 2, we give an overview of optimal designs for dose-finding studies that are ubiquitous in early clinical development. We assume the dose–response relationship can be adequately described by some nonlinear and/or heteroscedastic regression model, and the design space consists of a continuum of dose levels. In this case, the search of an optimal design involves determining the number of optimal design points, the location of these points in the design space, and the corresponding optimal allocation proportions. The solution can be mathematically complex and it frequently must be found by numerical methods. We shall discuss some statistical software tools available for this purpose. In Sect. 3, we discuss multi-objective optimal designs for parallel-group, randomized comparative clinical trials. Unlike the designs for early clinical development that are driven primarily by the goals of identifying some target dose(s) and estimation of the dose–response, randomized comparative studies are driven by hypothesis testing and statistical power considerations. A common goal is to achieve the desired level of statistical power while minimizing the total study size or the total expected number of treatment failures in the trial. The resulting optimal allocation designs often depend on the model parameters and call for response-adaptive randomization (RAR) for their implementation. We shall review some important recent advances in this field. In Sect. 4, we give an overview of optimal designs for population pharmacokinetic experiments, which are gaining increased popularity in modern clinical research. The optimal designs for such experiments involve careful balance between the selection of the PK sampling times and the number of subjects to include in the study to reduce study costs. Section 5 offers concluding remarks and outlines some important research in the field of optimal designs for clinical trials that are likely to gain attraction in the near future.

2 Optimal Designs for Dose-Finding Clinical Trials

The importance of dose-finding in clinical drug development cannot be underestimated [34, 35]. Identification of the "right" dose regimen(s) in early clinical studies can increase probability of success of subsequent randomized comparative clinical trials. A nice exposition of modern dose-finding methods can be found in [74]. Dose-finding designs can be categorized into three major types: phase I dose-escalation designs; phase I/II seamless designs; and phase II dose-ranging studies. In this section, we provide an overview of optimal designs for each of these types of studies.

2.1 Phase I Dose-Escalation Studies

The clinical part of any new drug development program starts with phase I studies to investigate safety, tolerability, and pharmacokinetics of the compound. Phase I first-in-human studies are commonly cast as dose-escalation designs. Study subjects are exposed to the drug at increased dose levels in a staggered manner: only when the previous dose is deemed as "safe" would the next cohort of subjects be assigned to the next dose level. The primary objective is to determine the maximum tolerated dose (MTD)—the highest dose level at which the risk of toxicity (side effects) is "acceptable"—which is thought to subsequently demonstrate therapeutic benefit (efficacy) in larger-scale studies.

The majority of innovative phase I dose-finding methodologies were developed in the context of cytotoxic anti-cancer compounds, where an inherent assumption is a monotone relationship between the dose and the risk of toxicity. The problem of finding the MTD can be then viewed as determination of a percentile of the dose–toxicity curve. In the literature, one can find numerous sequential design methodologies that can be useful for this purpose. These include nonparametric approaches, such as the up-and-down design [32]; parametric model-based approaches, such as the continual reassessment method [75] and escalation with overdose control [5]; semiparametric designs [23], etc. A recent review of many of these methods is given in [112].

In this paper, we focus on designs for estimating the parameters of interest in phase I dose-toxicity studies. These designs are rarely used in clinical practice because it is very difficult for Institutional Review Boards to justify a design that optimizes a statistical criterion and not take patient safety into account. The key merit of these optimal designs is that they provide important theoretical benchmarks to facilitate comparison with other more heuristically chosen designs.

To fix ideas, consider using a two-parameter logistic model in a dose–toxicity study. The design space is $\mathfrak{X} = \{x_1 < x_2 < \cdots < x_K\}$ comprising pre-specified doses to be investigated in the study. Assume the binary outcome ($Y = 1$ if toxicity, and $Y = 0$ otherwise). The model is

$$P(x_i) = \Pr(Y = 1|x_i) = \frac{1}{1 + \exp\{-(\alpha + \beta x_i)\}}, \quad i = 1, \ldots, K, \tag{4}$$

where α (intercept) and $\beta > 0$ (slope) are unknown parameters to be estimated based on data. Let $\gamma = \log\left(\frac{\Gamma}{1-\Gamma}\right)$, where $\Gamma \in (0, 1)$ is a predetermined constant (target

toxicity rate). Then, $D_{100\Gamma} = \frac{\gamma-\alpha}{\beta}$ is the 100Γ-th percentile of the curve $P(x)$; i.e., it is the dose for which $P\left(D_{100\Gamma}\right) = \Gamma$. In particular, if $\Gamma = 0.5$, $D_{50} = -\alpha/\beta$ is the 50th percentile (median).

For estimation of the model (4), we use the data $\mathcal{D} = \left\{\left(x_i, n_i, y_i\right),\ i = 1, \ldots, K\right\}$, where n_i is the number of subjects exposed at x_i (each subject in the study is assigned to only one dose), and $y_i = \sum_{j=1}^{n_i} Y_{ij}$ is the number of toxicities at x_i, such that $y_i \sim Binomial\left(n_i, P\left(x_i\right)\right)$, $i = 1, \ldots, K$. The maximum likelihood estimator (MLE) $(\hat{\alpha}, \hat{\beta})$ of (α, β) are the roots of the system of score equations $\frac{\partial}{\partial \boldsymbol{\theta}} \log \mathcal{L}(\boldsymbol{\theta}) = 0$, where $\log \mathcal{L}(\boldsymbol{\theta}) = \sum_{i=1}^{K} \left\{y_i \log P\left(x_i\right) + \left(n_i - y_i\right) \log\left(1 - P\left(x_i\right)\right)\right\}$ and $\boldsymbol{\theta}' = (\alpha, \beta)$. By the invariance property of MLEs, MLEs of other parameters can be readily obtained, for example, $\hat{D}_{100\Gamma} = \frac{\gamma-\hat{\alpha}}{\hat{\beta}}$ is the estimate of the 100Γ-th percentile of the logistic model.

The FIM for (α, β) given design $\xi = \left\{\left(x_i, \rho_i\right),\ i = 1, \ldots, K\right\}$ is $\mathbf{M}(\xi, \alpha, \beta) = n \sum_{i=1}^{K} \rho_i \boldsymbol{\mu}\left(x_i, \alpha, \beta\right)$, where $\boldsymbol{\mu}\left(x_i, \alpha, \beta\right) = \phi_i(\alpha, \beta) \begin{pmatrix} 1 & x_i \\ x_i & x_i^2 \end{pmatrix}$ and $\phi_i(\alpha, \beta) = e^{-\left(\alpha+\beta x_i\right)} / \left(1 + e^{-\left(\alpha+\beta x_i\right)}\right)^2$, $i = 1, \ldots, K$. Asymptotically, $(\hat{\alpha}, \hat{\beta})$ follows a bivariate normal distribution with mean (α, β) and variance–covariance matrix $\mathbf{M}^{-1}(\xi, \alpha, \beta)$. Once we have $(\hat{\alpha}, \hat{\beta})$ and $\mathbf{M}^{-1}\left(\xi, \hat{\alpha}, \hat{\beta}\right)$, we can construct (asymptotic) confidence intervals for various parameters of interest, including α, β, $P(x)$ (where x is not necessarily among the doses tested), $D_{100\Gamma}$, etc.

The next important question is: Which design is optimal for this model? Depending on the trial objectives, different criteria can be optimized. If we want to maximize efficiency in estimation of the entire dose–toxicity relationship, three possible criteria are the D-optimality $\det\{\mathbf{M}^{-1}(\xi, \alpha, \beta)\}$ (to minimize the volume of the confidence ellipsoid for (α, β)); the A-optimality $\text{trace}\{\mathbf{M}^{-1}(\xi, \alpha, \beta)\}$ (to minimize the sum of the lengths of the major axes in the confidence ellipsoid for (α, β)); and the E-optimality: maximum eigenvalue of $\mathbf{M}^{-1}(\xi, \alpha, \beta)$ (to minimize the length of the largest axis in the confidence ellipsoid for (α, β)). A comprehensive treatment of this problem is available, for instance, in [102], where the authors found that for a broad class of symmetric models (including the logistic model), and a class of different optimality criteria (including D-, A-, and E-optimality), the optimal design is a 2-point design, symmetric about D_{50}, with possibly unequal weights. Furthermore, Yang and Stufken [126] gave a general solution for nonlinear models with two parameters, including logistic, probit, double exponential and double reciprocal models for binary data, a log-linear Poisson model for count data, and the Michaelis–Menten model. Their results are applicable to any functions of the original parameters, any commonly used optimality criteria, and the design space can be restricted or non-restricted. That paper essentially unified most optimal design work for 2-parameter generalized linear models and other nonlinear models.

Let us describe one particularly insightful result, namely the structure of the locally D-optimal design for a 2-parameter logistic model. Minkin [70], Sitter and Wu [104], Sitter and Fainaru, [102], Matthew and Sinha [67], among others, reported that the locally D-optimal design is symmetric and equally supported at the 17.6th and 82.4th percentiles of the dose–response curve (4):

$$\xi_D^* = \left\{ \left(D_{17.6}, 0.5\right), \left(D_{82.4}, 0.5\right) \right\}, \tag{5}$$

where $D_{17.6} = \frac{-c^* - \alpha}{\beta}$, $D_{82.4} = \frac{c^* - \alpha}{\beta}$, and $c^* = \log\left(\frac{0.824}{0.176}\right) = 1.5434$. Moreover, the structure of the D-optimal design remains the same for many other estimation objectives [67]. If, say, θ_1 and θ_2 are two functions of α and β, then the information matrix for (θ_1, θ_2) is $\boldsymbol{J}\mathbf{M}(\xi, \alpha, \beta)\boldsymbol{J}'$, where the matrix $\boldsymbol{J}$ does not depend on the dose levels. Hence, the design (5) is also D-optimal, for instance, for the joint estimation of β (slope) and $D_{100\Gamma}$ (100Γ-th percentile), or for the joint estimation of $D_{100\Gamma_1}$ and $D_{100\Gamma_2}$, where $\Gamma_1, \Gamma_2 \in (0, 1)$ (two different percentiles).

From Eq. (5), one can see several limitations of the D-optimal design that makes its application in real clinical research problematic: i) The D-optimal design was obtained under the assumption of a 2-parameter logistic model, which may be misspecified in many ways; ii) even if the logistic model is plausible, the D-optimal design is a function of the model parameters (α, β) which are unknown at the outset; iii) the D-optimal design addresses the goal of estimating the entire dose–toxicity curve, whereas the most common goal of a phase I clinical trial is to cluster dose assignments at and around the target percentile; iv) the D-optimal design allocates 50% of the subjects to the dose with toxicity probability 17.6% (which may be viewed as too low from the clinical perspective) and 50% of subjects to the dose with toxicity probability 82.4% (which may be prohibitively toxic).

To address issue i), one can consider design optimization for more elaborate models, such as 3- or 4-parameter logistic [58], or even 5-parameter logistic [64]. Link functions other than logistic can be considered as well; for instance a semiparametric model can help alleviate the problem of mis-specification of the distributional assumption on the dose–toxicity curve [119]. One should be mindful, however, that phase I studies are typically small, and it may be problematic to fit complex models due to sparsity of the data. Therefore, striking the right balance between model parsimony and rigor is essential.

The issue ii)—the dependence of optimal designs on the true parameter values—is common to nonlinear (and heteroskedastic) models. It is referred to as *local optimality* of the designs. An experimenter may decide to implement the design for the best guess of the parameter values [22]; however, the efficiency of such a design may drop if the true values are different from the guessed ones. There are three approaches to mitigate this problem: adaptive designs, minimax designs, and Bayesian designs. The first approach is actually the core of any phase I dose-escalation study [112]. Since clinical trials are sequential experiments, one can use accumulating data for updating the dose–toxicity curve, and direct future dose assignments to the targeted optimal design. Various adaptive procedures can be constructed to approximate the targeted optimal design [45, 62, 68, 91, 103]. For instance, an ethically restricted sequential D-optimal design for model (4) can be constructed iteratively as follows [66, 119]. Based on data from n patients and design ξ_n after n allocations, obtain estimates $\left(\hat{\alpha}, \hat{\beta}\right)$ of (α, β) and update the feasible dose range to $\mathfrak{X}_R = \left\{x \in \mathfrak{X} : x \le \hat{\mu}_R\right\}$, where $\hat{\mu}_R = \frac{\gamma_R - \hat{\alpha}}{\hat{\beta}}$ and $\gamma_R = \log\left(\frac{\Gamma_R}{1-\Gamma_R}\right)$ for some predetermined $\Gamma_R \in (0, 1)$. Essentially, $\hat{\mu}_R$ is the estimated highest dose to which patients can be assigned safely. Then, the dose assignment for the $(n+1)$st patient, δ_{n+1}, is

determined as one that maximizes an incremental gain in information, subject to a constraint on the upper bound for the dose range:

$$\delta_{n+1} = \arg\max_{x_i \in \mathfrak{X}_R} \left| \mathbf{M}\left(\xi_n, \hat{\alpha}, \hat{\beta}\right) + \boldsymbol{\mu}\left(x_i, \hat{\alpha}, \hat{\beta}\right) \right|. \tag{6}$$

Note that the procedure (6) can be only implemented after initial reliable estimates $\left(\hat{\alpha}, \hat{\beta}\right)$ are available, which may be challenging in phase I trials. Therefore, a "start-up" procedure to ascertain initial data and model estimates must be chosen judiciously [62].

With the minimax approach, the idea is to find a design that is robust to the choice of the parameter values, by minimizing the maximum loss in efficiency with respect to the locally optimal designs over the range of potential model parameter values [56, 101]. For instance, a minimax D-optimal design problem is to find $\xi^*_{\mathrm{M}} = \arg\min_{\xi} \max_{(\alpha,\beta)\in\wp} \left|\mathbf{M}^{-1}(\xi, \alpha, \beta)\right|$, where $\wp$ denotes the pre-specified "region of robustness" of the design, reflecting uncertainty in the parameters.

With the Bayesian approach, we want to find a design that maximizes average efficiency with respect to the locally optimal designs for a given prior distribution of the parameters [19, 20]. A Bayesian D-optimal design problem is to find $\xi^*_{\mathrm{B}} = \arg\min_{\xi} \int \log\left|\mathbf{M}^{-1}(\xi, \boldsymbol{\theta})\right| g(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta}$, where $g(\boldsymbol{\theta})$ is a prior probability density for $\boldsymbol{\theta} = (\alpha, \beta)$.

Due to added model uncertainty, both minimax and Bayesian optimal designs typically have more support points and more complex structure than the locally optimal designs; they are usually not mathematically tractable and must be found numerically.

The issues iii) and iv) are closely related. The statistical goal of estimating the MTD (100Γ-th percentile of the dose–toxicity curve) and the "ethical" goal of treating the majority of study patients at the true MTD are in good correspondence; however, achieving these goals in practice may be problematic because the true MTD is unknown. It is intuitive that in order to estimate the target percentile with most precision, the design should assign all subjects to this unknown dose level. However, this is not always the case. Formally, if the parameter of interest is $D_{100\Gamma} = \frac{\gamma-\alpha}{\beta}$, then the asymptotic variance of $\hat{D}_{100\Gamma}$, the MLE of $D_{100\Gamma}$, can be approximated using delta method as $\mathrm{var}\left(\hat{D}_{100\Gamma}\right) \approx \boldsymbol{c}'\mathbf{M}^{-1}(\xi, \alpha, \beta)\boldsymbol{c}$, where $\boldsymbol{c}' = \left(\frac{\partial D_{100\Gamma}}{\partial \alpha}, \frac{\partial D_{100\Gamma}}{\partial \beta}\right) = -\frac{1}{\beta}\left(1, D_{100\Gamma}\right)$. The c-optimal design is one that minimizes $\mathrm{var}\left(\hat{D}_{100\Gamma}\right)$, i.e., $\xi^*_c = \arg\min_{\xi}\left\{\boldsymbol{c}'\mathbf{M}^{-1}(\xi, \alpha, \beta)\boldsymbol{c}\right\}$. The structure of the c-optimal designs under different distributions, including one in Eq. (4), can be found in [124]. It is either a one-point design centered at the percentile of interest, or a two-point design with one point at the lower part and the other point at the upper part of the dose–toxicity curve. The latter case, for ethical reasons, is infeasible in practice. To overcome this limitation, *restricted* optimal designs, where a restriction on the dose range reflects ethical constraints, have been proposed [45, 66, 85]. Ethically constrained Bayesian D- and c-optimal designs [45, 85] can be viable in practice due to their established theoretical properties [92] and availability of the corresponding statistical software [84].

Since most clinical trials are multi-objective experiments, it is problematic to have a single optimization criterion that would adequately fulfill the desired experimental objectives. Two common approaches for constructing multi-objective optimal designs are to find the *constrained* and the *compound* optimal designs. The equivalence of these two approaches was established by Cook and Wong [26] and Clyde and Chaloner [24]. Some examples of applications of constrained and compound optimal designs in clinical trials can be found in [7, 8, 109].

2.2 Phase I/II Efficacy–Toxicity Studies

A conventional path of developing a new cytotoxic drug in oncology involves two major steps: a phase I dose-escalation trial to identify the maximum tolerated dose (MTD), followed by a phase II trial to study the drug activity (therapeutic response) at the MTD. However, such a path may be inappropriate for development of a targeted therapy, such as a cytostatic agent in immuno-oncology, because such therapies have lower potential for toxicity than cytotoxic drugs, and their dose–efficacy curve may peak or reach plateau at doses below the MTD. The designs for targeted therapies require special considerations. An increasingly popular approach is the seamless phase I/II trial that includes both safety (toxicity) and therapeutic (efficacy) concerns in the dose-finding objectives [128]. An advantage of such an approach is that important, potentially correlated, clinical safety and efficacy outcomes are investigated jointly within the same trial, which can be more efficient than investigating each of these outcomes in separate studies [108].

Formulating a joint model for efficacy and toxicity is an important first step for a seamless phase I/II trial. Let us consider a simple case when the dose space is $\mathfrak{X} = \{x_1 < x_2 < \cdots < x_K\}$, and both efficacy and toxicity outcomes are binary: $Y_{\mathrm{T}} = 1(0)$ if toxicity (no toxicity); and $Y_{\mathrm{E}} = 1(0)$ if efficacy (no efficacy). Let $p(x) = \Pr(Y_{\mathrm{T}} = 1|x)$ and $q(x) = \Pr(Y_{\mathrm{E}} = 1|x)$ denote, respectively, the marginal probabilities of toxicity and efficacy at $x \in \mathfrak{X}$. Assume that both $p(x)$ and $q(x)$ are monotone increasing. The maximum tolerated dose MTD is defined by $\mathrm{MTD} = \max\{x \in \mathfrak{X} : p(x) \le \bar{p}_{\mathrm{T}}\}$ and the minimum efficacious dose MED is defined by $\mathrm{MED} = \min\left\{x \in \mathfrak{X} : q(x) \ge \underline{q}_{\mathrm{E}}\right\}$, for some user-specified thresholds $\bar{p}_{\mathrm{T}}, \underline{q}_{\mathrm{E}} \in (0, 1)$. A *therapeutic window* is the interval [MED, MTD] when $\mathrm{MED} < \mathrm{MTD}$; otherwise it is an empty set. Looking at the outcomes $(Y_{\mathrm{T}}, Y_{\mathrm{E}})$ jointly, there are several possibilities. For instance, a three-category model is obtained by defining a variable $Z = 0$, if $(Y_{\mathrm{T}}, Y_{\mathrm{E}}) = (0, 0)$ (no efficacy and no toxicity); $Z = 1$, if $(Y_{\mathrm{T}}, Y_{\mathrm{E}}) = (0, 1)$ (efficacy without toxicity); and $Z = 2$, if $(Y_{\mathrm{T}}, Y_{\mathrm{E}}) = (1, *)$ (toxicity). Such a trinomial outcome model is also known as a *contingent response* model, where an occurrence of toxicity makes efficacy irrelevant [83]. A four-category model assumes that any of the combinations $(Y_{\mathrm{T}}, Y_{\mathrm{E}}) = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$ can be observed. In either case, the outcome $(Y_{\mathrm{T}}, Y_{\mathrm{E}}) = (0, 1)$ is regarded as a "success," and one can define the probability of success at dose $x \in \mathfrak{X}$ as $s(x) = \Pr(Y_{\mathrm{E}} = 1|Y_{\mathrm{T}} = 0, x) \times \Pr(Y_{\mathrm{T}} = 0|x)$. A dose $d^* \in \mathfrak{X}$ that maximizes $s(x)$ is called the *most successful dose* (MSD). If, in addition,

it satisfies the safety constraint $\Pr\left(Y_{\mathrm{T}}=1|d^{*}\right)\leq\bar{p}_{\mathrm{T}}$, then it is called the safe MSD (sMSD). The objectives of a seamless phase I/II trial may be to identify sMSD (or stop the trial early if no such dose exists), and estimate it as accurately as possible. Locally optimal designs for estimation of various parameters in this context can be important calibrating tools for clinical investigators.

Optimal designs for various trinomial (contingent response) models were investigated by Fan and Chaloner [37, 38] and Rabie and Flournoy [82, 83]. In particular, Rabie and Flournoy [83] gave a comprehensive description of both D- and c-optimal designs for estimating MSD for contingent response models using different distributions for probabilities of toxicity and efficacy given no toxicity.

Optimal designs for bivariate binary models where all four outcomes are observable have also received attention in the literature [30, 31, 46]. It is instructive to consider a bivariate Gumbel model, which is a natural extension of a univariate logistic model (4). The marginal probabilities of efficacy and toxicity are modeled using the logistic distribution function given by $F(x)=\{1+\exp(-x)\}^{-1}$, i.e., $\pi_{\mathrm{E}}=\Pr\left(Y_{\mathrm{E}}=1|x\right)=F\left(\mu_{\mathrm{E}}+\beta_{\mathrm{E}}x\right)$ and $\pi_{\mathrm{T}}=\Pr\left(Y_{\mathrm{T}}=1|x\right)=F\left(\mu_{\mathrm{T}}+\beta_{\mathrm{T}}x\right)$ ($\beta_{\mathrm{E}},\beta_{\mathrm{T}}>0$ to ensure both have monotone increasing relationships). Let ρ be a parameter to characterize correlation between efficacy and toxicity ($|\rho|<1$). Then, the joint probability function can be written as

$$\begin{aligned}\pi_{y,z}(x,\boldsymbol{\theta})=\Pr\left(Y_{\mathrm{E}}=y,Y_{\mathrm{T}}=z|x,\boldsymbol{\theta}\right)&=\left\{\pi_{\mathrm{E}}\right\}^{y}\left\{\pi_{\mathrm{T}}\right\}^{z}\left\{1-\pi_{\mathrm{E}}\right\}^{1-y}\left\{1-\pi_{\mathrm{T}}\right\}^{1-z}\\&+(-1)^{y+z}\frac{e^{\rho}-1}{e^{\rho}+1}\pi_{\mathrm{E}}\pi_{\mathrm{T}}\left\{1-\pi_{\mathrm{E}}\right\}\left\{1-\pi_{\mathrm{T}}\right\}.\end{aligned}\tag{7}$$

The model (7) is defined by a 5-parameter vector $\boldsymbol{\theta}=\left(\mu_{\mathrm{E}},\mu_{\mathrm{T}},\beta_{\mathrm{E}},\beta_{\mathrm{T}},\rho\right)$. Other bivariate binary models are available in Ch. 6.5 of Fedorov and Leonov [40]. Locally D- and c-optimal designs for model (7) and some other models have been investigated Dragalin and Fedorov [30] and Dragalin et al. [31]. Many of these designs may assign high proportions to overly toxic and/or inefficacious doses. To address this limitation, Dragalin and Fedorov [30] proposed *penalized* optimal designs (see also [79]. The idea is to introduce a cost function that penalizes doses with high toxicity and low efficacy: $C(x,\boldsymbol{\theta})=\left\{\pi_{1,0}(x,\boldsymbol{\theta})\right\}^{-C_{\mathrm{E}}}\left\{1-\pi_{0,1}(x,\boldsymbol{\theta})-\pi_{1,1}(x,\boldsymbol{\theta})\right\}^{-C_{\mathrm{T}}}$, where $C_{\mathrm{E}},C_{\mathrm{T}}\geq 0$ are user-defined constants. The total cost of a K-point design ξ is $C_{\mathrm{tot}}(\xi,\boldsymbol{\theta})=\sum_{i=1}^{K}\rho_i C\left(x_i,\boldsymbol{\theta}\right)$ and the penalized D-optimal design is $\xi^{*}_{\mathrm{pen}}=\arg\min_{\xi}\left\{\log\frac{|\mathbf{M}^{-1}(\xi,\boldsymbol{\theta})|}{C_{\mathrm{tot}}(\xi,\boldsymbol{\theta})}\right\}$. Such a design provides maximum information per cost unit; in this particular case, the cost is the penalty for treating patients at highly toxic or inefficacious doses. Note that when $C_{\mathrm{E}}=C_{\mathrm{T}}=0$, the problem reduces to finding the D-optimal design; otherwise, we have a design that provides some trade-off between information and treatment goals. Likewise, penalized c-optimal or some other penalized letter optimal designs can be constructed.

A limitation of local optimality can be overcome by constructing *adaptive* penalized optimal designs. After initial pilot data have been ascertained, one obtains model parameter estimates, and performs dose assignments for subsequent patients adaptively, either sequentially or in cohorts, to approximate the targeted optimal

design. When properly calibrated, such adaptive designs show very reasonable, competitive performance, as shown both theoretically and through simulations [30, 31, 80].

In summary, we would like to highlight some further advantages and limitations of phase I/II optimal designs. One major advantage is that the approach is very flexible. In principle, bivariate models can be extended to accommodate several control variables (e.g., doses of different drugs) and important prognostic covariates. This can be useful in studies of drug combinations and in personalized dose-finding trials, where the drug effect is expected to vary across patient subgroups. The methodology is not limited to binary responses; Padmanabhan et al. [78] and Magnusdottir [63] considered the cases when efficacy and safety are continuous, or one is binary and the other one is continuous. One limitation of adaptive optimal phase I/II designs, which is shared with adaptive optimal designs for phase I toxicity studies, is that they may be difficult to pass Institutional Review Boards due to technical complexity of algorithms and potential ethical restrictions. An additional operational challenge is that the efficacy outcome is usually observed after some delay, which may make design adaptations problematic. Some methods to alleviate this problem, for example using Bayesian data augmentation algorithms, are discussed in [128].

2.3 Phase II Dose-Ranging Studies

Phase II drug development usually starts after phase I studies have demonstrated acceptable safety, tolerability, and PK properties of the investigational compound. The common objectives of phase II are two-fold: to assess the drug effect (clinical efficacy) in patients with the disease of interest, and to identify dose(s) with most promising benefit/risk ratio for testing in confirmatory phase III trials. Phase II trial designs are typically randomized, controlled studies involving several doses of an investigational drug, with sample sizes up to several hundred patients.

There are several approaches to phase II trial designs. The first one is based on multiple comparisons—the dose is regarded as a classification factor, and minimal (if any) assumptions are made on the underlying shape of the dose–response relationship. Let Y denote a continuous outcome variable, where large values of Y signify clinical efficacy. Let $0 < d_1 < d_2 < \cdots < d_K$ be the selected dose levels of the drug, with $d_0 = 0$ being the placebo. A simple statistical model for the outcome of interest is

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \tag{8}$$

where μ_i is the effect at dose d_i, ε_{ij}'s are independent error terms assumed to be normally distributed with mean 0 and variance σ_i^2, and n_i denotes the number of subjects at d_i for $i = 0, 1, \ldots, K$, $j = 1, \ldots, n_i$.

Various objectives can be formulated for model (8), e.g., testing the homogeneity hypothesis $H_0 : \mu_0 = \mu_1 = \cdots = \mu_K$ versus some alternative; identifying the minimum efficacious dose (MED), defined as $\text{MED} = \min\{d_i : \mu_i > \mu_0 + \Delta\}$ for some user-selected clinically relevant parameter $\Delta > 0$; estimating different treatment contrasts, etc. Depending on the study goals, various optimal designs can be

constructed. In this case, finding an optimal design means finding the vector of optimal treatment allocation proportions, which often times can be done explicitly, using Lagrange multiplier optimization. Sverdlov and Rosenberger [109] provided a comprehensive review of various single-objective and multi-objective optimal designs for model (8). For instance, the D_A-optimal design minimizing the volume of a confidence ellipsoid for a vector of treatment–placebo contrasts $(\mu_1 - \mu_0, \ldots, \mu_K - \mu_0)$ has an interesting structure: Its treatment allocation proportions are ordered consistently with the magnitude of treatment variances, such that more variable treatment groups receive higher proportions of subjects [122]. Other minimization objectives (and their combinations) will clearly result in different optimal designs. Therefore, a clear articulation of the experimental goals is important before the start of the study.

The second approach to phase II dose–response study designs is based on modeling. In contrast to Eq. (8), the drug effect is assessed using a regression model

$$Y_{ij} = f(d_i, \boldsymbol{\theta}) + \varepsilon_{ij}, \tag{9}$$

where $f(d_i, \boldsymbol{\theta})$ is the mean response at d_i (some (non)linear function), $\boldsymbol{\theta}$ are the parameters of interest, and ε_{ij}'s are independent (e.g., normal) error terms. The form of the dose–response should reflect the underlying biological mechanism of the drug effect, which clearly depends on the disease and the drug studied. For example, a very flexible class is a 4-parameter Emax model: $f(d_i, \boldsymbol{\theta}) = E_0 + E_{\max}\frac{d^r}{\mathrm{ED}_{50}^r + d^r}$, where $\boldsymbol{\theta}' = (E_0, \mathrm{ED}_{50}, E_{\max}, r)$. An advantage of using a regression model (9) over a saturated model (8) is that the dose is regarded as a continuous predictor for the mean response and regression modeling allows borrowing information across the range of doses to extrapolate the results beyond the doses actually studied. For (9), the *minimum efficacious dose* is defined as $\mathrm{MED}(\boldsymbol{\theta}) = \inf\{d \in (0, d_K] : f(d, \boldsymbol{\theta}) \geq f(0, \boldsymbol{\theta}) + \Delta\}$ and Δ is a user-specified positive constant. One potential limitation of the modeling approach is that it is model-dependent, and the model can be mis-specified in a number of ways, thereby complicating the design.

Optimal designs for dose–response studies driven by the modeling approach involve searches over a continuum of dose levels in the interval $(0, d_K]$. Once the study objectives have been identified and nominal values of $\boldsymbol{\theta}$ are available, we can use algorithms to generate locally optimal designs. Besides D-optimal designs, some useful selected references include c-optimal designs for estimating percentiles and minimum efficacious doses for various dose–response models [13, 27, 28, 77, 132], optimal designs for models with quadratic terms for the dose effect [36, 52, 53], optimal designs for estimating the interesting part of a dose-effect curve [69], etc. A limitation of local optimality can be overcome by means of adaptive designs. Some interesting simulation results can be found in the papers by a PhRMA adaptive dose-ranging studies working group [17, 29].

Since the majority of phase II studies are randomized, careful calibrations of a randomization procedure to implement the chosen optimal design is warranted. Ryeznik et al. [94, 95] investigated multi-stage adaptive D-optimal designs for dose–response studies with time-to-event outcomes. They found that both the choice of the allocation design and the randomization procedure can affect the quality of model estimates. For best performance, one should use a randomization procedure

that closely attains the targeted optimal design at each stage, especially when sample sizes are small.

Finally, the third approach to phase II dose–response studies is a combination of multiple comparisons (MCP) with modeling (Mod) techniques (MCP-Mod). The MCP-Mod, originally proposed by Bretz et al. [18] can efficiently handle model uncertainty at the design stage. The method is cast in two parts. The MCP part addresses a simple Yes/No question—whether there is any effect due to the drug. At the beginning of the trial, a set of candidate models that describe plausible dose–response relationships is postulated. When experimental data become available, the significance of dose–response is tested for each model, with appropriate adjustments for multiplicity. If the null hypothesis of a flat dose response cannot be rejected for any of the models, the procedure stops; otherwise, the Mod part commences—it involves modeling of the dose–response and subsequent application of various statistical inference procedures for the model(s) that is (are) most appropriate given the data. The MCP-Mod method has become increasingly popular in practice. Many clinical trial applications and numerous extensions of the original method have been developed; see Part III of O'Quigley et al. [74] for details. Furthermore, MCP–Mod received a qualification opinion from the European Medicines Agency in 2014 as a methodology that…*will promote better trial designs incorporating a wider dose range and increased number of dose levels* [35].

We would like to give one example of a successful phase II trial application using MCP-Mod. Selmaj et al. [97] conducted an adaptive dose-ranging, randomized, placebo-controlled phase II trial to evaluate safety, tolerability and efficacy of BAF312 (siponimod) in patients with relapsing-remitting multiple sclerosis (RRMS). The primary objective was to evaluate the dose–efficacy relationship among five doses of siponimod and placebo during 3 months of treatment in adult patients with RRMS. The primary endpoint was the number of combined unique active MRI lesions (CUAL) at 3 months, modeled using a negative binomial regression. The design included three main parts: stage 1, interim analysis, and stage 2. At stage 1, four doses were tested: placebo (0 mg), 0.5 mg, 2 mg, and 10 mg, with the total sample size of $n_1 = 188$ (47 patients per arm). At the interim analysis, three assessments were made: (1) analysis for futility (turned out to be negative); (2) calibration of the dose range by fitting six candidate dose–response models and choosing one that showed the highest correlation to the observed negative binomial fit, whereupon a 2-parameter Emax model was chosen, and additional two doses, 0.25 mg and 1.25 mg, were selected for testing in stage 2; and (3) sample size reassessment based on parameters not directly related to treatment effect (it turned out that increasing the sample size was unnecessary). At stage 2, three groups were tested: placebo (0 mg), 0.25 mg, and 1.25 mg ($n_2 = 109$ patients, 1:4:4 randomization). The final analysis was made based on pooled data from stage 1 and stage 2. The MCP step showed statistical significance of the Emax model ($P = 0.0001$) and the sigmoid Emax model ($P = 0.0115$). The Mod step based on the Emax model estimated the target dose range ED_{70}–ED_{90} as 0.96–3.7 mg. In all, a 2-stage adaptive design successfully achieved the trial objectives and the study allowed further development of siponimod, with a judicious selection of doses for confirmatory phase III testing.

3 Optimal Designs for Phase III Randomized Comparative Trials

Randomized comparative trials test the difference between two or more treatment groups. The design considerations include statistical power and other criteria, such as total sample size, study cost, number of treatment failures, etc. In this section, we give an overview of optimal allocation designs for two-arm and multi-arm comparative studies, and randomization procedures to implement these optimal designs in practice.

3.1 Two-Arm Trials

Consider a randomized clinical trial comparing the effects of two treatments, experimental (E) and control (C) with respect to some primary outcome variable (response). Let $Y_j(k)$ be the jth patient's potential response when the patient is randomized to the treatment group k, where $k=E, C$ and, assume that $Y_j(E) \sim N\left(\mu_E, \sigma^2\right)$ and $Y_j(C) \sim N\left(\mu_C, \sigma^2\right)$, where $\mu_E, \mu_C \in \mathbb{R}$ and $\sigma^2 > 0$. If $\Delta = \mu_E - \mu_C$, we wish to test the null hypothesis $H_0 : \Delta = 0$ versus the alternative hypothesis $H_1 : \Delta > 0$. For simplicity, we assume that σ^2 is known and without loss of generality, assume that it is unity. Then a one-sided α-sized z-test for testing treatment difference would reject H_0 if $Z = \frac{\hat{\Delta}}{\sqrt{\frac{1}{n_E}+\frac{1}{n_C}}} > z_\alpha$, where $\hat{\Delta}$ is the sample mean difference and z_α is the $100(1-\alpha)$th percentile of the standard normal distribution. If $\rho \in (0,1)$ is the allocation proportion to treatment group E and the total sample is n, we have $n_E \approx n\rho$ patients in treatment group E and $n_C \approx n(1-\rho)$ patients in treatment group C. A basic, yet important question is: What allocation maximizes the power of the test? In our case, the power of the z-test (for given values of α, Δ, ρ, and n) is $\boldsymbol{\Phi}\left(\Delta\sqrt{n\rho(1-\rho)} - z_\alpha\right)$, where $\boldsymbol{\Phi}(\cdot)$ is the standard normal distribution function. Clearly, for fixed $\Delta > 0$, n, and α, the power is maximized when $\rho = 1/2$, i.e., equal allocation with $n/2$ patients in each group E and C. This results also holds if σ^2 is unknown and the z-test is replaced by a two-sample t-test [6].

If the outcome variance is heterogeneous across the groups, equal allocation may not be optimal for maximizing power. This can be seen using an elegant general approach for optimizing a two-arm trial with binary responses as described in chapter 17 of Jennison and Turnbull [55]. Let p_k be the success rate for treatment k and $q_k = 1 - p_k$, $k = E, C$. If we have a sufficiently large sample size for a normal approximation to a binomial to hold, the Wald test statistic $W = \left(\hat{p}_E - \hat{p}_C\right)/\sqrt{\frac{\hat{p}_E\hat{q}_E}{n_E} + \frac{\hat{p}_C\hat{q}_C}{n_C}}$ can be used to test $H_0 : p_E = p_C$ versus $H_1 : p_E > p_C$. Here, $\hat{p}_k$ is a consistent estimate of p_k, $\hat{q}_k = 1 - \hat{p}_k$, $k = E, C$, and n_E and n_C are the two group sample sizes. The asymptotic power of the Wald test is a decreasing function of $\frac{p_E q_E}{n_E} + \frac{p_C q_C}{n_C}$, which is the variance of $\hat{p}_E - \hat{p}_C$ under the alternative hypothesis. The criterion we wish to minimize is $w_E n_E + w_C n_C$ for some suitably chosen weights $w_E, w_C > 0$ subject to the constraint that $\frac{p_E q_E}{n_E} + \frac{p_C q_C}{n_C} \leq L$ for some small constant $L > 0$. In fact, L should be set $\leq 1/2$, since for any $n_E \geq 1$ and

$n_C \geq 1$, one has $\frac{p_E q_E}{n_E} + \frac{p_C q_C}{n_C} \leq \frac{1}{4n_E} + \frac{1}{4n_C} \leq \frac{1}{2}$. A smaller value of L implies greater restriction on the asymptotic variance, and hence greater desired precision, which can be achieved by increasing the total sample size $n = n_E + n_C$. However, the optimal allocation proportion $\rho^* = \frac{n_E}{n_E + n_C}$ does not depend on L. Using Lagrange multiplier minimization, one can easily find that ρ^* in this case is given by

$$\rho^* = \frac{\sqrt{p_E q_E / w_E}}{\sqrt{p_E q_E / w_E} + \sqrt{p_C q_C / w_C}}. \tag{10}$$

If we set $w_E = w_C \equiv 1$, we are minimizing the total sample size of the study, in which case the optimal solution, $\rho_N^* = \sqrt{p_E q_E} / \left(\sqrt{p_E q_E} + \sqrt{p_C q_C} \right)$, is the *Neyman* allocation. Note that $\rho_N^* = 1/2$ if either $p_E = p_C$ or $p_E = 1 - p_C$. One limitation of Neyman allocation is that it is skewed to a less successful treatment arm if $p_E + p_C > 1$. If we set $w_E = q_E$ and $w_C = q_C$, then we are minimizing the total expected number of treatment failures in the study, and the optimal solution, given by Rosenberger et al. [87] is $\rho_{\text{RSIHR}}^* = \sqrt{p_E} / \left(\sqrt{p_E} + \sqrt{p_C} \right)$, which is always skewed in favor of a more successful treatment group.

The described approach was extended by many authors in a number of ways: for two-arm trials with binary responses and different study objectives [41, 127], trials with normal outcomes [14, 129], trials with survival outcomes [130], etc.

Another useful approach for obtaining trade-off between statistical inference and treatment goals in the binary response case was proposed in Baldi Antognini and Giovagnoli [7]. It is based on optimizing a weighted criterion

$$\Psi_\omega(\rho) = \omega \left\{ \frac{\Psi_1(\rho)}{\Psi_1^*} \right\} + (1 - \omega) \left\{ \frac{\Psi_2(\rho)}{\Psi_2^*} \right\},$$

where $\omega \in [0, 1]$ is a user-specified weight that determines trade-off between inference and treatment (e.g., $\omega = 0.8 \Rightarrow 80\%$ emphasis on inference and 20% on ethics), $\Psi_1(\rho) = \frac{p_E q_E}{\rho} + \frac{p_C q_C}{1-\rho}$ (scaled variance of the estimated difference in proportions, inferential criterion), $\Psi_1^* = \left\{ \sqrt{p_E q_E} + \sqrt{p_C q_C} \right\}^2$ (minimum value of $\Psi_1(\rho)$ for $\rho \in (0, 1)$), $\Psi_2(\rho) = q_E \rho + q_C (1 - \rho)$ (expected proportion of treatment failures, ethical criterion), and $\Psi_2^* = \min\{q_E, q_C\}$ (minimum value of $\Psi_2(\rho)$ for $\rho \in (0, 1)$). The optimal allocation, $\rho_\omega^* = \arg\min_\rho \Psi_\omega(\rho)$ is the unique solution in (0,1) of the following equation (to be solved numerically for given ω, p_E, p_C, q_E, q_C): $\frac{\omega}{1-\omega} \times \frac{p_E - p_C}{\min(q_E, q_C)} \times \left\{ \sqrt{\frac{p_C q_C}{p_E q_E}} + 1 \right\}^2 = \frac{\left\{ \frac{p_C q_C}{p_E q_E} - 1 \right\} \rho^2 + 2\rho - 1}{\{\rho(1-\rho)\}^2}$.

An excellent treatise of various multi-objective optimal allocation designs for two-arm comparative trials can be found in chapter 5 of Baldi Antognini and Giovagnoli [8]. Since many of these designs depend on model parameters, response-adaptive randomization (RAR) is needed to implement them in practice. We shall discuss RAR in Sect. 3.3.

Another consideration in optimizing treatment allocation ratio is the presence of important baseline covariates such as age, gender, disease severity, genetic signature,

etc., that are known to be correlated with the clinical outcome. For the classical linear regression model with constant variance and additive effects due to treatment and covariates, balance in both treatment totals and across the covariates is optimal for statistical inference [8]. However, for more complex settings, such as heteroscedastic and nonlinear models with possibly treatment-by-covariate interactions, balanced allocation may no longer be optimal [88]. Recently, some optimal designs for such complex scenarios have been proposed [1, 118]. These designs may be particularly useful in trials of precision medicine with potentially differential treatment effects across patient subgroups.

3.2 Multi-arm Trials

Randomized trials comparing the effects of several treatments in a single study are very common in clinical research. The treatment arms may be different doses of a drug, different drug combinations, or different intervention strategies. Designs for multi-arms trials require careful considerations, and the process of planning such trials is more complex than that in the two-arm case. As an example, suppose there are $K \geq 1$ experimental treatment groups $E_1, \ldots, E_K$ and a control group C, the outcome is binary (success or failure) and we want to assess performances using differences in their proportions. Suppose $Y_j(k)$, the jth patient's potential response to treatment k, has Bernoulli distribution with $E\big(Y_j(k)\big) = p_k$ and $\text{var}\big(Y_j(k)\big) = p_k q_k$, $k = 0, 1, \ldots, K$ (here $k = 0$ is the index for the control and $k = 1, \ldots, K$ is for the experimental treatments). The $(K+1)$-vector of true success probabilities is $\boldsymbol{p}' = \big(p_0, p_1, \ldots, p_K\big)$ and the K-vector of treatment contrasts is $\boldsymbol{p}_c' = \big(p_1 - p_0, \ldots, p_K - p_0\big) = \boldsymbol{A}'\boldsymbol{p}$, where $\boldsymbol{A}'$ is the appropriately chosen $K \times (K+1)$ contrast matrix. An allocation vector is $\boldsymbol{\rho}' = \big(\rho_0, \rho_1, \ldots, \rho_K\big)$, $\rho_k \in (0,1)$, $\sum_{k=0}^{K} \rho_k = 1$. For a total sample size n, $n_k \approx n\rho_k$ subjects would be randomized to group $k = 0, 1, \ldots, K$. The optimal value of $\boldsymbol{\rho}$ is to be determined based on the study goals. The MLE $\hat{\boldsymbol{p}}' = \big(\hat{p}_0, \hat{p}_1, \ldots, \hat{p}_K\big)$ has asymptotic variance–covariance matrix $\text{var}(\hat{\boldsymbol{p}}) = \text{diag}\left\{\frac{p_0 q_0}{n_0}, \frac{p_1 q_1}{n_1}, \ldots, \frac{p_K q_K}{n_K}\right\}$ and that for $\hat{\boldsymbol{p}}_c = \boldsymbol{A}'\hat{\boldsymbol{p}}$ is $\text{var}\big(\hat{\boldsymbol{p}}_c\big) = \text{diag}\left\{\frac{p_1 q_1}{n_1}, \ldots, \frac{p_K q_K}{n_K}\right\} + \frac{p_0 q_0}{n_0}\mathbf{1}\mathbf{1}'$, where $\mathbf{1}$ is the K-vector of ones.

For testing treatment difference, one may consider, for instance, a global hypothesis $H_0 : \boldsymbol{p}_c = \mathbf{0}$ versus $H_1 : \boldsymbol{p}_c \neq \mathbf{0}$. An appropriate test statistic is the Wald test $W_n = \hat{\boldsymbol{p}}_c'\big(\widehat{\text{var}}\big(\hat{\boldsymbol{p}}_c\big)\big)^{-1}\hat{\boldsymbol{p}}_c$, where $\widehat{\text{var}}\big(\hat{\boldsymbol{p}}_c\big)$ is a consistent estimator of $\text{var}\big(\hat{\boldsymbol{p}}_c\big)$. Assuming that n is sufficiently large, W_n follows asymptotically chi-squared distribution with $K-1$ degrees of freedom under H_0. Under H_1, the distribution is non-central chi-squared with $K-1$ degrees of freedom and the non-centrality parameter $\phi(\boldsymbol{n}) = \boldsymbol{p}_c'\big(\text{var}\big(\hat{\boldsymbol{p}}_c\big)\big)^{-1}\boldsymbol{p}_c = \sum_{i=1}^{K} \frac{n_i}{p_i q_i}\big(p_i - p_0\big)^2 - \left(\sum_{i=1}^{K} \frac{n_i}{p_i q_i}\big(p_i - p_0\big)\right)^2 \Big/ \sum_{i=0}^{K} \frac{n_i}{p_i q_i}$, which is a concave function of $\boldsymbol{n}' = \big(n_0, n_1, \ldots, n_K\big)$ with $\nabla\phi \geq 0$ [115].

One important observation is that there is a fundamental difference between the goals of estimation and testing, which will impact the choice of an optimal allocation design [96]. Efficient estimation of $\boldsymbol{p}_c$ calls for minimization of some function

of $\mathrm{var}(\hat{\boldsymbol{p}}_c)$ (e.g., determinant, trace, maximum eigenvalue, etc.), whereas maximization of power for testing $H_0 : \boldsymbol{p}_c = \mathbf{0}$ calls for maximization of the non-centrality parameter $\boldsymbol{p}_c'\left(\mathrm{var}(\hat{\boldsymbol{p}}_c)\right)^{-1}\boldsymbol{p}_c$. The resulting optimal allocations are, in general, quite different [109].

Second, even if an investigator has decided on hypothesis testing as the primary tool for inference, he or she may be interested in different kinds of research hypotheses (e.g., testing each experimental treatment vs. control rather than testing a global hypothesis of homogeneity); may prefer using other metrics for treatment difference (e.g., relative risk or odds ratio); and may wish to use different test statistics (other than Wald test). All these aspects will impact optimization and the structure of optimal allocation proportions [4, 10, 65, 96, 110].

Third, in addition to power maximization, some other objectives may be deemed relevant. For instance, in trials with grave outcomes, there is a strong ethical requirement to minimize the total number of treatment failures in the trial while maintaining sufficient power of the test. A constrained optimization approach for deriving optimal allocations that provide trade-off between statistical (power) and ethical goals for a multi-arm binary outcome trial was originally proposed by Tymofyeyev et al. [115], and subsequently extended by several authors in various contexts [9, 15, 54, 111, 131]. Baldi Antognini and Giovagnoli [8] also described in chapter 5 of their monograph a compound optimality approach that provides a trade-off among selected objectives by optimizing a weighted combination of the chosen criteria.

As in the two-arm case, optimal allocation designs for multi-arm trials frequently depend on the true values of model parameters and necessitate the use of response-adaptive randomization.

3.3 Response-Adaptive Randomization

Consider a clinical trial with $K \geq 2$ treatment groups, for which the vector of target allocation proportions has been derived according to some optimality criteria as $\boldsymbol{\rho}' = \left(\rho_1(\boldsymbol{\theta}), \ldots, \rho_K(\boldsymbol{\theta})\right)$, where $\rho_i(\boldsymbol{\theta}) \in (0,1)$, $\sum_{i=1}^{K} \rho_i(\boldsymbol{\theta}) = 1$ and $\boldsymbol{\theta}$ is the vector of model parameters (including treatment effects and variances). The total sample size (n) is fixed and predetermined. Eligible subjects enter the trial sequentially and must be randomized such that $N_i(n) \approx n\rho_i(\boldsymbol{\theta})$ subjects are assigned to the ith treatment group. If $\boldsymbol{\theta}$ were known, one could easily accomplish this by randomizing the jth subject to the ith treatment with probability $P_i(j) = \rho_i(\boldsymbol{\theta})$, $k = 1, \ldots, K$, $j \geq 1$. However, in practice $\boldsymbol{\theta}$ is unknown, which motivates introduction of response-adaptive randomization (RAR). The idea is to use accumulating data to sequentially estimate $\boldsymbol{\theta}$ and modify treatment randomization probabilities to direct the design to the desired target allocation [49].

Conceptually, RAR is implemented as follows: initial $m_0 < n$ patients are randomized according to some fixed (non-adaptive) procedure; e.g., using equal randomization. Assume that outcomes are ascertained without delay after randomization. After m assignments ($m \geq n_0$), one can obtain an estimate $\hat{\boldsymbol{\theta}}_m$, update the target allocation vector $\hat{\boldsymbol{\rho}}_m = \left(\rho_1(\hat{\boldsymbol{\theta}}_m), \ldots, \rho_K(\hat{\boldsymbol{\theta}}_m)\right)$ and randomize the next, $(m+1)$st

patient according to $\hat{\boldsymbol{\rho}}_m$. Let $\boldsymbol{N}(m) = \big(N_1(m), \ldots, N_K(m)\big)$ be the treatment group sample sizes after m assignments. Then, a general RAR procedure can be described by specifying randomization probabilities for the $(m+1)$st subject using information from the previous m subjects as follows:

$$\begin{aligned} P_i(m+1) &= \Pr\{(m+1)\text{st subject is assigned to the } i\text{th group}|\text{data}\} \\ &= \phi_i\left(\hat{\boldsymbol{\rho}}_m, \frac{\boldsymbol{N}(m)}{m}\right), \quad i = 1, \ldots, K, \end{aligned}$$

where the functions $\phi_i = \phi_i\left(\hat{\boldsymbol{\rho}}_m, \frac{N(m)}{m}\right)$, $0 < \phi_i < 1$, $\sum_{i=1}^{K} \phi_i = 1$ must be chosen judiciously.

For instance, one practical choice is the doubly adaptive biased coin design (DBCD) procedure of Hu and Zhang [50] obtained by setting $\phi_i\left(\hat{\boldsymbol{\rho}}_m, \frac{N(m)}{m}\right) \propto \rho_i\big(\hat{\boldsymbol{\theta}}_m\big) \cdot \left(\frac{\rho_i(\hat{\boldsymbol{\theta}}_m)}{N_1(m)/m}\right)^{\gamma}$, where $\gamma \geq 0$ is a user-defined parameter that controls the degree of randomness of the design. In practice, $\gamma = 2$ is recommended [86]. The DBCD has well-established theoretical properties: under widely satisfied conditions, the sample allocation proportions $\frac{N(m)}{m}$ converge to the target allocation $\boldsymbol{\rho}$ and asymptotically follow a multivariate normal distribution with a known covariance structure. Also, the MLE $\hat{\boldsymbol{\theta}}_m$ is strongly consistent for $\boldsymbol{\theta}$ and is asymptotically normally distributed. This implies that, theoretically, standard large sample estimators and tests can be used for statistical inference following DBCD, including other RAR procedures for which allocation proportions converge to the predetermined target proportions $0 < \rho_i(\boldsymbol{\theta}) < 1$, $i = 1, \ldots, K$. These results do not apply, however, to some RAR designs, such as Thompson's-type Bayesian RAR designs [113, 114] or randomly reinforced urn models [42] that target an extreme limiting allocation by skewing randomization to the superior treatment group when it exists. In practice, simulations should be used to evaluate performance of RAR in finite samples, and assess robustness of these procedures under various standard and worst-case scenarios, including cases when some underlying assumptions may be violated [89].

An alternative to the likelihood-based inference is the randomization-based inference [81, 90]. The idea is to treat responses as deterministic, and calculate the test statistic over the set of all possible randomization sequences induced by a given RAR procedure. The randomization-based P-value is then the sum of probabilities of randomization sequences that yield the value of the test at least as extreme as the one observed. Monte Carlo simulation can be used to obtain a consistent estimate of the randomization-based P-value. In the context of RAR, randomization-based tests have been found to be useful and more robust than the likelihood-based tests [43, 100, 116].

RAR has a long history in biostatistics literature. It includes a very broad family of randomization designs which cannot be comprehensively covered in any single paper. The readers are referred to recent monographs on this topic by Hu and Rosenberger [49], Atkinson and Biswas [2] and Baldi Antognini and Giovagnoli [8].

4 Optimal Designs for Population PK Experiments

In recent years, there has been an increasing interest in applying *pharmacometrics*—a branch of science that combines mathematical, statistical, and computational methods with pharmacology—to tackle various problems in drug development. Pharmacometrics can be used to *characterize, understand and predict a drug's pharmacokinetics (PK), pharmacodynamics (PD) and biomarker-outcome behavior* [33]. The PK part answers the question "what the body does to the drug," whereas the PD part explains "what the drug does to the body" [71]. As noted in [76], pharmacometric models may be *descriptive* (to characterize existing data) and *predictive* (to allow testing situations when data are not available via *model-based simulation*). The latter approach holds the promise to increase the number of successful clinical trials and to improve the efficiency of the informed decision making in drug development [44].

In this section, we discuss three important components of pharmacometrics: population modeling of PK/PD data, optimal designs for the population studies, and model-based adaptive optimal designs relevant to this setting. We also highlight useful software packages and some real-life applications.

4.1 Population Modeling of PK/PD Data

The *population modeling* approach was first proposed by Sheiner et al. [98], and it has been increasingly used in modern model-based drug development. This approach allows modeling of potentially sparse PK/PD data from individual subjects to allow estimation of the population parameters of interest while accounting for inter-individual variability of the observed responses. A statistical approach to handle population modeling is based on nonlinear mixed-effects models (NLMEM). With such an approach, a simple continuous response model can be defined as

$$\boldsymbol{y}_i = \boldsymbol{f}\left(\boldsymbol{t}_i, \boldsymbol{x}_i, \boldsymbol{\theta}, \boldsymbol{\eta}_i\right) + \boldsymbol{\varepsilon}_i, i = 1, \ldots, n, \tag{11}$$

where $\boldsymbol{f}(\cdot)$ is some nonlinear (vector) function, and for the ith individual, we have $\boldsymbol{y}_i$ is a vector of responses (e.g., some efficacy measurement), $\boldsymbol{t}_i$ is a vector of sampling time points, $\boldsymbol{x}_i$ is a vector of administered dose levels, $\boldsymbol{\theta}$ is a vector of typical parameter values, $\boldsymbol{\eta}_i \sim MVN(0, \boldsymbol{\Omega})$ is a vector of inter-individual variabilities (IIV's), and $\boldsymbol{\varepsilon}_i \sim MVN(0, \boldsymbol{\Sigma})$ is a vector of measurement errors. In a more general setting, model (11) may also include important individual covariates, and the residual term $\boldsymbol{\varepsilon}_i$ may be modeled as a function of other components of the model (11).

The first problem that arises in population modeling is to estimate the parameters $(\boldsymbol{\theta}, \boldsymbol{\Omega}, \boldsymbol{\Sigma})$. The maximum likelihood estimate $(\boldsymbol{\theta}^*, \boldsymbol{\Omega}^*, \boldsymbol{\Sigma}^*)$ is found as

$$(\boldsymbol{\theta}^*, \boldsymbol{\Omega}^*, \boldsymbol{\Sigma}^*) = \arg\max_{\boldsymbol{\theta}, \boldsymbol{\Omega}, \boldsymbol{\Sigma}} \sum_{i=1}^{n} \log\left(\iint \ell_i\left(\boldsymbol{y}_i, \boldsymbol{\eta} | \boldsymbol{\theta}, \boldsymbol{\Sigma}\right) \cdot p(\boldsymbol{\eta} | \boldsymbol{\Omega}) \mathrm{d}\boldsymbol{\eta}\right), \tag{12}$$

where $\ell_i(\cdot)$ is the individual likelihood, and $p(\cdot)$ is the probability density of the IIV's, given the population parameters $(\boldsymbol{\Omega})$. The integral in Eq. (12) is the marginal likelihood of ith individual which explains the individual's contribution to the

population likelihood. There is no closed-form solution to the problem in Eq. (12), so numerical methods are used. The dimensionality of the problem can be very large. For instance, if there are p typical values and p random effects, then we have to estimate, in general, $p + p(p+1)/2$ parameters. If, additionally, there are n individuals, then n diagonal elements of $\mathbf{\Sigma}$ have to be estimated as well. The most challenging part of finding a solution to Eq. (12) is an approximation of the marginal likelihood, which is also essential for calculating the FIM to find the population optimal designs. Below are some numerical methods suitable for this purpose, with more details in [117].

- *First-order (FO)* method [12] was originally implemented in NONMEM [11], the most popular commercial software to deal with NLMEM. The method linearizes the model around the means of individual parameters, $\boldsymbol{\eta}_i = 0$, and the marginal likelihood is then calculated assuming the observed data are normally distributed with the mean and the variance of the linearized model.
- *First-order conditional estimation (FOCE)* method was first introduced in [61], and implemented in NONMEM. First, the algorithm searches for the solution to the optimization problem: $\boldsymbol{\eta}^* = \arg\max_{\boldsymbol{\eta}} \ell_i\left(\boldsymbol{y}_i, \boldsymbol{\eta}|\boldsymbol{\theta}, \mathbf{\Sigma}\right) \cdot p(\boldsymbol{\eta}|\mathbf{\Omega})$. Then, the linearization around $\boldsymbol{\eta}^*$ is done, and the marginal likelihood is then calculated assuming the observed data are normally distributed with the mean and the variance which are linear in $\boldsymbol{\eta}^*$.
- *Laplace integration* method is more computationally intensive compared to the FO and FOCE approaches due to calculations of the Hessian of the individual likelihoods, but the approximated result is more accurate.
- *Stochastic and Monte Carlo* methods are based on sampling techniques from probability distribution $p(\boldsymbol{\eta}|\mathbf{\Omega})$. These methods are rather slow to converge. NONMEM provides several stochastic algorithms to calculate likelihood, and they include Monte Carlo importance sampling method [93] and the stochastic approximation of expectation minimization (SAEM) algorithm [57]. The latter algorithm was originally developed and implemented in a commercial software Monolix (http://lixoft.com/products/monolix/), which is another popular software for NLMEM with a free license for use in academia. Today, it is the only algorithm used for MLE in Monolix. The method simulates individual data using a Markov Chain, then, at the E-step, the stochastic approach approximates the likelihood, and, at the M-step, the parameters are updated to maximize the likelihood. The procedure is repeated iteratively until some user-specified convergence criterion or criteria are met.

It is worth mentioning some other useful software tools for NLMEM, such as SAS PROC NLMIXED; Perl-speaks-NONMEM (PsN) [59, 60]; open source R packages such as nlme, nlmer, saemix, brms [105], nlmixr (http://nlmixr.org/wp/), mrgsolve (https://mrgsolve.github.io/), PopED [73], etc. A comparison of five different software tools (PFIM, PkStaMp, PopDes, PopED, and POPT) for design evaluation in population PK/PD studies was done in [71].

4.2 Population Optimal Designs

Population PK/PD experiments are quite complex and require careful design considerations. Optimal design techniques can be applied to enhance planning of a PK/PD experiment under a given nonlinear mixed-effects model. Elements to be optimized may include sampling times, sampling frequency, sampling cost, dose, etc. Population optimal designs can be useful, for instance, in the following cases:

- When a study sample size is small (as in many PK studies), population optimal designs can help understand a typical pattern of PK over time and the uncertainty in the observations.
- Population optimal designs can reduce the number of sampling times, which may translate into savings in the study cost.
- Population optimal designs may help improve existing therapies or diagnostics, and provide recommendations for efficient dose regimen(s).
- In pediatric studies, population optimal designs may help extrapolate results from the trials in adults onto children population(s).

Most of the population optimal designs are obtained under local optimality conditions. Data from previous studies can be used to build initial designs, and then both the model and the design can be calibrated during the study via simulations.

The most common application of optimal designs for population PK studies is optimization of the sampling schedule. In Fedorov and Leonov [40], the authors considered an example of a population D-optimal design to estimate the parameters of a two-compartment PK model with bolus input x_0 (cf. Equation (7.9) in [40], §7.2.2, p. 192). The model describes the amount of the drug at a given time in the central and peripheral compartments. In terms of Eq. (11), the parameters are: θ_1 (plasma clearance), θ_2 (volume of distribution), $\Omega_{11} = \mathrm{var}(\eta_{i1})$, $\Omega_{22} = \mathrm{var}(\eta_{i2})$, $\Omega_{12} = \mathrm{cov}(\eta_{i1}, \eta_{i2})$ (η_{i1} and η_{i2} are the ith individual's effects that explain IIV in clearance and volume, respectively), $\Sigma_1 = \mathrm{var}(\varepsilon_{i1})$, and $\Sigma_2 = \mathrm{var}(\varepsilon_{i2})$ (ϵ_{i1} and ϵ_{i2} are residual error terms). The initial non-optimal study design consisted of 16 sampling points. Several D-optimal designs with different numbers of sampling points (from 5 to 8) were obtained and then compared to the original 16-point design. The 8-point design was, overall, 16% less informative; however, the increase in the number of samples from 8 to 16 did not affect the precision of all the estimates of the parameters $\theta_1, \theta_2, \Omega_{11}, \Omega_{22}, \Omega_{12}$, and Σ_1, except Σ_2. In addition, when the total cost of collecting and analyzing samples was taken into consideration, the cost-efficiency of a design with a fewer number of sampling time points became apparent. The optimal design was found to be

$$\xi_{\mathrm{D}}^{*} = \left\{ \begin{array}{ll} x_1^{*} = (5\mathrm{min}; 15\mathrm{min}; 144\mathrm{h}) & x_2^{*} = (5\mathrm{min}; 15\mathrm{min}; 30\mathrm{min}; 84\mathrm{h}; 144\mathrm{h}) \\ w_1^{*} = 0.1 & w_2^{*} = 0.9 \end{array} \right\},$$

which means that 10% of the subjects should be randomized to a 3-sample sequence x_1^{*} and 90% of patients should be randomized to a 5-sample sequence x_2^{*}.

The above method is not limited to finding optimal sampling times. Other design variables, such as dose, can be taken into account as well. For instance, Nyberg et al. [72] investigated optimization of doses on a continuous scale, and sampling time points for three PK/PD models: (1) intravenous PK, E-max PD model; (2) oral PK, E-max PD model; and (3) intravenous PK with Michaelis–Menten elimination. Two approaches to optimization were compared: simultaneous (times and doses are optimized at the same time), and sequential (doses are optimized for given fixed times, and then times are optimized given the optimal doses from the first stage; or, alternatively, times are optimized in the first stage, and then optimization is done for doses). Both D-optimal (locally optimal) and ED-optimal (globally optimal) designs were investigated. Another important problem is optimization of the treatment period length in a study of disease progression [48]. ED-optimal design applied to three disease progression models with different drug effects—a symptomatic effect, a disease-modifying effect, and a combination of both effects—demonstrated high power in discriminating between the models.

More complex experiments may call for optimization of several objectives simultaneously. For example, Hennig et al. [47] found D-, Ds-, and EDs-optimal designs to improve a pre-transplant dose-finding of an immunosuppressant drug ciclosporin. The following variables were optimized simultaneously, using a published ciclosporin population PK model as prior information: the sampling times, the dose of ciclosporin, the timing of the second dose, the infusion duration, and the administration order. The original design was reduced from 22 to 6 samples per patient and both doses (intravenous oral) were administered within 8 hours. The loss in efficiency of the optimal design with reduced samples compared to the original rich design was found to be minimal.

Another instructive example is Silber et al. [99], where the authors considered optimization of the intravenous glucose tolerance test in patients with type 2 diabetes. It was found that the Ds-optimal design, in contrast to the standard design, could improve the insulin modified intravenous glucose tolerance test, with possibly fewer number of samples. Optimization of sampling times resulted in the largest improvement, followed by the insulin dose. The reduction in the total sample time resulted only in a minor loss in efficiency. The predicted uncertainty of parameter estimates was low in all tested cases, despite the reduction in the number of samples per subject. All computations were done using PopED [73].

4.3 Adaptive Optimal Designs

Model-based adaptive optimal designs (MBAOD) for population experiments are novel and there are many open research problems on this topic. These designs essentially attempt to overcome potential non-robustness to changes in the parameter values of locally optimal designs. One example of a MBAOD can be found in Strömberg [106], where the adaptive design with the FDA stopping criteria was applied in a bridging study from adults to children. It was shown that the MBAOD requires fewer children to fulfill the precision criteria than the sample size obtained from the traditional estimation methodologies. The power for a non-adaptive optimal

design was lower than the required target value of 80%. Another interesting example is described in Strömberg and Hooker [107], where the authors compared robust (global) optimality with local optimality for the MBAOD. It was shown that optimizing design by using global optimality criteria is more flexible, and that the MBAOD may be less sensitive to mis-specification in the prior information available at the design stage. The MBAOD methodology was applied to a simulated PK/PD study with a concentration from a one-compartment first-order absorption PK model driving the population effect response in a sigmoidal Emax PD model. A stopping criterion was introduced to obtain an accurate effect prediction using MBAOD based on minimizing the expected uncertainty in the effect response of the typical individual. Simulations showed that by using a robust optimality criterion in MBAODs, one could reduce the number of required adaptations and improve the practicality of adaptive trials using optimal design.

The R package MBAOD (https://github.com/andrewhooker/MBAOD) provides tools for simulating clinical or pre-clinical trials based on predefined adaptation and optimization rules. This package can be used to plan and evaluate the predicted effectiveness of an upcoming trial, and it can be also used to optimize any specific cohort of an ongoing study.

5 Discussion

In this paper, we gave an overview of optimal designs for various problems from clinical drug development. Overall, optimal designs serve at least two important purposes. First and foremost, they provide benchmarks for judging alternatives. If a simple heuristic procedure is shown to be robust and nearly as efficient as the optimal one, its use may be well justified in a given trial. However, if a simple procedure exhibits high loss in efficiency, then alternatives should be considered. Second, adaptive designs using either stage-wise or sequential calibration of the model can be constructed to approximate optimal designs, and potentially achieve study goals with a reduced sample size. However, implementation of such adaptive optimal designs requires time, careful planning, and frequently extra resources to properly implement interim analyses and design adaptations. Some of these optimal designs are computationally challenging to find and as noted by Yang et al. [125] and Cheng and Yang [21], developing novel and efficient algorithms to generate them in practice is an important contribution to the field.

Optimal designs for phase I dose–toxicity studies have been well researched in the literature. Statistical software (web-based) for constructing optimal designs for various nonlinear models, including the logistic model (4) is available [51, 121]. Another class of optimal designs is the seamless phase I/II dose-finding trials where efficacy and toxicity are considered simultaneously. Optimization of such trials can help identify doses with favorable benefit/risk ratio more efficiently than in separate dose–toxicity and dose–efficacy studies. Adaptive optimal designs for phase I and phase I/II studies have been developed, but are rarely used in practice, because they are unlikely to receive approval by Institutional Review Boards.

Phase II clinical trials commonly investigate dose–response relationships using multiple comparisons (MCP), modeling techniques (Mod), or, most recently, using a combination approach MCP-Mod [18]. Optimal designs for such studies have been developed to address various research questions, such as testing the presence of dose–response, estimation of a minimum efficacious dose, etc. Adaptive optimal designs can be cast in two or more stages and they have been developed and implemented in practice [16, 17]. Since most phase II studies are randomized, controlled, parallel-group designs, careful choice of a randomization procedure that adjusts optimal allocation ratio at interim analyses is essential [95].

Phase III trials are randomized and adequately well-controlled studies designed to test specific clinical research hypotheses. Such designs typically use 1:1 randomization in the two-arm case, but recently, some statistical research has been done to develop optimal allocation designs for randomized multi-arm clinical trials with multiple objectives. These objectives may be to minimize the expected number of treatment failures while maintaining power. Such optimal designs can be implemented using RAR with established properties. Despite that optimal RAR designs can potentially balance, to some degree, the competing goals of experimentation and treatment while maintaining integrity and validity of the trial results, they are very rarely used in phase III confirmatory settings.

A novel, interesting and promising research area is optimal designs for population experiments. The applications are numerous, including optimization of PK sampling times, optimization of dosing and frequency of treatment regimens, bridging population studies (e.g., from adults to children), etc. A major challenge is the high dimensionality and complexity in the computation and optimization. Special software packages are available [71]. In the future, we expect continuation of research and applications of optimal designs and model-based adaptive optimal designs for population studies.

Acknowledgements The authors would like to thank three anonymous reviewers for their constructive comments that have helped improve the original version of the manuscript. Wong was partially supported by a grant from the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R01GM107639. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Compliance with Ethical Standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Atkinson A (2015) Optimum designs for two treatments with unequal variances in the presence of covariates. Biometrika 102(2):494–499
2. Atkinson A, Biswas A (2014) Randomised response-adaptive designs in clinical trials. Chapman & Hall/CRC Press, Boca Raton
3. Atkinson A, Donev A, Tobias R (2007) Optimum experimental designs, with SAS. Oxford University Press, New York

4. Azriel D, Feigin PD (2014) Adaptive designs to maximize power in clinical trials with multiple treatments. Seq Anal 31:60–86
5. Babb J, Rogatko A, Zacks S (1998) Cancer phase I clinical trials: efficient dose escalation with overdose control. Stat Med 17:1103–1120
6. Baldi Antognini A (2008) A theoretical analysis of the power of biased coin designs. J Stat Plan Inference 138:1792–1798
7. Baldi Antognini A, Giovagnoli A (2010) Compound optimal allocation for individual and collective ethics in binary clinical trials. Biometrika 97(4):935–946
8. Baldi Antognini A, Giovagnoli A (2015) Adaptive designs for sequential treatment allocation. CRC Press, Boca Raton
9. Baldi Antognini A, Novelli M, Zagoraiou M (2018) Optimal designs for testing hypothesis in multiarm clinical trials. Stat Methods Med Res. https://doi.org/10.1177/0962280218797960
10. Bandyopadhyay U, Bhattacharya R (2018) An optimal three treatment allocation for binary treatment responses. Stat Biopharm Res 10(4):287–300
11. Bauer RJ (2018) NONMEM user guide. Introduction to NONMEM 7.4.3. Gaithersburg, MD
12. Beal SL, Sheiner LB (1982) Estimating population kinetics. Crit Rev Biomed Eng 8(3):195–222
13. Biedermann S, Dette H, Zhu W (2007) Compound optimal designs for percentile estimation in dose–response models with restricted design intervals. J Stat Plan Inference 137:3838–3847
14. Biswas A, Mandal S (2004) Optimal adaptive designs in phase III clinical trials for continuous responses with covariates. In: Di Bucciano A, Lauter H, Wynn HP (eds) mODa7—Advances in model-oriented design and analysis. Physica-Verlag, Heidelberg, pp 51–58
15. Biswas A, Mandal S, Bhattacharya R (2011) Multi-treatment optimal response-adaptive designs for phase III clinical trials. J Korean Stat Soc 40:33–44
16. Bornkamp B, Bretz F, Dette H (2011) Response-adaptive dose-finding under model uncertainty. Ann Appl Stat 5(2B):1611–1631
17. Bornkamp B, Bretz F, Dmitrienko A, Enas G, Gaydos B, Hsu CH, König F, Krams M, Liu Q, Neuenschwander B, Parke T, Pinheiro J, Roy A, Sax R, Shen F (2007) Innovative approaches for designing and analyzing adaptive dose-ranging trials. J Biopharm Stat 17(6):965–995
18. Bretz F, Pinheiro J, Branson M (2005) Combining multiple comparisons and modeling techniques in dose-response studies. Biometrics 61:738–748
19. Chaloner K, Larntz K (1989) Optimal Bayesian design applied to logistic regression experiments. J Stat Plan Inference 21:191–208
20. Chaloner K, Verdinelli I (1995) Bayesian experimental design: a review. Stat Sci 10:273–304
21. Cheng Q, Yang M (2019) On multiple-objective optimal designs. J Stat Plan Inference 200:87–101
22. Chernoff H (1953) Locally optimal designs for estimating parameters. Ann Math Stat 24:586–602
23. Clertant M, O'Quigley J (2018) Semiparametric dose finding methods. J R Stat Soc B 79(5):1487–1508
24. Clyde M, Chaloner K (1996) The equivalence of constrained and weighted designs in multiple objective design problems. J Amer Stat Assoc 91:1236–1244
25. Cook RD, Fedorov V (1995) Constrained optimization of experimental design. Statistics 26:129–178
26. Cook RD, Wong WK (1994) On the equivalence of constrained and compound optimal designs. J Am Stat Assoc 89:687–692
27. Dette H, Bretz F, Pepelyshev A, Pinheiro J (2008) Optimal designs for dose-finding studies. J Am Stat Assoc 103:1225–1237
28. Dette H, Kiss C, Bevanda M (2010) Optimal designs for the emax, log-linear, and exponential models. Biometrika 97(2):513–518
29. Dragalin V, Bornkamp B, Bretz F, Miller F, Padnamabhan SK, Patel N, Perevozskaya I, Pinheiro J, Smith JR (2010) A simulation study to compare new adaptive dose-ranging designs. Stat Biopharm Res 2(4):487–512
30. Dragalin V, Fedorov V (2006) Adaptive designs for dose-finding based on efficacy–toxicity response. J Stat Plan Inference 136:1800–1823
31. Dragalin V, Fedorov V, Wu Y (2008) Two-stage designs for dose-finding that accounts for both efficacy and toxicity. Stat Med 27:5156–5176
32. Durham SD, Flournoy N, Rosenberger WF (1997) A random walk rule for phase I clinical trials. Biometrics 53:745–760
33. Ette EI, Williams PJ (2007) Pharmacometrics: the science of quantitative pharmacology. Wiley, New York

34. European Medicines Agency (1994) ICH E4: Dose-response information to support drug registration. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E4/Step4/E4_Guideline.pdf. Accessed 9 Feb 2019
35. European Medicines Agency (2014) Qualification opinion of MCP-Mod as an efficient statistical methodology for model-based design and analysis of phase II dose finding studies under model uncertainty. https://www.ema.europa.eu/documents/regulatory-procedural-guideline/qualification-opinion-mcp-mod-efficient-statistical-methodology-model-based-design-analysis-phase-ii_en.pdf. Accessed 10 Feb 2019
36. Fackle Fornius E, Nyquist H (2009) Using the canonical design space to obtain c-optimal designs for the quadratic logistic model. Commun Stat Theory Methods 39(1):144–157
37. Fan SK, Chaloner K (2001) Optimal designs for a continuation-ratio model. In: Atkinson AC, Hackl P, Müller WG (eds) mODa 6—advances in model-oriented design and analysis. Physica-Verlag, Heidelberg, pp 77–86
38. Fan SK, Chaloner K (2004) Optimal designs and limiting optimal designs for a trinomial response. J Stat Plan Inference 126:347–360
39. Fedorov V (2010) Optimal experimental design. WIREs Comput Stat 2:581–589
40. Fedorov V, Leonov S (2014) Optimal design for nonlinear response models. CRC Press, Boca Raton
41. Feng C, Hu F (2018) Optimal response-adaptive designs based on efficiency, ethic, and cost. Stat Interface 11:99–107
42. Flournoy N, May C, Secchi P (2012) Asymptotically optimal response-adaptive designs for allocating the best treatment: an overview. Int Stat Rev 80(2):293–305
43. Galbete A, Rosenberger WF (2016) On the use of randomization tests following adaptive designs. J Biopharm Stat 26(3):466–474
44. Gobburu JV (2010) Pharmacometrics 2020. J Clin Pharmacol 50:151–157
45. Haines L, Perevozskaya I, Rosenberger WF (2003) Bayesian optimal designs for phase I clinical trials. Biometrics 59:591–600
46. Heise MA, Myers RH (1996) Optimal designs for bivariate logistic regression. Biometrics 52(2):613–624
47. Hennig S, Nyberg J, Fanta S, Backman JT, Hoppu K, Hooker AC, Karlsson MO (2012) Application of the optimal design approach to improve a pretransplant drug dose finding design for ciclosporin. J Clin Pharmacol 52(3):347–360
48. Hennig S, Nyberg J, Hooker AC, Karlsson MO (2009) Trial treatment length optimization with an emphasis on disease progression studies. J Clin Pharmacol 49(3):323–335
49. Hu F, Rosenberger WF (2006) The theory of response-adaptive randomization in clinical trials. Wiley, New York
50. Hu F, Zhang L-X (2004) Asymptotic properties of doubly adaptive biased coin designs for multi-treatment clinical trials. Ann Stat 32(1):268–301
51. Hu J, Zhu W, Su Y, Wong WK (2010) Controlled optimal design program for the logit dose response model. J Stat Softw 35:6
52. Hyun SW (2013) Optimal designs for a probit model with a quadratic term. Stat Biopharm Res 5(1):18–26
53. Hyun SW (2014) Optimal designs for estimating the ED50 when response functions have a downturn. Stat Biopharm Res 6(1):9–15
54. Jeon Y, Hu F (2010) Optimal adaptive designs for binary response trials with three treatments. Stat Biopharm Res 2:310–318
55. Jennison C, Turnbull B (2000) Group sequential methods with applications to clinical trials. Chapman & Hall/CRC, New York
56. King J, Wong WK (2000) Minimax D-optimal designs for the logistic model. Biometrics 56:1263–1267
57. Kuhn E, Lavielle M (2005) Maximum likelihood estimation in nonlinear mixed effects models. Comput Stat Data Anal 49(4):1020–1038
58. Li G, Majumdar D (2008) D-optimal designs for logistic models with three and four parameters. J Stat Plan Inference 138:1950–1959
59. Lindbom L, Pihlgren P, Johnsson EN (2005) PsN Toolkit—a collection of computer intensive statistical methods for non-linear mixed effects modeling using NONMEM. Comput Methods Programs Biomed 79(3):241–257

60. Lindbom L, Ribbing J, Johnsson EN (2003) Perl-speaks-NONMEM (PsN)—a Perl module for NONMEM related programming. Comput Methods Programs Biomed 75(2):85–94
61. Lindstrom MJ, Bates DM (1990) Nonlinear mixed effects models for repeated measures data. Biometrics 46(3):673–687
62. Liu G, Rosenberger WF (2006) Sequential designs for logistic phase I clinical trials. J Biopharm Stat 16(5):605–621
63. Magnusdottir BT (2013) c-optimal designs for the bivariate Emax model. In: Usiński D, Atkinson AC, Patan M (eds) mODa 10—advances in model-oriented design and analysis. Springer, Berlin, pp 153–161
64. Manukyan Z, Rosenberger WF (2010) D-optimal design for a five-parameter logistic model. In: Giovagnoli A, Atkinson AC, Torsney B (eds) mODa 9—advances in model-oriented design and analysis. Physica-Verlag, Heidelberg, pp 17–24
65. Marschner IC (2007) Optimal design of clinical trials comparing several treatments with a control. Pharm Stat 6:23–33
66. Mats VA, Rosenberger WF, Flournoy N (1998) Restricted optimality for phase I clinical trials. In: Rosenberger WF, Flournoy N, Wong WK (eds) New developments and applications in experimental design, vol 34. Lecture Notes—Monograph Series. Institute of Mathematical Statistics, Hayward, pp 50–61
67. Matthew T, Sinha BK (2001) Optimal designs for binary data under logistic regression. J Stat Plan Inference 93:295–307
68. McLeish DL, Tosh DH (1990) Sequential design in bioassay. Biometrics 46:103–116
69. Miller F, Guilbaud O, Dette H (2007) Optimal designs for estimating the interesting part of a dose–effect curve. J Biopharm Stat 17(6):1097–1115
70. Minkin S (1987) Optimal designs for binary data. J Am Stat Assoc 82:1098–1103
71. Nyberg J, Bazzoli C, Ogungbenro K, Aliev A, Leonov S, Duffull S, Hooker AC, Mentré F (2015) Methods and software tools for design evaluation in population pharmacokinetics-pharmacodynamics studies. Br J Clin Pharmacol 79(1):6–17
72. Nyberg J, Karlsson MO, Hooker AC (2009) Simultaneous optimal experimental design on dose and sample times. J Pharmacokinet Pharmacodyn 36:125–145
73. Nyberg J, Ueckert S, Strömberg EA, Hennig S, Karlsson MO, Hooker AC (2012) PopED: an extended, parallelized, nonlinear mixed effects models optimal design tool. Comput Methods Programs Biomed 108(2):789–805
74. O'Quigley J, Iasonos A, Bornkamp B (eds) (2017) Methods for designing, monitoring, and analyzing dose-finding trials. CRC Press, Boca Raton
75. O'Quigley J, Pepe M, Fisher L (1990) Continual reassessment method: a practical design for phase I clinical studies in cancer. Biometrics 46:33–48
76. Owen JS, Fielder-Kelly J (2014) Introduction to pharmacokinetic/pharmacodynamic analysis with nonlinear mixed effects models. Wiley, New York
77. Padmanabhan SK, Dragalin V (2010) Adaptive Dc-optimal designs for dose finding based on a continuous efficacy endpoint. Biom J 52(6):836–852
78. Padmanabhan SK, Hsuan F, Dragalin V (2010) Adaptive penalized D-optimal designs for dose finding based on continuous efficacy and toxicity. Stat Biopharm Res 2(2):182–198
79. Parker S, Gennings C (2008) Penalized locally optimal experimental designs for nonlinear model. J Agric Biol Environ Stat 13(3):334–354
80. Pronzato L (2010) Penalized optimal adaptive designs for dose finding. J Stat Plan Inference 140:283–296
81. Proschan MA, Dodd DE (2019) Re-randomization tests in clinical trials. Stat Med 38(12):2292–2302
82. Rabie H, Flournoy N (2004) Optimal designs for contingent response models. In: Di Bucciano A, Lauter H, Wynn HP (eds) mODa 7—advances in model-oriented design and analysis. Physica-Verlag, Heidelberg, pp 133–142
83. Rabie H, Flournoy N (2013) Optimal designs for contingent response models with application to toxicity–efficacy studies. J Stat Plan Inference 143:1371–1379
84. Rosenberger WF, Canfield GC, Perevozskaya I, Haines LM, Hausner P (2005) Development of interactive software for Bayesian optimal phase 1 clinical trial design. Drug Inf J 39:89–98
85. Rosenberger WF, Haines LM, Perevozskaya I (2001) Constrained Bayesian optimal designs for phase I clinical trials: continuous dose space. In: Atkinson AC, Hackl P, Müller WG (eds) mODa 6—advances in model-oriented design and analysis. Physica-Verlag, Heidelberg, pp 225–233

86. Rosenberger WF, Hu F (2004) Maximizing power and minimizing treatment failures in clinical trials. Clin Trials 1:141–147
87. Rosenberger WF, Stallard N, Ivanova A, Harper CN, Ricks ML (2001) Optimal adaptive designs for binary response trials. Biometrics 57:909–913
88. Rosenberger WF, Sverdlov O (2008) Handling covariates in the design of clinical trials. Stat Sci 23(3):404–419
89. Rosenberger WF, Sverdlov O, Hu F (2012) Adaptive randomization for clinical trials. J Biopharm Stat 22(4):719–736
90. Rosenberger WF, Uschner D, Wang Y (2019) Randomization: the forgotten component of the randomized clinical trial. Stat Med 38:1–12
91. Roth K (2012) Sequential designs for dose escalation studies in oncology. Commun Stat Simul Comput 41(7):1131–1141
92. Roy A, Ghosal S, Rosenberger WF (2009) Convergence properties of sequential Bayesian D-optimal designs. J Stat Plan Inference 139:425–440
93. Rubinstein RY, Kroese DP (2017) Simulation and the Monte Carlo method. Wiley, New York
94. Ryeznik Y, Sverdlov O, Hooker A (2018) Adaptive optimal designs for dose-finding studies with time-to-event outcomes. AAPS J 20(1):24
95. Ryeznik Y, Sverdlov O, Hooker A (2018) Implementing optimal designs for dose–response studies through adaptive randomization for a small population group. AAPS J 20(5):85
96. Schou IM, Marschner IC (2017) Design of clinical trials involving multiple hypothesis tests with a common control. Biom J 59(4):636–657
97. Selmaj K, Li DK, Hartung HP, Hemmer B, Kappos L, Freedman MS, Stüve O, Riekmann P, Montalban X, Ziemssen T, Auberson LZ, Pohlmann H, Mercier F, Dahlke F, Wallström E (2019) Siponimod for patients with relapsing-remitting multiple sclerosis (BOLD): an adaptive, dose-ranging, randomised, phase 2 study. Lancet Neurol 12(8):756–767
98. Sheiner LB, Rosenberg B, Marathe VV (1977) Estimation of population characteristics of pharmacokinetic parameters from routine clinical data. J Pharmacokinet Biopharm 5:445–479
99. Silber HE, Nyberg J, Hooker AC, Karlsson MO (2009) Optimization of the intravenous glucose tolerance test in T2DM patients using optimal experimental design. J Pharmacokinet Pharmacodyn 36:281–295
100. Simon R, Simon NR (2010) Using randomization tests to preserve type 1 error with response-adaptive and covariate-adaptive randomization. Stat Probab Lett 81(7):767–772
101. Sitter RR (1992) Robust designs for binary data. Biometrics 48(4):1145–1155
102. Sitter RR, Fainaru I (1997) Optimal designs for the logit and probit models for binary data. Can J Stat 25(2):175–190
103. Sitter RR, Forbes BE (1997) Optimal two-stage designs for binary response experiments. Stat Sin 7:941–955
104. Sitter RR, Wu CFJ (1993) Optimal designs for binary response experiments: Fieller, D, and A criteria. Scand J Stat 20(4):329–341
105. Stegmann G, Jacobucci R, Harring JR, Grimm KJ (2018) Nonlinear mixed-effects modeling programs in R. Struct Eq Model Multidiscip J 25(1):160–165
106. Strömberg EA (2016) Applied adaptive optimal design and novel optimization algorithms for practical use. Uppsala University. Retrieved from simulated model based adaptive optimal design using FDA stopping criteria. An adults to children bridging study example: https://www.page-meeting.org/pdf_assets/7437-MBAOD_simulation_ES_2.pdf
107. Strömberg EA, Hooker AC (2017) The effect of using a robust optimality criterion in model based adaptive optimization. J Pharmacokinet Pharmacodyn 44:317–324
108. Sverdlov O, Gao L (2017) Phase I/II dose-finding designs with efficacy and safety endpoints. In: O'Quigley J, Iasonos A, Bornkamp B (eds) Methods for designing, monitoring, and analyzing dose-finding trials. CRC Press, Boca Raton, FL, pp 81–107
109. Sverdlov O, Rosenberger WF (2013) On recent advances in optimal allocation designs in clinical trials. J Stat Theory Pract 7(4):753–773
110. Sverdlov O, Ryeznik Y (2019) Implementing unequal randomization in clinical trials with heterogeneous treatment costs. Stat Med 38:2905–2927
111. Sverdlov O, Tymofyeyev Y, Wong WK (2011) Optimal response-adaptive randomized designs for multi-armed survival trials. Stat Med 30:2890–2910
112. Sverdlov O, Wong WK, Ryeznik Y (2014) Adaptive clinical trial designs for phase I cancer studies. Stat Surv 8:2–44

113. Thall PF, Wathen JK (2007) Practical Bayesian adaptive randomisation in clinical trials. Eur J Cancer 43:859–866
114. Thompson WR (1933) On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. Biometrika 25(3/4):285–294
115. Tymofyeyev Y, Rosenberger WF, Hu F (2007) Implementing optimal allocation in sequential binary response experiments. J Am Stat Assoc 102:224–234
116. Villar SS, Bowden J, Wason J (2018) Response-adaptive designs for binary responses: How to offer patient benefit while being robust to time trends? Pharm Stat 17:182–197
117. Wang Y (2007) Derivation of various NONMEM estimation methods. J Pharmacokinet Pharmacodyn 34(5):575–593
118. Wang Y, Ai M (2016) Optimal designs for multiple treatments with unequal variances. J Stat Plan Inference 171:175–183
119. Warfield J, Roy A (2013) A semiparametric sequential algorithm for estimation of dose–response curve. Seq Anal 32:196–213
120. Wong WK (1999) Recent advances in multiple-objective design strategies. Stat Neerl 53:257–276
121. Wong WK (2013) Web-based tools for finding optimal designs in biomedical studies. Comput Methods Progr Biomed 111:701–710
122. Wong WK, Zhu W (2008) Optimum treatment allocation rules under a variance heterogeneity model. Stat Med 27:4581–4595
123. Woodcock J, LaVange L (2017) Master protocols to study multiple therapies, multiple diseases, or both. New Engl J Med 377:62–70
124. Wu CFJ (1988) Optimal design for percentile estimation of a quantal response curve. In: Dodge Y, Fedorov V, Wynn HP (eds) Optimal design and analysis of experiments. Elsevier Science, North Holland, pp 213–223
125. Yang M, Biedermann S, Tang E (2013) On optimal designs for nonlinear models: a general and efficient algorithm. J Am Stat Assoc 108(504):1411–1420
126. Yang M, Stufken J (2009) Support points of locally optimal designs for nonlinear models with two parameters. Ann Stat 37(1):518–541
127. Yi Y, Yuan Y (2013) An optimal allocation for response-adaptive designs. J Appl Stat 40(9):1996–2008
128. Yuan Y, Nguyen HQ, Thall PF (2016) Bayesian designs for phase I-II clinical trials. CRC Press, Boca Raton
129. Zhang L, Rosenberger WF (2006) Response-adaptive randomization for clinical trials with continuous outcomes. Biometrics 62:562–569
130. Zhang L, Rosenberger WF (2007) Response-adaptive randomization for survival trials: the parametric approach. Appl Stat 56(2):153–165
131. Zhu H, Hu F (2009) Implementing optimal allocation for sequential continuous responses with multiple treatments. J Stat Plan Inference 139:2420–2430
132. Zhu W, Wong WK (2001) Bayesian optimal designs for estimating a set of symmetrical quantiles. Stat Med 20:123–137

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.