technische universität dortmund

fakultät statistik

# Optimal design theory of dose-response experiments in toxicology

Dissertation
in Fulfillment of the Requirements for the Degree of
*Doktor der Naturwissenschaften*

Submitted to the
Department of Statistics
of the
TU Dortmund University

by
Leonie Schürmeyer
on
April 15th, 2025

Referees:
JProf. Dr. Kirsten Schorning
Prof. Dr. Jan Georg Hengstler
Prof. Dr. Jörg Rahnenführer

Date of Oral Examination:
May 14th, 2025

# Abstract

New research approaches of sciences like statistics and toxicology have been coming up through rapid development in the last years. However, those research approaches often refer to only one of the two sciences, not considering important aspects of the other discipline. Especially in the crucial part of planning an experiment, the laboratory routine in toxicology does not consider optimal design theory of statistics, although there is already much research present, which could help to improve the research results. This demonstrates a huge gap between practical applications in toxicological research and existing statistical theory. On the one hand, this gap exists due to the missing statistical methods specifically tailored to toxicological applications, and on the other hand, if those statistical methods exist, they are not reported in a clear manner for non-statisticians. The consequences for toxicological experiments are a waste of observations, or even worse animals, and non-optimal results in terms of precision. Therefore, this is an important aspect in the field of statistics and toxicology, which needs to be addressed. Optimal design approaches specifically tailored to toxicology must be developed and reported appropriately.

This cumulative thesis is based on three works, that all present optimal design approaches for diverse toxicological applications. The first manuscript highlights the importance of considering optimal design approaches for classical cytotoxicity experiments in a user-friendly manner. Here, different optimal design approaches are compared to typically used designs in practice based on an extensive case study. Moreover, a guideline for cytotoxicity experiments and an R-Shiny software tool are presented, which both facilitate the planning of upcoming cytotoxicity experiments. In the second manuscript a new design approach for the precise estimation of effective dose sets in drug combination studies is developed. For that matter, the performance of the corresponding developed criterion is investigated in a simulation study based on various scenarios including a case study. Finally, a new design approach for the analysis of high-dimensional gene-expression data is developed in the third manuscript. While two of the manuscripts are already published, the second one is attached in its current, unpublished form.

# Acknowledgments

Zunächst möchte ich meinen aufrichtigen Dank an meine Betreuerin Kirsten Schorning aussprechen. Sie hat mich in allen Situationen stets unterstützt und gefordert, wodurch sie das Beste aus mir herausgeholt hat. Ihre innovativen Ideen und ihr unermüdliches Engagement sowie ihre Leidenschaft für die Forschung begeistern mich immer wieder.

Ebenfalls möchte ich Jörg Rahnenführer danken, der mir mit seinem konstanten Optimismus und seiner Unterstützung in jeder Phase zur Seite stand. Mein Dank gilt auch Jan Hengstler, der mir durch zahlreiche Gespräche und seine eigene Begeisterung für die Toxikologie ganz neue Perspektiven und Aspekte der Forschung eröffnet hat. Der interdisziplinäre Austausch hat mir ermöglicht, sowohl fachlich als auch überfachlich sehr dazuzulernen.

Besonders hervorheben möchte ich das gesamte Umfeld des GRKs. Der Zusammenhalt unter den Kolleginnen und Kollegen hat mich in schwierigen Phasen sehr gestärkt. Viele meiner Kolleginnen und Kollegen sind durch zahlreiche Konferenzen und außeruniversitären Treffen zu Freunden geworden. Ein ganz besonderer Dank gilt meiner direkten Arbeitskollegin Julia Duda. Durch unsere Freundschaft und die vielen herzerwärmenden Gespräche während des "Rollercoasters" der Promotion ist mir der Arbeitsalltag oft viel leichter gefallen.

Mein tiefster Dank gilt meinem Verlobten Nils, der mich in jeder Phase meiner Promotion unterstützt, mir stets Mut zugesprochen und immer an mich geglaubt hat.

Zuletzt möchte ich meiner Familie danken, die mich bedingungslos unterstützt hat. Besonders meiner Mutter, ohne deren Hilfe ich nie so weit hätte kommen können, sowie meiner Schwester Anna, die mir in allen Lebenslagen eine große Stütze ist.

# List of Publications

This cumulative thesis is based on the following three manuscripts:

**Article 1:** Schürmeyer, L., Peng, C., Albrecht, W., Brecklinghaus, T., Baur, P., Hengstler, J. G., & Schorning, K. (2025). Design of optimal concentrations for in vitro cytotoxicity experiments. Archives of Toxicology, 99, 357–376. https://doi.org/10.1007/s00204-024-03893-1

Contribution of the author:

Kirsten Schorning and the author of this thesis had the central idea for this project. Furthermore, the author implemented nearly all data analyses, performed nearly all simulation studies and implemented the corresponding Shiny App. The guideline was developed from worthwhile discussions with Kirsten Schorning and Jan Hengstler. Additionally the author of this thesis made the first draft of the manuscript in close cooperation with Kirsten Schorning and Jan Hengstler. Constructive comments of the co-authors were incorporated.



**Article 2:** Schürmeyer, L., Sandig, L., Hezler, L., Igl, B. W., & Schorning, K. Optimal designs for identifying effective doses in drug combination studies. Submitted. A preprint is also available in arXiv (https://doi.org/10.48550/arXiv.2506.05913).

Contribution of the author:

The original idea for this project was proposed by Kirsten Schorning. The author of this thesis developed central theorems and proofs, implemented all data analyses and wrote the first draft of the manuscript supervised by Kirsten Schorning. Constructive comments of the co-authors were incorporated. The manuscript is attached in its current version.

**Article 3:** Schürmeyer, L., Schorning, K., & Rahnenführer, J. (2023). Designs for the simultaneous inference of concentration–response curves. BMC Bioinformatics 24, 393. https://doi.org/10.1186/s12859-023-05526-3

Contribution of the author:

Kirsten Schorning and Jörg Rahnenführer proposed the first idea for this project. The author of this thesis performed all data analyses and simulation studies, added an own design approach and wrote the first draft of the manuscript. Discussion and revising the manuscript was done together with Kirsten Schorning and Jörg Rahnenführer.

Further publications:

Kappenberg, F., Duda, J.C., Schürmeyer, L., Gül, O., Brecklinghaus, T., Hengstler, J. G., Kirsten Schorning & Jörg Rahnenführer (2023). Guidance for statistical design and analysis of toxicological dose–response experiments, based on a comprehensive literature review. Archives of Toxicology, 97, 2741–2761. https://doi.org/10.1007/s00204-023-03561-w

# Contents

# Part I.

# Introduction

# 1. Motivation

One of the most famous quotes in toxicology, attributed to Paracelsus, is: "The dose alone makes a thing [not] a poison" (originally in German: "Allein die Dosis macht, dass ein Ding (k)ein Gift sei") (Paracelsus, 1965). This quote highlights the crucial aspect of dose-response analysis, specifically regarding how dose levels influence the (toxic) effects of a substance. Although the assertion that only the dose affects toxicity is now considered outdated (Bödeker and Moebus, 2024), as other factors like e.g. exposure time can also influence toxicity, it nonetheless illustrates a fundamental principle of toxicological analysis. In most contexts, the dose of a substance remains the primary factor of interest, prompting researchers to investigate the relationship between different dose levels and their corresponding responses precisely.

The dose-response relationship is investigated in various different fields. One important application area is the development of new drugs, where specifically the "best" dose for a therapy is of interest (Ting et al., 2017). Besides, toxicity is often studied in preclinical settings or outside of pharmaceutical contexts (Hothorn, 2016). All fields share a common goal: to measure the dose-response relationship as precisely as possible and to extract the most information from it.

For that matter, alert thresholds or effective doses are a primary concern, to detect which dose level leads to a specific amount of the maximal effect. For instance, in a typical toxicological cell experiment, where cell viability is measured, an alert threshold can be defined by the absolute effective dose $ED_{50}$, which refers to the dose at which exactly 50% of the cells die.

Given the numerous possible applications mentioned above, dose-response analysis and the ongoing improvement of the associated methods are very important. In this thesis three current challenges related to optimal design theory in toxicology are presented and discussed, new methods are developed to address the corresponding research gaps and improve future experiments.

The first important challenge is referring to the design of typical dose-response experiments in general. The current state of the statistical methodological research regarding the design of experiments is not yet reflected in the laboratory routine (see Kappenberg et al., 2023). Kappenberg et al. (2023) reviewed all papers of 2021 in three journals selected out of the top toxicological journals and examined the methods for planning, executing and analyzing dose-response analyses. From a total of 5 670 dose-response analyses, not one referred to some statistical consideration of the experimental design. Furthermore, the applied experimental designs were often not clearly stated at all. Instead, laboratory routines typically employ equidistant or log-equidistant designs based on dilution factors, despite the fact that arbitrary concentration choices are feasible in cytotoxicity experiments. In contrast to this, there is extensive theoretical research on optimal designs (see Pukelsheim (2006), Dette et al. (2008) among many others) that demonstrates how fewer observations are needed in an experiment, when planning the experimental design in an optimal way. Hence, optimal design of experiments lead to reduced costs and, in animal studies, a decrease in the number of animals required.

Due to the fact that statistical concepts are often very complex and not necessarily customized for toxicological contexts, they are often not known in practice and additionally quite difficult to apply for practitioners. Furthermore, there are very few user-friendly tools available to implement the concepts of optimal experimental design (Holland-Letz and Kopp-Schneider, 2021). This large gap between practical application and existing theory in development must be bridged by more toxicology-related statistical research.

Recently, interest has shifted from examining the effects of a single substance only to exploring the combined effects of two substances, leading to the second important aspect of this thesis dealing with dose-combination experiments. Such preclinical dose-combination studies serve as the foundation for corresponding clinical trials in humans (Chou, 2008). In these cases, while the dose-response relationship for each individual substance is often well understood, the combined effect remains unknown. There are different ways to describe the interaction of two substances. Established strategies are for example isobologram or combination index analysis as well as response surface models (see e.g., Zhao et al. (2010), Foucquier and Guedj (2015) and Holland-Letz and Kopp-Schneider (2018)). While isobole analysis and combination index methods reduce the interaction analysis to a one-dimensional linkage, response surface analysis allows for the investigation of the dose-response shape across all dose combinations within the design space. Hence, Zhou et al. (2024) propose modeling drug combination response data with response surface models.

Research has already been conducted on how to optimally design drug-combination studies based on the classical D-optimality criterion (see e.g. Papathanasiou et al., 2019), which aims for precise estimation of the overall response surface. Similar to analyses involving a single substance, specific threshold levels, referred to as effective doses $\mathrm{ED}_p$, are also of interest in drug-combination settings. These effective doses correspond to achieving a specific percentage $p\%$ of the maximum effect within the design space of interest. Due to the two-dimensional nature of the problem, multiple combinations of doses can yield this specific amount $p$, resulting in contour lines on the response surface. The key question then arises: Which design should be selected when specific threshold levels are being targeted?

The third challenge covered in this thesis arises from high-dimensional toxicological data sets. In the last decades the scope in toxicological research enhanced with the field of gene expression analysis leading to a large amount of data. In gene expression analysis, the focus is on how information from genes is utilized to synthesize specific products within an organism. Here, the translation of the genetic information into such products, e.g. proteins, is of interest. In detail, the genetic information, contained in each cell of an organism and referred to as deoxyribonucleic acid (DNA), compresses the information for the translation of products in each cell and therefore contributes to the phenotype of the organism (Berg et al., 2018). Therefore, gene expression serves as a link between the genotype - the genetic information - and the resulting phenotype, which corresponds to the observable characteristics. By analyzing such genetic information, we can gain a deeper understanding of biological processes and diseases.

According to Baldi and Hatfield (2011), DNA microarray analyses were primarily conducted for this purpose, allowing for the simultaneous examination of all genes in a cell using a gene chip. Under predefined conditions, the level of gene expression for each gene is determined simultaneously. This leads to new challenges in the design aspect, since one design is required for evaluating thousands of genes simultaneously. Here, a precise estimation of the concentration-response relationship of all genes is of interest. Another challenging aspect after conducting the experiment is the large amount of data due to the high-dimensionality.

The works presented in this thesis address the three challenges mentioned earlier in various ways. To tackle these challenges, the design of experiments is explored in one-dimensional, two-dimensional, and high-dimensional settings in each article, respectively. The first article highlights the importance of optimal designs in practice, presenting them in a user-friendly manner. It focuses on cytotoxicity experiments, where multiple concentration choices are available for designing experiments. However, practitioners often rely on dilution factors when selecting the design of the experiment. This article compares commonly used design approaches in practice with those derived from optimal design theory through an extensive case study. Through data-driven analyses and visualizations, the importance of employing optimal designs is demonstrated for practitioners. Additionally, the article provides guidelines for upcoming cytotoxicity experiments and introduces a user-friendly software tool to facilitate the calculation of optimal designs.

The second article tackles the challenge of making appropriate design choices for drug-combination studies. A new design criterion is introduced, aimed at achieving precise estimations of effective dose levels within a two-dimensional framework. This article develops concepts from optimal design theory, including equivalence theorems, efficiency bounds and analytical results. Additionally, the new design approach is compared to established factorial and ray designs as well as D-optimal designs across various scenarios, including a detailed case study.

Finally, the third paper addresses the challenge of designing toxicological experiments involving high-dimensional gene-expression data. In this work, concepts from optimal design theory are adapted to high-dimensional contexts. A new criterion is developed specifically for estimating a large number of concentration-response curves. The newly developed design, called simultaneous inference design, is compared to various other approaches, including established methods such as log-equidistant and equidistant designs. This comparison is based on a real case study and includes both theoretical evaluations based on efficiencies and practical assessments conducted through a large simulation study.

The thesis is structured as follows: Chapter 2 presents the statistical background necessary for understanding the three articles in detail. The concepts described here are well-established and have been developed prior to this research. All newly contributed methods are based on these concepts. Chapter 3 offers summaries of all articles included in this thesis. Finally, Chapter 4 provides a comprehensive discussion of these contributions. Additionally, all articles are attached at the end.

# 2. Statistical Methods

In the following chapter, the central statistical methods employed in this thesis are presented to provide the statistical background necessary for understanding all three articles.

Throughout this thesis, the terms "dose" and "concentration" are used interchangeably. From an applied perspective, these terms describe different biological contexts. However, in terms of the mathematical relationship within the model, there is no distinction between dose and concentration. Consequently, the term "dose" is mostly used within this thesis.

## 2.1. Dose-response modeling

In numerous scientific fields, including toxicology and clinical research, it is essential to accurately determine the relationship between the dose of a substance and its effect (see Proctor et al. (2017), Vorrink et al. (2018), Brecklinghaus et al. (2022) among many others). The effect, which can be any metric effect of interest, is here referred to as response. For instance, in toxicology, one often measures cell viability, expressed as the percentage of living cells. To analyze the relationship of the dose and its response, dose-response experiments are conducted.

Dose-response experiments persist of $n$ different measurement levels $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathcal{X}$, where $\mathcal{X}$ defines the design space. At each measurement level $\boldsymbol{x}_i$ for $i = 1, \ldots, n$ multiple measurements $r_i$ can be obtained, while the total number of observations in such an experiment equals $N = \sum_{i=1}^{n} r_i$. This thesis considers one-dimensional measurement levels corresponding to a single dose level $x_i \in [0, x_{\max}] \subset \mathbb{R}$ of a substance $X$ of interest, as well as two-dimensional measurement levels represented by $\boldsymbol{x}_i = (v_i, w_i) \in [0, v_{max}] \times [0, w_{\max}] \subset \mathbb{R}^2$, which refer to dose combination levels of the substances $V$ and $W$, respectively. The design spaces are defined as either one-dimensional or two-dimensional and are bounded by the placebo dose(s) of 0 and the maximum considered dose levels $x_{\max}, v_{\max}, w_{\max}$. For simplicity, we refer to these as dose-response experiments, which may pertain to either one-dimensional doses or two-dimensional dose combinations, depending on the context.

The experimental design of a dose-response experiment can then be denoted as an exact design (Fedorov and Leonov, 2013), with

$$\xi_N = \begin{pmatrix} \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_n \\ \frac{r_1}{N} & \cdots & \frac{r_n}{N} \end{pmatrix}.$$

In the design context the measurement levels $\boldsymbol{x}_i$ are also called support points of the design $\xi_N$ with corresponding weights $\frac{r_i}{N}$ for $i = 1, \ldots, n$ and $\sum_{i=1}^{n} \omega_i = 1$.

Assuming that the weights of an exact design converge asymptotically to weights $\omega_i \in (0, 1]$ with $\lim_{N \to \infty} \frac{r_i}{N} = \omega_i$ for all $i = 1, \ldots, n$, we can denote the (continuous) design $\xi$ as a probability measure with mass $\omega_i$ at each design point $\boldsymbol{x}_i \in \mathcal{X}$, as indicated by Kiefer

(1974). This is represented by

$$\xi = \begin{pmatrix} \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_n \\ \omega_1 & \cdots & \omega_n \end{pmatrix}.$$

The distinction between a design $\xi$ and the exact design $\xi_N$ lies in the fact that $\xi$ is independent of the total number of observations $N$ and the weights $\omega_i$ vary continuously within the interval $(0, 1]$.

To gather information about the effect of the substance(s) of interest after conducting a dose-response experiment, the following functional relationship is assumed:

$$Y_{ij} = \eta(\boldsymbol{x}_i, \boldsymbol{\theta}) + \varepsilon_{ij} \qquad i = 1, \ldots, n, \quad j = 1, \ldots, r_i, \quad \boldsymbol{\theta} \in \boldsymbol{\Theta} = \mathbb{R}^p, \tag{2.1}$$

which corresponds to a (non)linear regression model, where $\varepsilon_{ij}$ are independent random variables such that $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ with $\sigma^2 > 0$ unknown. The function $\eta : \mathcal{X} \times \mathbb{R}^p \to \mathbb{R}$ describes the assumed dose-response shape and $\boldsymbol{\theta}$ the corresponding $p$-dimensional model parameter.

For specific experimental data the corresponding dose-response shape can be estimated based on the Maximum Likelihood Estimation of the model parameter $\boldsymbol{\theta}$ (Casella and Berger, 2024). The Maximum Likelihood Estimation method aims for the best-fitting estimates by maximizing the likelihood of the functional relationship, thereby ensuring the highest probability that the observed data points align with the model.

First the likelihood needs to be specified. For the regression model in (2.1) the corresponding likelihood is defined as

$$\mathcal{L}(\boldsymbol{\theta}, \sigma^2 | \boldsymbol{y}, \boldsymbol{x}) = \prod_{i=1}^{n} \prod_{j=1}^{r_i} f(y_{ij} | \boldsymbol{x}_i, \boldsymbol{\theta}) = \prod_{i=1}^{n} \prod_{j=1}^{r_i} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{\big(y_{ij} - \eta(\boldsymbol{x}_i, \boldsymbol{\theta})\big)^2}{2\sigma^2} \right\},$$

since we assume that the response values $Y_{ij}$ are independent and follow a normal distribution $\mathcal{N}(\eta(x_i, \boldsymbol{\theta}), \sigma^2)$ for $i = 1, \ldots, n$ with density function $f(y_{ij} | \boldsymbol{x}_i, \boldsymbol{\theta})$. Here, the vector of responses is denoted as $\boldsymbol{y} = (y_{11}, \ldots, y_{n\, r_n})$ and the vector of dose levels is denoted as $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$. Let $\hat{\boldsymbol{\theta}}$ define a function of $\boldsymbol{y}$ and $\boldsymbol{x}$, where the likelihood attains its maximum with $\hat{\boldsymbol{\theta}}(\boldsymbol{y} | \boldsymbol{x})$ for a fixed $\boldsymbol{y}$. Mathematically this means

$$\hat{\boldsymbol{\theta}}(\boldsymbol{y} | \boldsymbol{x}) = \arg\max_{\boldsymbol{\theta} \in \Theta} \mathcal{L}(\boldsymbol{\theta}, \sigma^2 | \boldsymbol{y}, \boldsymbol{x}).$$

Following Casella and Berger (2024) the Maximum Likelihood Estimator (short: MLE) of $\boldsymbol{\theta}$ then is defined as $\hat{\boldsymbol{\theta}}(\boldsymbol{Y} | \boldsymbol{x})$ for the sample of random variables $\boldsymbol{Y} = (\boldsymbol{Y}_{11}, \ldots, \boldsymbol{Y}_{nr_n})$. The corresponding Maximum Likelihood estimate is attained with the realized value of the estimator based on the experimental data.
Note that $\sigma$ also needs to be estimated for the Maximum Likelihood estimate $\hat{\boldsymbol{\theta}}$, but can be estimated separately (for details see Casella and Berger, 2024). For simplicity, we will denote the MLE in the following as $\hat{\boldsymbol{\theta}}$ instead of $\hat{\boldsymbol{\theta}}(\boldsymbol{Y} | \boldsymbol{x})$.

Besides, different methods to estimate the model parameter $\boldsymbol{\theta}$ exits. Apart from the Maximum Likelihood estimation another popular approach is the Least-Squares estimation (short: LS estimation). When estimating the model parameter with the LS estimation, the parameter estimate $\boldsymbol{\theta}_{LS}$ shows the smallest quadratic deviance of the considered data points (Fedorov and Leonov, 2013). In case of normally distributed data points, like we assume in (2.1), the MLE and ordinary Least-Squares estimator coincide.

Another property of the MLE is that, under standard regularity conditions and in case of identifiability of $\boldsymbol{\theta}$, it can be assumed that the MLE $\hat{\boldsymbol{\theta}}$ is asymptotically normally distributed with:

$$\sqrt{N}\big(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\big) \xrightarrow{\mathcal{D}} \mathcal{N}\big(0, M^{-1}(\xi, \boldsymbol{\theta})\big), \tag{2.2}$$

while $\mathcal{D}$ stands for convergence in distribution (Jennrich, 1969). On the one hand, this implies that the asymptotic mean is equal to zero, indicating that the Maximum Likelihood Estimator is consistent. On the other hand, it signifies that the asymptotic variance of the Maximum Likelihood Estimator $\hat{\boldsymbol{\theta}}$ is directly related to the corresponding information matrix, which is defined as

$$M(\xi, \boldsymbol{\theta}) = \frac{1}{\sigma^2} \int_{\mathcal{X}} \frac{\partial}{\partial \boldsymbol{\theta}} \eta(\boldsymbol{x}, \boldsymbol{\theta}) \left( \frac{\partial}{\partial \boldsymbol{\theta}} \eta(\boldsymbol{x}, \boldsymbol{\theta}) \right)^T d\xi(\boldsymbol{x}),$$

for normally distributed errors $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ as assumed in (2.1). It is important to note that the information matrix depends on both the parameter $\boldsymbol{\theta}$ and the chosen design $\xi$.

A specific example of dose-response modeling is the four-parameter log-logistic model (short: 4pLL-model), which is commonly used in the context of one-dimensional dose-response analysis. Following Ritz et al. (2015) the 4pLL-model is defined as

$$\eta(x, \boldsymbol{\theta}) = c + \frac{d - c}{1 + \exp\big(b(\log(x) - \log(e)\big)}, \quad \boldsymbol{\theta} = (b, c, d, e)^T, \tag{2.3}$$

where $x$ refers to the dose level within the design space $\mathcal{X} \subset \mathbb{R}$. The lower and upper asymptotes are described by $c$ and $d$, respectively, whereas the steepness of the curve is described with parameter $b$. The parameter $e$ refers to the effective concentration $\mathrm{EC}_{50}$, the dose where 50% of the maximal effect is achieved. For practical applications, the 4pLL-model is implemented in the R package `drc` (Ritz et al., 2015). Here, a least-squares estimation of the model parameter is used for estimating the model parameter.

The 4pLL-model is often referred to as the sigmoid Emax model, which represents a different parameterization of this model. According to Macdougall (2006), the sigmoid Emax model is defined as follows:

$$\eta(x, \boldsymbol{\theta}) = \mathrm{E}_0 + \frac{x^h \cdot \mathrm{E}_{\max}}{x^h + \mathrm{EC}_{50}^h}, \qquad \boldsymbol{\theta} = (\mathrm{E}_0, \mathrm{E}_{\max}, \mathrm{EC}_{50}, h)^T, \tag{2.4}$$

where $\mathrm{E}_0$ corresponds to the placebo effect, $\mathrm{E}_{\max}$ denotes the maximal effect, $\mathrm{ED}_{50}$ represents the effective dose, where $p$ of the maximal effect is achieved, and the steepness is described with $h$. Although the model formulas appear different, both models (2.3) and (2.4) yield the same model curve (see Kappenberg (2021) for details). In practice, the sigmoid Emax model can be fitted using the `DoseFinding` package in R (Bornkamp et al., 2024). Note that due to different optimization functions used in both R packages there can be numerical differences for the models.

Given that the biological relationship frequently results in a sigmoidal pattern, and the 4pLL-model (also known as the sigmoid Emax model) effectively captures this behavior with its sigmoidal shape, it is one of the most widely used dose-response models (Goutelle et al., 2008). However, alternative models such as linear, exponential, Emax, or quadratic models are also potential options (see e.g. Bretz et al., 2005).

It should be noted that calculating estimates for nonlinear regression functions can lead to difficulties. In the mentioned examples of the 4pLL-model/ sigmoid Emax model, it may

occur that the corresponding likelihood cannot be computed due to underlying computationally singular information matrices when the slope parameter is chosen to be very high (Mielke, 2016). Additionally, it should be mentioned that the likelihood changes very little numerically at higher slope parameters, as the likelihood function exhibits an extremely flat slope around its maximum (see again Mielke (2016)). Consequently, it is sensible to constrain the slope parameter when estimating parameters in practice.

## 2.2. Optimal design theory

A main challenge before conducting an experiment in general is how the corresponding design should be chosen. The field of optimal design of experiments provides statistical methods to calculate optimal designs for diverse experimental settings. In this thesis optimal design theory for dose-response modeling (see Section 2.1) is presented.

When an experiment is planned the main intention is to maximize the information gain, which corresponds to achieving most precise results. Since we assume an underlying functional relationship between dose and response with (2.1), this means a precise estimation of the model implicating a precise estimation of the corresponding model parameter $\boldsymbol{\theta}$. Since the primary objective of estimating is to achieve an unbiased estimate, which is asymptotically accomplished by (2.2), the next goal is to be most precise. Thus, a minimal variance is of interest.

The link between the asymptotic variance of the MLE and the corresponding information matrix (see (2.2) for details) can be utilized to identify optimal design choices. Those result in minimal variance referred to minimizing the inverse of the information matrix $M^{-1}(\xi, \boldsymbol{\theta})$ or equally maximizing the information gain by maximizing $M(\xi, \boldsymbol{\theta})$. Here, the challenge occurs, that the variance in fact can be expressed via the information matrix, but is is not directly clear when this variance gets minimal (or the information maximal), since there is no total order for matrices. There exits only a partial order, called Loewner ordering. According to Pukelsheim (2006), the Loewner ordering is defined on the set of $k \times k$ symmetric matrices denoted with $\text{Sym}(k)$. Two matrices $A, B \in \text{Sym}(k)$ can be ordered with the Loewner ordering denoted with $\geq_L$ if the following holds

$$A \geq_L B \quad \Leftrightarrow \quad A - B \geq_L 0 \quad \Leftrightarrow \quad A - B \in \text{NND}(k),$$

where $\text{NND}(k)$ denotes the set of non-negative definite $k \times k$ matrices. The problem is, there are matrices, which are not comparable with the Loewner ordering given that neither $A \leq_L B$ or $B \leq_L A$ holds true. Due to this lack of a total order for matrices, the minimal variance based on the information matrix cannot be calculated directly. Criteria are used to specify an order of such matrices by utilizing their properties.

On one hand such criteria need to offer the ability to link the matrices to real values, such that we can use the order of the real numbers to identify a design with a corresponding information matrix as optimal, and on the other hand those criteria need to provide a unique extreme value such that the optimal design can be identified explicitly. Mathematically this means optimality criteria are based on concave or convex functions depending on the context. If minimizing the variance expressed with $M^{-1}(\xi, \boldsymbol{\theta})$ is of interest, a convex criterion can provide a unique minimum. Pukelsheim (2006) emphasizes that the primary aspect of optimal design of experiments is maximizing information gain, which is linked to maximizing the information matrix $M(\xi, \theta)$. Consequently, a criterion, which aims for such a maximization results, should be a concave function to provide a unique maximum.

For simplicity, the following section only refers to concave criteria that maximize information in terms of the information matrix $M(\xi, \boldsymbol{\theta})$ itself. Due to the functional relationship between the asymptotic variance and the inverse of the information matrix, all concepts of concave criteria can also be transferred to this relationship and corresponding convex criteria.

Following Pukelsheim (2006) optimal design criteria are defined as functions that assign a real value to each information matrix based on its characteristics. The set of all possible information matrices is denoted with the following set

$$\mathcal{M} := \{M(\xi, \boldsymbol{\theta}) : \xi \in \Xi\},$$

where $\Xi$ corresponds to the set of all possible designs for the design space $\mathcal{X}$. The set $\mathcal{M}$ corresponds to a convex subset of the set of non-negative definite $p \times p$ matrices $\mathrm{NND}(p)$ (Pukelsheim, 2006). Specifically, a criterion $\varphi$ is defined as a concave and differentiable function $\varphi : \mathrm{NND}(p) \to \mathbb{R}$. Given the inherent order of real numbers and the concavity of the criterion, the design leading to the unique maximal information (expressed via the information matrix) can be calculated explicitly based on such criteria. Therefore, the objective to maximize the information based on the criterion $\varphi$ on all possible information matrices yields the optimal design. More precisely, this means a design $\xi^*$ is $\varphi$-optimal, if and only if

$$\xi^* = \arg\max_{\xi \in \Xi} \ \varphi(M(\xi, \boldsymbol{\theta})).$$

Following Pukelsheim (2006), a concave criterion $\varphi$ is directly related to the convex analog with $1/\varphi$. Thus, the concavity of the information criterion $\varphi$ (or indirect the convexity of $1/\varphi$) allows for the application of convex optimization methods to identify the optimal design within the convex set of all possible information matrices $\mathcal{M}$.

In the case of a linear regression function $\eta(\boldsymbol{x}, \boldsymbol{\theta}) = \boldsymbol{\theta}^T g(\boldsymbol{x})$, with $\boldsymbol{g}(\boldsymbol{x}) = \big(g_1(\boldsymbol{x}), \ldots, g_p(\boldsymbol{x})\big)$ being a vector of basis functions, like e.g. polynomials, the information matrix does not depend on the parameter $\boldsymbol{\theta}$. This is because the defining gradient $\frac{\partial}{\partial \boldsymbol{\theta}} \eta(\boldsymbol{x}, \boldsymbol{\theta}) = g(\boldsymbol{x})$ does not depend on $\boldsymbol{\theta}$ in this context. However, when the function $\eta$ in (2.1) is nonlinear, this dependence arises, and the information matrix becomes a function of $\boldsymbol{\theta}$. For this scenario, Chernoff (1953) introduced the concept of locally optimal designs. The optimal design $\xi^*_{\boldsymbol{\theta}}$ is calculated based on an initial estimate of the parameter $\boldsymbol{\theta}$. This approach yields the theoretically most precise estimation when this initial estimate coincides with the true value $\boldsymbol{\theta}$. Conversely, if the parameter is misspecified, the corresponding locally optimal design $\xi^*_{\boldsymbol{\theta}}$ may lead to a poor performance.

The calculation of optimal designs can be quite complicated due to the underlying convex optimization problem. Therefore, it is beneficial to verify whether a design is $\varphi$-optimal in a more pragmatic manner for practical purposes. In this context, equivalence theorems are highly useful for assessing optimality. These show that the maximum has been found for $\xi^*$ with $M(\xi^*, \boldsymbol{\theta})$. Since the concavity of the criterion $\varphi$ ensures that the maximum is unique, it only needs to be shown that it is indeed a maximum. For a maximum all directional derivatives must be less than or equal to zero.

Mathematically, this means, according to Silvey (1980) a design $\xi^*$ is $\varphi$-optimal, if and only if the Fréchet derivative of $\varphi$ evaluated in $M^* = M(\xi^*, \boldsymbol{\theta})$ is less or equal to zero for all directions $E$, with $E$ being an information matrix of the Dirac measure at $\boldsymbol{x}_0$, denoted as $\delta_{\boldsymbol{x}_0}$, i.e. $E = \boldsymbol{x}_0 \boldsymbol{x}_0^T \in \{M(\delta_{\boldsymbol{x}_0}, \boldsymbol{\theta}) : \boldsymbol{x}_0 \in \mathcal{X}\} =: \mathcal{D}$. In detail, this means

$$F_\varphi\big(M^*, E\big) = \lim_{\varepsilon \to 0} \varphi\big((1-\varepsilon)M^* + \varepsilon E\big) - \varphi\big(M^*\big) \leq 0 \ \ \forall E \in \mathcal{D}. \tag{2.5}$$

It is sufficient to show the inequality for all $E \in \mathcal{D}$ since the set of all possible information matrices $\mathcal{M}$ is the convex hull of the set of all Dirac information matrices $\mathcal{D}$.

The quality of a design $\xi$ in comparison to a $\varphi$-optimal design $\xi^*$ in general can be quantified with the corresponding efficiency. Following Pukelsheim (2006), the corresponding efficiency is defined by

$$\mathrm{Eff}_\varphi(\xi, \boldsymbol{\theta}) := \frac{\varphi(\xi, \boldsymbol{\theta})}{\varphi(\xi^*, \boldsymbol{\theta})},$$

whose value lies in $[0, 1]$ by definition. With the efficiency, the ability of design $\xi$ to utilize the information can be expressed in a percentage.

In most application scenarios, it can be assumed that there is variability among different experiments, which may arise from different laboratory settings or day-to-day effects. Since designing new dose-response experiments requires an initial estimate of the model parameter $\boldsymbol{\theta}$, relying on a rough estimate often results in misspecifications of the true parameter due to this variability. Consequently, this highlights the necessity for more robust designs.

For a predefined prior distribution $\pi$ of $\boldsymbol{\theta} \in \Theta$ (Pseudo)-Bayesian designs can assure more robustness based on the different parameter values within the considered distribution. Following Pronzato and Walter (1985) and Chaloner and Larntz (1989) a design $\xi$ is Bayesian $\varphi$-optimal for the criterion $\varphi$, if it maximizes

$$\Psi(\xi) := \int_\Theta \varphi(\xi, \boldsymbol{\theta})\ d\pi(\boldsymbol{\theta})$$

for all possible designs $\xi \in \Xi$ in the design space.

After introducing the key concepts of optimal design theory in general for a criterion $\varphi$, we will now focus on a specific example: the D-optimality criterion, which is often addressed in this thesis. Following Chernoff (1953) and Pukelsheim (2006), a design $\xi_D$ is called locally D-optimal for estimating the parameter $\boldsymbol{\theta} \in \mathbb{R}^p$, if it maximizes the concave functional

$$\varphi_D(\xi, \boldsymbol{\theta}) := \det\big(M(\xi, \boldsymbol{\theta})\big)^{1/p}$$

among all designs $\xi$ on the design space $\mathcal{X}$. Here, maximizing the determinant of the information matrix corresponds to maximizing the information gain in general. Therefore, the D-optimality aims for a precise estimation of the complete parameter $\boldsymbol{\theta}$.

Furthermore the relationship of concave and convex criteria can be illustrated easily based on the D-optimality criterion. According to the definition of the D-optimality, the criterion is concave and aims for maximizing the information gain, resulting in a unique maximum accomplished by the D-optimal design $\xi_D$. If instead minimizing the variance is of interest, the criterion can easily be transformed to a convex criterion, which aims for minimizing the determinant of the dispersion matrix with $\det(M^{-1}(\xi, \boldsymbol{\theta}))^{1/p}$ and leads to the same unique minimum $\xi_D$ (Pukelsheim, 2006).

In practical applications the calculation of such D-optimal designs is often based on meta heuristic algorithms, which do not guarantee convergence to an extreme value. Therefore, it is desirable to verify whether a design is locally D-optimal with a corresponding equivalence theorem. For that matter, it can be validated if a design $\xi_D$ is locally D-optimal

by checking whether the inequality

$$d(\boldsymbol{x}, \xi_D, \boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \eta^T(\boldsymbol{x}, \boldsymbol{\theta}) M^{-1}(\xi_D, \boldsymbol{\theta}) \frac{\partial}{\partial \boldsymbol{\theta}} \eta(\boldsymbol{x}, \boldsymbol{\theta}) - p \leq 0 \tag{2.6}$$

is satisfied for all $\boldsymbol{x} \in \mathcal{X}$ (see Fedorov and Leonov, 2013). In other words this shows that the D-optimal design $\xi_D$ maximizes the criterion $\varphi_D$ via calculating the Fréchet derivative of $\varphi_D$. In detail, the inequality (2.6) is an example of inequality (2.5).

The quality of another design $\xi$ compared to the D-optimal design $\xi_D$ can then be assessed by its $D$-efficiency, which is defined by

$$\text{Eff}_{\varphi_D}(\xi, \boldsymbol{\theta}) := \frac{\varphi_D(\xi, \boldsymbol{\theta})}{\varphi_D(\xi_D, \boldsymbol{\theta})},$$

whose value lies in $[0, 1]$ by definition. The better the design $\xi$ is in terms of the D-optimality criterion, the greater is its efficiency. Additionally, the efficiency of the D-optimality is directly linked to the portion of additional observations needed for the non-optimal design compared to the optimal design (Fedorov and Leonov, 2013). In detail $(1/\text{Eff}_{\varphi_D}(\xi, \theta) - 1) \times 100$ % more observations are needed for a non-optimal design compared to the D-optimal design $\xi$. For instance this corresponds to 25% more observations needed for a non-optimal design with a $D$-efficiency of 0.8.

## 2.3. Particle swarm optimization

The particle swarm optimization (short: PSO) is an iterative and metaheuristic algorithm for optimization problems. It is often used for complex optimization problems, like they can occur in optimal design theory (Chen et al., 2022). Originally the idea of the PSO was proposed by Kennedy and Eberhart (1995) and is based on nature's concepts. To be precise, the behavior of a swarm, e.g. birds searching for the best nesting site, serves as the basis for the concept of the PSO and also contributes to its name. Just as in the example of birds searching for the best nesting site, the PSO seeks an optimum in a given optimization problem. For that matter, two fundamental ideas are the basis of the PSO. On the one hand, the position of objects of the swarm, which in the example correspond to individual birds, influences the PSO, while each object's position could be the potential optimal solution. On the other hand, the movement/velocity of the swarm objects is part of the PSO. This movement is influenced by different factors: the own velocity of each object (bird), the velocity of the swarm (bird flock) and random components. Due to this concept the movement of the total swarm should be directed towards the optimum, which then can be identified.

Nowadays, there are also other metaheuristic algorithms defined on similar principles as the PSO, e.g. the imperialist competitive algorithm (see Atashpaz-Gargari and Lucas, 2007 for details), which is also frequently used for in terms of the calculation of optimal designs (see e.g. Masoudi et al., 2017). However, for the sake of brevity, this thesis will focus on explaining the PSO.

According to Clerc (2006), the PSO of an optimization problem, based on a criterion of interest, can be mathematically described based on the following principles. Unless stated otherwise, the following refers to Clerc (2006).

First, the number of considered objects in the swarm $P$, in the following denoted as particles, and the number of iterations $T$ needs to be fixed in advance. The PSO then

includes the number of particles specified in each iteration step, while it ends with the last iteration step. Consequently, the higher both numbers are, the more likely it is that the optimum can be found by the PSO. However, the number of iterations as well as the number of particles influences the runtime of the algorithm, which can exceed quickly. Therefore, depending on the considered optimization problem, a tradeoff between a high number of particles and iterations as well as a short runtime needs to be found.

In case of a $z$-dimensional problem in the problem space $\mathcal{Z} = [x_{\min}, x_{\max}]^z \in \mathbb{R}^z$, each article has positions $\boldsymbol{x}_i$ with $\boldsymbol{x}_i = (x_{i,1}, \ldots, x_{i,z})$ and velocities $\boldsymbol{v}_i$ with $\boldsymbol{v}_i = (v_{i,1}, \ldots, v_{i,z})$. In the beginning both the positions as well as the velocities of each particle are assigned randomly based on a uniform distribution in $[(x_{\min} - x_{\max})/2, (x_{\max} - x_{\min})/2]$.

In the progress of the algorithm the position of each particle changes based on certain impacts due to its velocity. For that matter, following Kennedy and Eberhart (1995) two components, a cognitive component and a social component, influence the velocity of the particles similar to the phenomenon in nature. The cognitive component incorporates the best criterion value of particle $i$ achieved until iteration $t \in \{1, \ldots, T\}$ denoted with $p_{\text{best}\,i,d}(t)$, where $d \in \{1, \ldots, z\}$, and therefore shows the cognitive learning from former iterations of the particle. The social component is based on the best criterion value of all particles achieved until iteration $t$, denoted with $g_{\text{best}\,d}(t)$. In each iteration, the quality of each particle's position is evaluated based on the criterion of interest. The best value for each particle, as well as the swarm's overall best value, is updated in each iteration. Each particle position is then updated in the iteration step with incorporating both the social and the cognitive component via

$$x_{i,d}(t+1) = x_{i,d}(t) + v_{i,d}(t+1)$$

with

$$\begin{aligned} v_{i,d}(t+1) = v_{i,d}(t) \\ &+ c_1 \cdot \text{Rand}\big(0,1\big) \cdot \big(p_{\text{best}\,i,d}(t) - x_{i,d}(t)\big) \\ &+ c_2 \cdot \text{Rand}\big(0,1\big) \cdot \big(g_{\text{best}\,d}(t) - x_{i,d}(t)\big). \end{aligned}$$

With $c_1 \cdot \text{Rand}(0,1)$ and $c_2 \cdot \text{Rand}(0,1)$ the velocity is also influenced randomly with a random number denoted with $\text{Rand}(0,1)$. According to Clerc (2006), the constant values $c_1$ and $c_2$ can be specified individually for each optimization problem. However, values leading to a better and faster implementation can only be found via experimental trial.

For practical applications it is also desirable to restrict the movement of the particles to the search space $\mathcal{Z}$ of interest. It should be prohibited that particles move outside the search space, with adjusting the particles $x_{i,d}(t) \notin [x_{\min}, x_{\max}]$ in each iteration $t$ with

$$\begin{aligned} &v_{i,d}(t) \leftarrow 0 \\ &x_{i,d}(t) < x_{\min} \Rightarrow x_{i,d}(t) \leftarrow x_{\min} \\ &x_{i,d}(t) > x_{\max} \Rightarrow x_{i,d}(t) \leftarrow x_{\max}. \end{aligned}$$

On the basis of this principles the position of the particles are adjusted in each iteration, while the final result of the PSO is given in the last iteration $T$ with the positions $d = 1, \ldots, D$ leading to the best value $g_{\text{best}\,d}(T)$, respectively.

Based on the random movement of the particles often a broad exploration of the search space is possible. Therefore, there is also a high probability for finding the optimal criterion

value. However, there is no convergence guarantee of the PSO. Nevertheless, the PSO is a good method to solve non-linear and multidimensional optimization problems efficiently (Shi et al., 2019). Especially, in terms of optimization problems regarding optimal designs, the optimality of the final PSO result can also be validated via an equivalence theorem (see Section 2.2 for details).

# 3. Summary of the Articles

## 3.1. Article 1: Design of optimal concentrations for in vitro cytotoxicity experiments

The first article provides important aspects of research in statistics and toxicology by emphasizing the importance of optimal designs in practice. There exists a huge gap between the current status in applicable statistical methods and the everyday laboratory routine. A recent literature review of Kappenberg et al. (2023) for three top toxicological journals, showed that no statistical considerations were made in laboratory routine in terms of the design choice. Often equidistant or log-equidistant designs based on specific dilution factors are used in practice, although other concentration choices would be realizable in cytotoxicity experiments. In contrast to this, there is already much research present, showing that optimal designs lead to the most precise estimation by using less observations than non-optimal designs.

The aim of this paper was to bridge the gap between theory and toxicological application. We emphasize the importance of optimal designs in practice by showing the impact of different design choices in a practical way based on an extensive case study. Furthermore, we provide a guideline for the design of upcoming cytotoxicity experiments and a user-friendly software tool for a direct calculation of the corresponding optimal choice of concentrations. Two different scenarios are considered: one in which prior knowledge of the substance of interest is directly available, and another in which there is no prior knowledge about the substance itself, but information exists regarding related substances.

In detail, we analyzed an extensive case study of the toxic substance valproic acid (short: VPA) persisting of four experiments with 50 concentrations each for the scenario with pre-existing knowledge of the test substance. This data set, which is extremely large and untypical in a toxicological setting, was used to specify a reliable reference curve for the concentration-response relationship of VPA via a four-parametric log-logistic model (4pLL-model). The specified references curve enabled the comparison of different choices of concentrations by calculating the corresponding precision to this reference curve. To be precise, in a simulation study new cytotoxicity experiments were imitated by sampling data points from the large VPA data set in each simulation step, so we maintain a realistic selection of observations. Two often used design approaches in laboratory routine, i.e. a log-equidistant design approach and an approach based on biological arguments, both based on dilution factors, were compared to a Bayesian D-optimal design approach (short called: Bayesian design) from optimal design theory. Furthermore, the different design approaches were compared for different numbers of concentrations, i.e. four to seven, and different numbers of technical replicates at the specific concentrations, i.e. three or six each. The precision of the considered designs was then measured for the new model fit of each simulation step by comparing this model fit to the reference curve via the Root Mean Squared Error (short: RMSE).

The results clearly show that the Bayesian optimal design approach leads to the highest precision to the reference curve for all numbers of concentrations. Additionally, the Bayesian design with seven concentrations shows also the highest precision of the alert concentration $EC_{50}$ compared to the reference $EC_{50}$. The log-equidistant design leads to a much less precise estimation and high variability when incorporating four to six concentrations. It improves when the seventh concentration is added. The design based on biological experience is also inferior to the Bayesian design approach in terms of general precision to the reference curve and a precise estimation of the $EC_{50}$.

Besides, we analyzed how many observations are needed for the different designs to achieve a precise estimation of the $EC_{50}$. For all numbers of concentrations the Bayesian design requires the smallest sample size, so the smallest number of technical replicates at each concentration, while the biologically motivated design and the log-equidistant design require more than 2 or 4 times more observations in total to achieve the same precision, respectively.

In addition to the analysis of the different design approaches in general, we investigated with another simulation study, how a sequential setup of an experiment influences the precision of the different design approaches. Here, the sequential approach consists of a pre-experiment and a main experiment. For the pre-experiment, all three different design approaches are considered, while the main experiment is designed based on the information gathered from the pre-experiment using a Bayesian design, as this method has demonstrated the best results in previous analyses. The analysis demonstrates that the results are not improved by a sequential approach, if the experiment is planned in an optimal way in the first step already, i.e. for the VPA data set the Bayesian or the log-equidistant design with seven concentrations. If instead the design based on biological arguments is used, an improvement by the sequential approach is clearly visible. The sequential approach tends to be more advantageous when the design of the pre-experiment is suboptimal.

Moreover, we analyzed how different design approaches perform for new substances, where only limited prior knowledge is available based on other related substances but not for the substance itself. A large data set of 104 different but related substances was used to calculate a Bayesian design for the substance VPA, without knowledge of the substance itself. A theoretical simulation study then compared the ability to estimate the VPA dose-response relationship of the Bayesian design without prior knowledge of VPA to the three different design approaches. As well as in the scenario with pre-existing knowledge, the Bayesian design even without prior knowledge of VPA shows a higher precision than the design based on biological arguments, both in terms of general precision as well as the precision of the $EC_{50}$. Besides, we analyzed the performance of the Bayesian design based on the 104 substances for all substances considered in the data set. Our results show, that it outperforms the traditional used design approaches for the majority of substances in this scenario.

Based on both scenarios and their corresponding analyses, we developed a guideline for upcoming in vitro cytotoxicity experiments. To ensure an easy calculation of the optimal concentrations, we also provide a useful software tool in form of an R Shiny App. In this R Shiny App users can calculate optimal concentrations for upcoming cytotoxicity experiments themselves via providing basic prior knowledge, like e.g. the maximal considered concentration and information about the effective concentrations $EC_{10}$ and $EC_{50}$. Additionally the user is able to compare the calculated optimal concentrations to other concentration levels in terms of their quality with an interactive tool.

In summary, the first article highlights the importance of employing optimal design approaches in practice. Additionally, it establishes a crucial foundation for bridging the gap between theory and practical application by presenting the results in a user-friendly manner and offering valuable tools.

## 3.2. Article 2: Optimal designs for identifying effective doses in drug combination studies

The second article addresses optimal design theory for drug combination studies. In such drug-drug-interaction studies, the dose-response relationship of the single drug therapies are often well known, while the interaction effect remains unknown. There are different ways to describe the interaction effect of two substances, established methods are isobologram analysis or methods based on the combination index (see Lee et al. (2007) for details). Furthermore, response surface models as proposed by Zhou et al. (2024) can be used to model the response effect in drug combination studies. Due to the two-dimensional setup of such studies new challenges arise because of the more complex setup than in classical one-dimensional dose-response experiments.

For this research article response surface models of the following form are considered to model the common dose-combination relationship of two substances $C$ and $D$ with concentrations $c$ and $d$:

$$\eta\big((c,d),\theta\big) = \theta_0 + \eta_C(c,\theta_C) + \eta_D(d,\theta_D) + \gamma\eta_C(c,\theta_C)\eta_D(d,\theta_D)\,, \quad \theta = (\theta_0,\theta_C,\theta_D,\gamma)^T,$$

where $\theta_0$ corresponds to the common placebo effect, $\eta_C(\cdot,\theta_C)$ and $\eta_D(\cdot,\theta_D)$ denote the regression functions of the substances $C$ and $D$, respectively. The interaction effect is defined with $\gamma \in [-1,1]$, indicating a positive interaction of the substances for $\gamma > 0$, a negative interaction for $\gamma < 0$ and no interaction effect for $\gamma = 0$.

Due to the more complex model setup compared to classical one-dimensional dose-response analyses, this increased complexity also raises additional challenges in the design aspect. There is already research present for classical D-optimal designs for drug combination studies (Papathanasiou et al., 2019), which aims to achieve precise estimations for the entire dose-response relationship of both substances. Similar to one-dimensional dose-response analyses in such analyses also effective doses are of interest, indicating at which doses a specific amount of the maximal effect is achieved. In contrast to the one-dimensional setting, where the effective dose can be identified with one specific dose, in the dose-combination setting a set of multiple dose-combinations leads to a specific effect level of interest. The corresponding response effect at those effective doses results in contour lines on the response surface.

We propose a design criterion, called Multivariate Effective Dose (short: MED) criterion aiming for a precise estimation of those contour lines and therefore the precise estimation of the effective doses in the two-dimensional drug combination setting. In detail, the new developed design approach minimizes the asymptotic variance at contour lines of interest. We establish the design criterion for $k \geq 1$ contour lines of interest. Besides, we provide an equivalence theorem and efficiency bounds for validating the optimality of MED designs in practice.

Furthermore, we compare the performance of the newly developed MED design approach to typically established design approaches in practice. Here, factorial designs and ray

designs, as well as D-optimal design approaches, are evaluated against the MED design through efficiencies for a theoretical comparison and across various simulation scenarios. For both, a real case study is used as a basis, while the simulation study was also extended with additional scenarios. All scenarios vary in their dose-combination relationship, where different dose-response models can be incorporated, as well as their contour levels of interest.

In detail, the theoretical comparison via efficiencies based on the case study shows the superiority of the MED design to the other designs. Especially the considered factorial and ray designs lead to an inferior performance, although the considered factorial design uses even more dose combination levels than the MED design. The D-optimal design shows a reasonable performance depending on the choice of contour levels.

Additionally to the theoretical comparison, the performance of the design approaches was also investigated in practical simulation studies. In particular, different dose-combination relationships as well as different effect levels of interest were considered in three different scenarios: One with no visible interaction effect within the designs space, a second one with a clear positive interaction effect, resulting in a monotonic dose-combination relationship, and a third scenario with a negative interaction effect. The last scenario exhibits a non monotonic dose-combination relationship and to that effect a more complex structure of the contour lines of interest.

The different considered designs exhibit varying performances depending on the simulation scenario. Each design approach is compared using an equal number of observations in each simulation scenario to ensure comparability, while the number and selection of dose combination levels varies among them based on the design approach. In all scenarios the MED design approach outperforms the other considered design approaches. In cases with monotonic response surfaces within the design space, e.g. the first and second scenario, the D-optimal design showed a worse performance compared to the locally MED-optimal design, but it was also able to estimate the multivariate effective doses of interest accurately. In the situation where contour lines were achieved only at high dose combinations, the D-optimal design struggled to capture effects as effectively as the locally MED-optimal design. Furthermore, the considered ray designs show a clearly worse performance in estimating the effect levels than the MED design in most scenarios.

However, in the third scenario, which assumes a negative interaction effect resulting in contour lines divided in distinct sets, the considered ray designs as well as the considered factorial design show a reasonable performance. For this scenario, the D-optimal design demonstrates an inferior ability to estimate the contour lines of interest compared to all other design approaches, even for higher sample sizes.

Additionally, we propose a robust design approach for our new developed MED criterion. Often, dose-combination experiments are conducted for two drugs, which are well-known before the experiment, but their interaction effect remains unknown. Therefore, we investigated the performance of the locally MED-optimal designs as well as a Bayesian version for different values of the interaction effect $\gamma$. Our results actively demonstrate, using the robust Bayesian approach, when the interaction effect is uncertain before conducting the experiment and the contour lines of interest vary depending on the interaction effect $\gamma$. In this case a misspecification of the interaction effect can lead to less effectively estimating the contour lines of interest. If sufficient prior knowledge of the interaction effect is known beforehand or the multivariate effective doses of interest are not affected much by the value of $\gamma$, a locally MED-optimal design is sufficient even if the interaction effect is slightly misspecified.

Overall, the second article introduces a novel design approach for drug-combination studies and addresses the previously mentioned second research gap by presenting new methodologies relevant to current drug-combination research. The proposed design approach for estimating the multivariate effective doses outperforms the established design approaches in various scenarios. Additionally, it is highly flexible and can be applied to various settings, involving different interaction effects and diverse dose-response shapes for the individual substances. We also provide code for practitioners to facilitate its implementation.

## 3.3. Article 3: Designs for the simultaneous inference of concentration–response curves

The third article focuses on optimal experimental designs for high-dimensional gene expression data. In these analyses, large numbers of genes are evaluated simultaneously, which represents new challenges in the design process. Specifically, it considers the context of microarray analyses, where thousands of genes are assessed concurrently on a single gene chip. The challenge lies in the fact that theoretically, each gene would require its own optimal design, i.e. individual concentration levels. However, the experimental setup with all genes on one chip does not allow for different settings for each gene. Instead, it requires a single set of concentrations to analyze all genes at once. This raises the question of how to select concentrations for such microarray analyses to accurately estimate the concentration-response relationship for the majority of genes. This article contributes to the field of optimal design of experiments by proposing a specific design criterion that aims for a precise estimation of all considered genes simultaneously.

Therefore, we developed a new design criterion based on the concept of Bayesian optimal designs (see Section 2.2 for details) for a precise estimation of a various number of concentration-response curves. In our criterion a prior distribution $\pi$ needs to be specified, which represents weights for different concentration-response relationships with incorporating their nonlinear concentration-response parameter estimates $\boldsymbol{\theta}$ in the set $\Theta_G$ of all considered concentration-response curves. The design for the simultaneous inference is then defined as:

$$\psi(\xi, \pi) = \sum_{\theta \in \Theta_G} \pi(\theta) \mathrm{Eff}_D\left(\xi, \theta\right), \tag{3.1}$$

aiming for high $D$-efficiency values on all considered concentration-response curves with corresponding parameters in $\Theta_G$.

Furthermore, another new design approach based on the cluster algorithm $k$-means is proposed. In this method, the optimal concentrations from the considered concentration-response curves are used to compute a common design for all curves collectively by clustering these optimal concentrations.

The newly developed design approaches are compared to other existing design methods both theoretically, in terms of efficiencies, and practically, through a simulation study based on a case study by Krug et al. (2013). Here, 15 233 genes evaluated for the substance valproic acid (VPA) are considered in the analysis. To describe each gene's dose-response relationship a four-parametric log-logistic (4pLL) model was fitted for each gene.

In this analysis, the focus is on evaluating the performance of different design approaches in accurately assessing the majority of genes. Established design approaches, such as equidistant and log-equidistant designs, are included in the comparison alongside the

original design used in the case study. The newly developed criterion for the simultaneous inference was used to calculate an optimal design, called simultaneous inference design, which was constructed based on a prior distribution of all considered genes in the data set. For that matter, in a first step the dose-response relationship of all considered genes was investigated in terms of their $EC_{50}$ values and the steepness of the corresponding curve. It was analyzed which parameter combinations of the genes were highly represented in the data set based on a $5 \times 5$ division for the parameter combinations of interest. Here, the majority of all genes showed parameters combinations in only seven of 25 considered areas. Consequently, one representative gene was included for each considered area in the parameter set $\Theta_G$. Additionally, the corresponding weights of the corresponding prior distribution $\pi$ were calculated based on the relative frequency of genes in those areas.

The simultaneous inference design, which was calculated based on the criterion in (3.1), outperformed all other considered design approaches in the theoretical as well as the practical evaluations. While other methods, particularly the $k$-means approach and the equidistant design, demonstrated competitive performance relative to the simultaneous inference design, the original design performed worse, and the log-equidistant design exhibited very limited effectiveness. Although the simultaneous inference design for the case study was based solely on seven representative genes, it outperformed the other designs, highlighting a significant advantage over the $k$-means approach, which incorporates all optimal designs for all considered genes.

This article clearly demonstrates the importance of incorporating optimal design theory in high-dimensional settings. We present a criterion for the simultaneous inference of multiple concentration-response curves. A significant advantage of this newly developed criterion is its flexibility, allowing it to be adapted to various contexts. Furthermore, the design criterion for the simultaneous inference can be employed if specific gene groups are of interest by incorporating specific information in the prior distribution for the design criterion.

# 4. Discussion and Outlook

In this thesis, three manuscripts dealing with optimal design theory in the context of dose-response analysis in toxicology were presented. In detail, different aspects of optimal design theory regarding dose-response analysis are investigated. The first manuscript investigates classical cytotoxicity experiments focused on the effects of a single substance. The second manuscript explores the two-dimensional case and its corresponding optimal designs for the combination of two substances. Finally, the last paper examines high-dimensional cases related to gene expression analysis of one substance but a large number of genes. Each project presents its own challenges and opportunities for further development.

The first paper highlights the importance of considering optimal designs in practice. A Bayesian design approach from optimal design theory is compared to traditionally used designs in practice, like log-equidistant designs or designs based on biological arguments. The superiority of the Bayesian design is demonstrated based on an extensive case study of valproic acid (VPA). In detail, this large data set is used to specify a reference curve by fitting a four-parametric log-logistic (4pLL) model. The 4pLL model was selected based on prior knowledge of the dose-response relationship of VPA and its status as the most widely used model in dose-response analyses (Goutelle et al., 2008). Here, other model choices are also possible. Additionally, one could consider model selection techniques like the multiple comparison and modeling approaches proposed by Bretz et al. (2005) or even model averaging methods (Claeskens and Hjort, 2008). In this context, model selection involves choosing a specific model from a set of predefined candidate models, while model averaging results in a weighted sum of prespecified candidate models. In case of model selection, the prior knowledge which serves as a basis for the Bayesian optimal design, would change according to the selected model. However, the methods presented in the paper can be easily adapted to accommodate different model choices. When using model averaging to specify the reference curve for VPA, the process becomes slightly more complex but can still be adapted via incorporating this in an adjusted prior for a Bayesian design or with using related methods from optimal design theory, such as composite designs (see e.g. Pukelsheim, 2006). Nonetheless, the advantages of optimal designs should also be evident with these alternative approaches, as the theory of optimal design has already been shown to yield more precise estimations for optimal designs compared to non-optimal ones.

Furthermore, the viability of the data points is assumed to be characterized with homoscedastic errors across all concentrations when fitting the dose-response relationship in the first article. Typically, variability tends to be higher at lower concentrations and decreases noticeably at higher concentrations. One could incorporate this via using mixed-effect models and consider optimal design theory based in this context (see e.g. Seurat et al., 2021). In order to illustrate the idea of optimal experimental designs in general and bridge the gap between theory and practice without introducing excessive complexity we decided to provide information for the most commonly used and known model. For similar reasons, we also decided to base all concepts on the D-optimality criterion. Other criteria like $A$-optimality or $c$-optimality (Pukelsheim, 2006) could also be relevant in such contexts (see e.g. Holland-Letz, 2017).

The second article provides a new design criterion, which enables a precise estimation of multivariate effective doses in drug-combination studies. Here, the criterion is developed based on the assumption of response-surface modeling, where the common effect is expressed via the response effect of the single substances and an additional interaction effect of the product of both effects. This enables a precise estimation of the response effect within the whole designs space, not only at the measured dose-combinations. Nevertheless, there are various methods to model the joint effect of two substances, such as isobolograms or combination indices (see e.g., Zhao et al. (2010) and Foucquier and Guedj (2015)). While these methods lack the ability to estimate the effects for all possible combinations of both substances, they are less complex due to their reduction to a one-dimensional linkage. For instance, Holland-Letz and Kopp-Schneider (2018) provide optimal design theory for the combination index of two substances. Additionally, the joint effect of substances can be modeled using one-dimensional dose-response models for fixed proportions of both substances. Holland-Letz et al. (2020) demonstrate that the combined effect can be approximated using commonly employed log-logistic models, provided that either full additivity or a reasonably simple interaction can be assumed for the mixture. This approach could also serve as a basis for investigating mixtures involving more than two substances.

Due to the complexity of the response surface model assumed in Article 2, extending the approach to accommodate higher numbers of substances may not be feasible. However, it offers the explicit advantage of estimating the response effects in detail across the entire design space and therefore enables the estimation of all possible multivariate effective dose combinations. Furthermore, based on its two-dimensional setting and its flexibility, the assumed model can be adapted to other contexts where relationships can be expressed through regression functions. Consequently, optimal designs could be explored also for specific effects based on dose and exposure time as two influencing variables in the proposed approach in Article 2.

The third article demonstrates the relevance of considering optimal design approaches also for high-dimensional settings. Here, microarray gene expression analyses are considered, where thousand of genes are evaluated simultaneously on gene chips. This kind of analyses claim one design for the entire analysis, although each gene could be evaluated most precisely with its own optimal design. Therefore, we developed a design for the simultaneous inference of a large number of concentration-response curves, aiming for a precise estimation of the majority of all considered genes. We also analyzed the performance of the simultaneous inference design compared to traditional used designs like equidistant and log-equidistant designs with theoretical comparisons via efficiencies and practical comparisons via a simulation study, both based on a case study proposed by Krug et al. (2013). For both comparisons, the theoretical and practical analysis, the simultaneous design showed superior performance.

The developed criterion for the simultaneous inference is based on prior knowledge of the concentration-response relationship of the genes. Therefore, a sufficient prior knowledge is required for the calculation of the simultaneous inference design, which may limit its usage. Nevertheless, if prior knowledge is available, e.g. for reproducibility analyses or based on results from literature, the design leads to more precise results than traditionally used designs. For the example of VPA it especially outperformed the log-equidistant design.

Furthermore, this paper only includes genes in the analysis that exhibited a reasonable dose-response signal modeled using the 4pLL model. Similar to the one-dimensional analysis in the first article, this could be extended to various dose-response shapes via model

selection or model averaging techniques (see again Claeskens and Hjort, 2008). The proposed criterion for simultaneous inference of concentration-response curves can be adapted to incorporate different dose-response models, demonstrating its flexibility in this context. Additionally, the structure of the criterion allows for emphasizing genes with specific parameter constellations. In this context, it would be interesting to analyze the performance of the criterion when specific gene groups are of interest.

Recently, another method known as RNA sequencing has emerged for analyzing gene expression data. This technique measures the entire transcriptome, which includes all transcripts derived from DNA. In contrast, microarray analyses focus only on predefined genes. Rao et al. (2019) highlights the advantages and disadvantages of both methods. Although the benefits of RNA sequencing are considerable, including its ability to provide a more comprehensive analysis of the entire transcriptome, this also results in much larger data sets. Additionally, there is a lack of optimized and standardized protocols for RNA sequencing (see again Rao et al. (2019)). Therefore, microarray analyses remain relevant today and are addressed within this thesis, even though RNA sequencing has become the predominant method for gene expression analysis.

Summarizing, all three proposed articles provide a strong foundation for advancing research in both statistics and toxicology. The relevance of optimal designs of experiments is demonstrated, ranging from one-dimensional experiments to high-dimensional analyses. In addition, further interesting ideas for future research projects have emerged from each of the projects.

# Bibliography


Atashpaz-Gargari, E., & Lucas, C., (2007): Imperialist competitive algorithm: An algorithm for optimization inspired by imperialistic competition, *2007 IEEE congress on evolutionary computation*, 4661–4667, https://doi.org/10.1109/CEC.2007.4425083

Baldi, P., & Hatfield, G. W., (2011): *DNA microarrays and gene expression: From experiments to data analysis and modeling*, Cambridge university press.

Berg, J. M., Tymoczko, J. L., Gatto, G. J., & Stryer, L., (2018): *Stryer Biochemie* (Vol. 8), Springer Berlin Heidelberg, https://doi.org/10.1007/978-3-662-54620-8

Bödeker, W., & Moebus, S., (2024): Das Paracelsus-Prinzip „Allein die Dosis macht, dass ein Ding (k) ein Gift sei"–ein nicht mehr zeitgemäßer Lehrsatz und ätiologischer Mythos, *Prävention und Gesundheitsförderung*, *19*(2), 165–170, https://doi.org/10.1007/s11553-023-01037-4

Bornkamp, B., Pinheiro, J., Bretz, F., Sandig, L., & Thomas, M., (2024): *DoseFinding: Planning and Analyzing Dose Finding Experiments* [R package version 1.2-1], https://CRAN.R-project.org/package=DoseFinding

Brecklinghaus, T., Albrecht, W., Duda, J., Kappenberg, F., Gruendler, L., Edlund, K., Marchan, R., Ghallab, A., Cadenas, C., Rieck, A., Vartak, N., Tolosa, L., Castell, J. V., Gardner, I., Halilbasic, E., Trauner, M., Ullrich, A., Zeigerer, A., Demirci Turgunbayer, Ö., ... Hengstler, J. G., (2022): In vitro/in silico prediction of drug induced steatosis in relation to oral doses and blood concentrations by the Nile Red assay, *Toxicology Letters*, *368*, 33–46, https://doi.org/10.1016/j.toxlet.2022.08.006

Bretz, F., Pinheiro, J. C., & Branson, M., (2005): Combining multiple comparisons and modeling techniques in dose-response studies, *Biometrics*, *61*(3), 738–748, https://doi.org/10.1111/j.1541-0420.2005.00344.x

Casella, G., & Berger, R., (2024): *Statistical inference*, CRC press.

Chaloner, K., & Larntz, K., (1989): Optimal Bayesian design applied to logistic regression experiments, *Journal of Statistical Planning and Inference*, *21*(2), 191–208, https://doi.org/10.1016/0378-3758(89)90004-9

Chen, P.-Y., Chen, R.-B., & Wong, W. K., (2022): Particle swarm optimization for searching efficient experimental designs: A review, *Wiley interdisciplinary reviews: computational statistics*, *14*(5), e1578, https://doi.org/10.1002/wics.1578

Chernoff, H., (1953): Locally optimal designs for estimating parameters, *Annals of Mathematical Statistics*, *24*, 586–602, https://doi.org/10.1214/aoms/1177728915

Chou, T.-C., (2008): Preclinical versus clinical drug combination studies, *Leukemia & Lymphoma*, *49*(11), 2059–2080, https://doi.org/10.1080/10428190802353591

Claeskens, G., & Hjort, N. L., (2008): *Model Selection and Model Averaging*, Cambridge University Press.

Clerc, M., (2006): *Particle swarm optimization*, ISTE.

Dette, H., Bretz, F., Pepelyshev, A., & Pinheiro, J., (2008): Optimal Designs for Dose-Finding Studies, *Journal of the American Statistical Association*, *103*(483), 1225–1237, https://doi.org/10.1198/016214508000000427

Fedorov, V. V., & Leonov, S. L., (2013): *Optimal design for nonlinear response models*, CRC Press.

Foucquier, J., & Guedj, M., (2015): Analysis of Drug Combinations: Current Methodological Landscape, *Pharmacology Research & Perspectives*, *3*(3), https://doi.org/10.1002/prp2.149

Goutelle, S., Maurin, M., Rougier, F., Barbaut, X., Bourguignon, L., Ducher, M., & Maire, P., (2008): The Hill equation: a review of its capabilities in pharmacological modelling, *Fundamental & clinical pharmacology*, *22*(6), 633–648, https://doi.org/10.1111/j.1472-8206.2008.00633.x

Holland-Letz, T., & Kopp-Schneider, A., (2021): An R-shiny application to calculate optimal designs for single substance and interaction trials in dose response experiments, *Toxicology letters*, *337*, 18–27, https://doi.org/10.1016/j.toxlet.2020.11.018

Holland-Letz, T., & Kopp-Schneider, A., (2018): Optimal experimental designs for estimating the drug combination index in toxicology, *Computational Statistics & Data Analysis*, *117*, 182–193, https://doi.org/10.1016/j.csda.2017.08.006

Holland-Letz, T., (2017): On the combination of c-and D-optimal designs: General approaches and applications in dose–response studies, *Biometrics*, *73*(1), 206–213, https://doi.org/10.1111/biom.12545

Holland-Letz, T., Leibner, A., & Kopp-Schneider, A., (2020): Modeling dose–response functions for combination treatments with log-logistic or Weibull functions, *Archives of toxicology*, *94*, 197–204, https://doi.org/10.1007/s00204-019-02631-2

Hothorn, L., (2016): *Statistics in toxicology using R*, CRC Press Boca Raton, FL.

Jennrich, R. I., (1969): Asymptotic properties of non-linear least squares estimators, *The Annals of Mathematical Statistics*, *40*(2), 633–643, https://doi.org/10.1214/aoms/1177697731

Kappenberg, F., (2021): *Statistical approaches for calculating alert concentrations from cytotoxicity and gene expression data* [Doctoral dissertation, Dissertation, Dortmund, Technische Universität, 2021].

Kappenberg, F., Duda, J. C., Schürmeyer, L., Gül, O., Brecklinghaus, T., Hengstler, J. G., Schorning, K., & Rahnenführer, J., (2023): Guidance for statistical design and analysis of toxicological dose–response experiments, based on a comprehensive literature review, *Archives of Toxicology*, 1–21, https://doi.org/10.1007/s00204-023-03561-w

Kennedy, J., & Eberhart, R., (1995): Particle swarm optimization, *Proceedings of ICNN'95-international conference on neural networks*, *4*, 1942–1948, https://doi.org/10.1109/ICNN.1995.488968

Kiefer, J., (1974): General Equivalence Theory for Optimum Designs (Approximate Theory), *The Annals of Statistics*, *2*(5), 849–879, https://doi.org/10.1214/aos/1176342810

Krug, A. K., Kolde, R., Gaspar, J. A., Rempel, E., Balmer, N. V., Meganathan, K., Vojnits, K., Baquié, M., Waldmann, T., Ensenat-Waser, R., Jagtap, S., Evans, R. M., Julien, S., Peterson, H., Zagoura, D., Kadereit, S., Gerhard, D., Sotiriadou, I., Heke, M., … Sachinidis, A., (2013): Human embryonic stem cell-derived test systems for developmental neurotoxicity: A transcriptomics approach, *Archives of toxicology*, *87*(1), 123–143, https://doi.org/10.1007/s00204-012-0967-3

Lee, J. J., Kong, M., Ayers, G. D., & Lotan, R., (2007): Interaction index and different methods for determining drug interaction in combination therapy, *Journal of biopharmaceutical statistics*, *17*(3), 461–480, https://doi.org/10.1080/10543400701199593

Macdougall, J., (2006): Analysis of dose–response studies—Emax model. In *Dose finding in drug development* (pp. 127–145), Springer.

Masoudi, E., Holling, H., & Wong, W. K., (2017): Application of imperialist competitive algorithm to find minimax and standardized maximin optimal designs, *Computational Statistics & Data Analysis*, *113*, 330–345, https://doi.org/10.1016/j.csda.2016.06.014

Mielke, T., (2016): Study designs for the estimation of the hill parameter in sigmoidal response models, *mODa 11-Advances in Model-Oriented Design and Analysis: Proceedings of the 11th International Workshop in Model-Oriented Design and Analysis held in Hamminkeln, Germany, June 12-17, 2016*, 183–190, https://doi.org/10.1007/978-3-319-31266-8_21

Papathanasiou, T., Strathe, A., Overgaard, R. V., Lund, T. M., & Hooker, A. C., (2019): Optimizing dose-finding studies for drug combinations based on exposure-response models, *The AAPS Journal*, *21*, 1–11, https://doi.org/10.1208/s12248-019-0365-3

Paracelsus, T., (1965): Die dritte Defension wegen des Schreibens der neuen Rezepte, *Septem Defensiones*, *1538*, 508–513.

Proctor, W. R., Foster, A. J., Vogt, J., Summers, C., Middleton, B., Pilling, M. A., Shienson, D., Kijanska, M., Ströbel, S., Kelm, J. M., Morgan, P., Messner, S., & Williams, D., (2017): Utility of spherical human liver microtissues for prediction of clinical drug-induced liver injury, *Archives of Toxicology*, *91*, 2849–2863, https://doi.org/10.1007/s00204-017-2002-1

Pronzato, L., & Walter, E., (1985): Robust experiment design via stochastic approximation, *Mathematical Biosciences*, *75*(1), 103–120, https://doi.org/10.1016/0025-5564(85)90068-9

Pukelsheim, F., (2006): *Optimal design of experiments*, SIAM.

Rao, M. S., Van Vleet, T. R., Ciurlionis, R., Buck, W. R., Mittelstadt, S. W., Blomme, E. A., & Liguori, M. J., (2019): Comparison of RNA-seq and microarray gene expression platforms for the toxicogenomic evaluation of liver from short-term rat toxicity studies, *Frontiers in genetics*, *9*, 636, https://doi.org/10.3389/fgene.2018.00636

Ritz, C., Baty, F., Streibig, J. C., & Gerhard, D., (2015): Dose-Response Analysis Using R, *PLOS ONE*, *10*, https://doi.org/10.1371/journal.pone.0146021

Seurat, J., Tang, Y., Mentré, F., & Nguyen, T. T., (2021): Finding optimal design in nonlinear mixed effect models using multiplicative algorithms, *Computer Methods and Programs in Biomedicine*, *207*, 106126, https://doi.org/10.1016/j.cmpb.2021.106126

Shi, Y., Zhang, Z., & Wong, W. K., (2019): Particle swarm based algorithms for finding locally and Bayesian D-optimal designs, *Journal of Statistical Distributions and Applications*, *6*(1), 3, https://doi.org/10.1186/s40488-019-0092-4

Silvey, S. D., (1980): *Optimal Design*, Chapman; Hall.

Ting, N., Chen, D.-G., Ho, S., & Cappelleri, J. C., (2017): *Phase II clinical development of new drugs*, Springer.

Vorrink, S. U., Zhou, Y., Ingelman-Sundberg, M., & Lauschke, V. M., (2018): Prediction of drug-induced hepatotoxicity using long-term stable primary hepatic 3D spheroid cultures in chemically defined conditions, *Toxicological Sciences*, *163*(2), 655–665, https://doi.org/10.1093/toxsci/kfy058

Zhao, L., Au, J. L., & Wientjes, M. G., (2010): Comparison of methods for evaluating drug-drug interaction, *Frontiers in bioscience (Elite edition)*, *2*, 241, https://doi.org/10.2741/e86

Zhou, Y., Sloan, A., Menon, S., & Wang, L., (2024): Combination MCP-mod for two-drug combination dose-ranging studies, *Journal of Biopharmaceutical Statistics*, *35*(2), 257–270, https://doi.org/10.1080/10543406.2024.2311254

# Part II.

# Publications

# 1. Article: Design of optimal concentrations for in vitro cytotoxicity experiments

Archives of Toxicology
https://doi.org/10.1007/s00204-024-03893-1

IN VITRO SYSTEMS

# Design of optimal concentrations for in vitro cytotoxicity experiments

**Leonie Schürmeyer[1] · Chen Peng[2,3] · Wiebke Albrecht[2] · Tim Brecklinghaus[2] · Pauline Baur[1] · Jan G. Hengstler[2] · Kirsten Schorning[1]**

Received: 14 August 2024 / Accepted: 14 October 2024


## Abstract

Concentration-dependent cytotoxicity experiments are frequently used in toxicology. Although it has been reported that an adequate choice of concentrations improves the quality of the statistical inference substantially, a recent literature review of three major toxicological journals has shown that the corresponding methods are rarely used in toxicological practice. In this study the performance of different sets of concentrations, also called designs, are analyzed, while the overall goal is to promote the advantages of optimal design procedures and to present a user-friendly guideline for planning new cytotoxicity concentration-response experiments. We compare the frequently used log-equidistant design to a Bayesian design, which is constructed by methods of optimum design theory. Using both a dense data set of concentration-cytotoxicity data of valproic acid (VPA) and regular assay data of 104 substances, the performance of the different designs is analyzed in two scenarios, where detailed previous knowledge on VPA is available or not. The results show that it is critical to apply a specific design strategy to determine optimal concentrations for cytotoxicity testing. In particular, the Bayesian design technique with and without incorporating pre-existing knowledge of a specific test substance resulted in a more precise statistical inference than the other used designs. Finally, we present a guideline for upcoming experiments and an accessible user-friendly Shiny app (see http://shiny.statistik.tu-dortmund.de:8080/app/occe).

**Keywords** Optimal design · *D*-optimal design · Bayesian optimal design · Concentration-response experiments

## Introduction

Concentration-dependent cytotoxicity tests are frequently used in toxicology (Vorrink et al. 2018; Brecklinghaus et al. 2022a; Proctor et al 2017; Khetani et al. 2013; Gu et al. 2018). These tests usually serve to determine the $EC_{50}$-value as the concentration of a substance that reduces vitality to 50% of solvent controls. In further studies, also the lowest concentrations that begin to cause cytotoxicity are analyzed by determining $EC_{10}$-values (Brecklinghaus et al. 2022b; Ghallab et al 2022). An important challenge of in vitro cytotoxicity testing is the choice of adequate test concentrations. Here, two scenarios should be differentiated, depending

Jan G. Hengstler and Kirsten Schorning contributed equally to this work.

✉ Leonie Schürmeyer
schuermeyer@statistik.tu-dortmund.de

Chen Peng
cp9109@163.com

Wiebke Albrecht
albrecht@ifado.de

Tim Brecklinghaus
brecklinghaus@ifado.de

Pauline Baur
baur@statistik.tu-dortmund.de

Jan G. Hengstler
hengstler@ifado.de

Kirsten Schorning
schorning@statistik.tu-dortmund.de

[1] Department of Statistics, TU Dortmund University, Vogelpothsweg 87, Dortmund 44227, North Rhine-Westphalia, Germany

[2] Leibniz Research Centre for Working Environment and Human Factors at the Technical University of Dortmund (IfADo), Ardeystraße 67, Dortmund 44139, North Rhine-Westphalia, Germany

[3] Department of Microbiology, School of Medicine, Jiangsu University, Zhenjiang 212013, China

Published online: 16 November 2024

on the availability of previous established $EC_{10}$- and $EC_{50}$ -values.

If $EC_{10}$- and $EC_{50}$-values are known from previous studies this can be used to plan the cytotoxicity experiment as a one-step procedure. We compare three procedures to select sets of experimental concentrations, also called designs, when prior knowledge on the $EC_{10}$ and $EC_{50}$ is available. First, the log-equidistant design technique can be applied, where the highest solubility or maximal concentration is multiplied or divided by the same factor. Second, the log-equidistant design technique can be combined with biological arguments so that, e.g., around the $EC_{50}$-value smaller distances of concentrations are chosen. Third, the pre-existing knowledge about the $EC_{10}$- and $EC_{50}$-value can serve as a basis for a determination of optimal test concentrations using a (pseudo) Bayesian design technique (see Chaloner and Larntz 1989). This technique has been developed within the broad research field of optimal design of experiments (ODE), which provides methods on planning and conducting experiments in a way that maximizes the efficiency of the data-generating process. Hereby, the major goal is to obtain the most informative and precise statistical results with the smallest necessary resources. In the context of toxicological experiments, this means identifying the smallest necessary number of different concentrations, their allocations and the necessary number of replicates.

Despite these advantages, the techniques provided by ODE, in particular the Bayesian technique, are rarely used for planning toxicological experiments. Instead, the log-equidistant technique is frequently applied to obtain sets of test concentrations in practice (see Kappenberg et al. 2023). This is also due to the fact, that the influence of the test concentrations on the quality of data and thus on the quality of the statistical analysis is usually unknown. As a result, the design of test concentrations is often neglected during the planning of an experiment (see again Kappenberg et al. 2023). Besides, the limited availability of online tools for the calculation of optimal test concentrations based on Bayesian or other ODE techniques makes its application difficult for practitioners (Holland-Letz and Kopp-Schneider 2021).

To compare the performance of the three different design techniques in concentration-cytotoxicity experiments, reference cytotoxicity data for a single test substance was produced with an unusually high number of concentrations and biological replicates. The test substance under consideration was valproic acid (VPA). VPA is a medium soluble compound with well-documented human hepatotoxicity and high clinical relevance. Besides it has been used as a positive control in previous studies (Albrecht et al. 2019). The curve fitted to this atypically large data set is assumed to represent the true concentration-cytotoxicity relationship of VPA, hereafter referred to as reference curve. Then the quality of each considered design was assessed by comparing it against the reference curve, thereby using only a subset of the full VPA data set that corresponds to the design.

An additional research question of this study was the analysis of a sequential approach with a pre-experiment followed by a main experiment. Here, the data from the pre-experiment serve as prior information for the determination of the design that should be used in the main experiment. Each of the considered design techniques was used to plan the pre-experiment. Based on the corresponding subsets of the large data set, it was investigated, if the experiment could be improved by the sequential approach. Moreover, it was analyzed which of the design techniques led to the best results if used as a design in the pre-experiment.

So far, the proposed design techniques all require pre-existing knowledge about the properties of the considered test substance. However, this information may not be available in some situations, for example when a new substance is tested. To address such research situations, a second scenario was considered where available data of other, related test compounds were used as less specific prior knowledge. More precisely, already existing concentration-dependent cytotoxicity data of 104 other compounds that had been tested by the same cytotoxicity assay were used to construct (less specific) prior information about the test substance. This prior information was used to apply the different design techniques. In particular, a Bayesian design was constructed that could be used both for experiments on the test substance VPA itself and for experiments on all other considered substances.

Finally, we present a guideline with easy-to-apply methods for one-step and sequential strategies of toxicity testing. Furthermore, we provide a Shiny app for a user-friendly calculation of optimal concentrations for upcoming experiments.

## Materials and methods

### Test compounds

For this study a compound set with a total of 104 compounds, composed mostly of pharmaceuticals, was utilized. Details on the compounds, supplier, utilized solvents and concentrations are provided in Supplement Data.

### Cytotoxicity testing in HepG2 cells

The HepG2 cell line (ATCC number: HB-8065™) was used for the cytotoxicity tests as described previously (Brecklinghaus et al. 2022a; Albrecht et al. 2019). Briefly, the cells were cultured in 4.5 g/L glucose Dulbecco's modified eagle's medium (PAN Biotech GmbH, Aidenbach, Germany P04-04500) supplemented with 10% heat inactivated fetal calf

serum (PAN Biotech GmbH, Aidenbach, Germany 3702-P103009) and 100 U/ml Penicillin/0.1 mg/ml Streptomycin (PAN Biotech GmbH, Aidenbach, Germany P06-07100). Approximately 15 000 cells were seeded in black 96-well plates (Greiner bio-one, Frickenhausen, Germany, REF 655986) coated with 0.25 mg/ml rat tail collagen (Roche Diagnostic Mannheim, 10 mg, Cat. No. 11171179001). No cells were seeded into the outermost columns and rows. These wells were filled with phosphate buffered saline to counteract evaporation effects. After 16–20 h incubation at 37 °C and 5% $CO_2$ the medium was exchanged and exposure to the test compounds was started. For each test compound a vehicle control was included. The cells were exposed to the test compounds for 48 h at 37 °C and 5% $CO_2$. At the end of the exposure period, the medium was removed, the cells were washed three times with phosphate buffered saline and 100 $\mu$l/well of 1:5 CellTiter-Blue (short: CTB) reagent (Promega, Cat. No. G8081) in medium were added. As a background control wells with CTB mixture without cells was included. The cells were further incubated at 37 °C and 5% $CO_2$ until a color change from blue to purple was observed in the vehicle controls. The fluorescence was measured with the Tecan Infinite M200 Pro plate reader (software i-control, version 1.7.1.12) utilizing the excitation wavelength 540 nm and emission wavelength 594 nm. Prior to curve fitting the mean of the background control wells was subtracted from all samples. Each compound was tested in at least three biological replicates with at least three technical replicates. The raw data for all compounds are provided in Supplement Data. In the following analysis the dense data set of VPA is referred to as VPA data set, besides the collection of data of 104 compounds investigating Drug-induced liver injury (short: DILI) effects is referred to as DILI data set. A part of the DILI data set was already published before (Albrecht et al. 2019). Since the blue resazurin contained in the CellTiter-Blue reagent is metabolized by vital cells to the pink resorufin, the concentration dependent change in the measured fluorescence corresponds to a change in viability of the cells.

## Statistical methods

All analyses were performed using the statistical software R, version 4.2.2 (R Core Team 2022) and Julia, version 1.9.3 (Bezanson et al. 2017), especially the Julia package Kirstine.jl (Sandig 2024).

### Modelling cytotoxic concentration-response data

The concentration-response relationships of the substances investigated within the paper are expected to follow a sigmoidal course, so that the corresponding concentration-response data is always fitted to a four-parametric log-logistic model, hereinafter also referred to as 4pLL-model. Following Ritz et al. (2015), a 4pLL-model consists of four parameters and is defined by the non-linear regression function

$$\eta(x,\theta) = c + \frac{d-c}{1+\exp\left(b(\log(x)-\log(e))\right)}, \quad \theta = (b,c,d,e), \tag{1}$$

where $x$ is a concentration within the concentration range $\mathcal{X} = [0, x_{\max}]$. The lower and upper asymptotes are represented by the parameters $c$ and $d$, whereas the parameter $b$ describes the steepness of the curve. Furthermore, the parameter $e$ describes the turning point of the curve, i.e. the concentration on the S-shaped curve halfway between $c$ and $d$.

Let now $\{x_1, \ldots, x_K\}$ be a set of selected concentrations, where $x_k \in \mathcal{X} = [0, x_{\max}]$ and $K \in \mathbb{N}$. This set will be denoted by design in the following. We assume that $n$ replicates $y_{k1}, \ldots, y_{kn}$ are produced at each concentration $x_k$ of the design during the experiment such that the resulting concentration-response dataset is given by $\{(x_k, y_{kj}) | k = 1, \ldots, K, j = 1, \ldots, n\}$.

For the fit of a 4pLL model to the dataset, least squares (LS) estimation is used. More precisely, the LS estimate $\hat{\theta}$ is defined as the parameter value that minimizes the quadratic distance between the data and the function in (1), i.e.

$$\hat{\theta} \text{ minimizes } \sum_{k=1}^{K}\sum_{j=1}^{n}(y_{kj} - \eta(x_k,\theta))^2 .$$

The fitted 4pLL function is then given by $\eta(x, \hat{\theta})$. To measure the precision of the fitted 4pLL function, the covariance matrix of the LS-estimate $\hat{\theta}$, denoted by Cov $(\hat{\theta})$, is considered. As the direct calculation of this covariance matrix is difficult due to the non-linearity of the 4pLL function, an approximation is used. Following Jennrich (1969), the covariance matrix can be approximated by the inverse of the information matrix $M(x_1, \ldots, x_K, \theta)$ which is given by

$$M(x_1,\ldots,x_K,\theta) = \sum_{k=1}^{K} \frac{\partial}{\partial\theta}\eta(x_k,\theta)\frac{\partial}{\partial\theta}\eta(x_k,\theta)^{\top} . \tag{2}$$

Here, $\frac{\partial}{\partial\theta}\eta(x_k, \theta)$ denotes the gradient of the regression function, i.e. of the function given in (1), and $\theta$ is the unknown true parameter of the model. Consequently, the precision of the LS estimate $\hat{\theta}$ (and thus the precision of the model fit) depends both on the unknown parameter $\theta$ and the concentrations involved in the design. Note that the variance does not depend directly on the responses, but on the concrete model assumption made in advance. Therefore, no observations are needed to determine concentrations that result in a precise estimator of $\theta$ if prior knowledge of the model parameter $\theta$ is available. For the sake of simplicity, we first assume that the unknown parameter $\theta$ is given by a fixed

value $\theta_0 \in \mathbb{R}^4$. Then, the target of optimal design theory is to determine the design $\{x_1, \dots, x_K\}$, that results in the best precision of the LS estimate given that $\theta = \theta_0$. This optimization problem is equivalent to the maximization of the information matrix $M(x_1, \dots, x_K, \theta_0)$ with respect to $x_1, \dots, x_K$ and $K \in \mathbb{N}$. As matrices cannot be ordered, real-valued functions of the information matrix $M(x_1, \dots, x_K, \theta_0)$ have to be incorporated to solve the optimization problem. Different real-valued functions can be used here, whereas the most popular real-valued function is the determinant (see for instance Pukelsheim 2006). More precisely, a design is called locally $D$(eterminant)-optimal if it maximizes

$$\log(\det(M(x_1, \dots, x_K, \theta_0)))$$

among all designs on the concentration range $\mathcal{X} = [0, x_{\max}]$ (see Chernoff 1953 among many others). The $D$-optimal design minimizes the volume of the confidence ellipsoid for the complete parameter $\theta$ (see again Pukelsheim 2006). Kiefer (1974) proved that the $D$-optimal design also minimizes the maximum (asymptotic) variance of the estimated regression curve within $\mathcal{X}$, i.e. the expression

$$\max_{x \in \mathcal{X}} \frac{\partial}{\partial \theta} \eta^\top(x, \theta) M^{-1}(x_1, \dots, x_K, \theta) \frac{\partial}{\partial \theta} \eta(x, \theta) .$$

Consequently, optimizing based on the $D$-optimality criterion results in a design that has a good overall performance.

If a real-valued function of the parameter $\theta$, i.e. $\mu(\theta)$, is of special interest, optimality criteria can be defined to measure the approximate variance of the corresponding estimator $\mu(\hat{\theta})$. More precisely, a design is then called locally $c$-optimal (with $c = \frac{\partial}{\partial \theta} \mu(\theta)$) if it minimizes its asymptotic variance

$$\frac{\partial}{\partial \theta} \mu^\top(\theta) M^-(x_1, \dots, x_K, \theta) \frac{\partial}{\partial \theta} \mu^\top(\theta)$$

among all designs on the concentration range $\mathcal{X}$. Here $M^-$ denotes the generalized inverse of $M$. Although using $c$-optimal designs results in the precise estimation of $\mu(\theta)$, an estimation of the complete parameter $\theta$ might not be possible if this design is used. For instance, the $c$-optimal design for estimating the effect at concentration $x_0$ consists of one concentration, namely $x_0$. While the observations at concentration $x_0$ can be used to estimate $\eta(x_0, \theta)$, they are insufficient to fit the complete regression curve $\eta(x, \theta)$ (Silvey 1980). One possibility to overcome this problem is to consider compound optimality criteria (McGree et al. 2008) combine different optimality criteria, such as the $D$-optimality criterion and a $c$-optimality criterion of interest. However, the compound criteria can become very complex and the $c$-optimality criterion has to be defined individually. Due to these pitfalls, we decided to use the $D$-optimality criterion for the further analysis.

Note that $D$-optimal designs depend on the choice of the parameter value $\theta_0$ and are therefore called **locally** $D$-optimal designs. When less specific prior knowledge about the unknown parameter $\theta$ is available, the concept of locally $D$-optimal designs can be extended to Bayesian $D$-optimal designs (Chaloner and Larntz 1989). Assume prior information about $\theta$ is available in the form of different potential values, e.g., $\theta_1, \dots, \theta_P$. Then a design is called Bayesian $D$-optimal (for the prior $\theta_1, \dots, \theta_P$) if it maximizes the function

$$\sum_{j=1}^{P} \frac{1}{P} \log(\det(M(x_1, \dots, x_K, \theta_j)))$$

with respect to $x_1, \dots, x_K$ and $K \in \mathbb{N}$. Consequently, the resulting Bayesian $D$-optimal design maximizes the average of all potential determinants of the different possible information matrices. While we restrict ourselves to the case, where the potential parameter values obtain uniform weight (i.e. $\frac{1}{P}$), other weightings are possible using more complex prior distributions. During the study, the calculation of Bayesian $D$-optimal designs using a uniform prior on a given set of potential parameter values is called Bayesian design technique.

In cytotoxicity-testing, it is of interest at which concentrations a prespecified fraction of cells is damaged. Under the assumption that the concentration-response relationship is described by a 4pLL-model, these concentrations can be identified using the model fit of the data. More precisely, standardizing the upper asymptote $d$ to 100%, as it is often done in cytotoxicity experiments (see Kappenberg et al. 2020), the concentration at which the fitted regression function $\eta(x, \hat{\theta})$ attains 50% defines the value, at which exactly 50% of all cells were damaged. This value is denoted by the (absolute) $EC_{50}$-value. Note that the definition can easily be extended. For other percentages $p\% \in [0, 100]\%$, the (absolute) $EC_p$-value is defined by the concentration at which the fitted regression function $\eta(x, \hat{\theta})$ equals $p\%$, i.e. at which exactly $p\%$ of the cells were damaged.

## Results

### Design of an experiment with pre-existing knowledge about VPA

#### Construction of the concentrations for VPA data

Before a cytotoxicity experiment is conducted, the set of concentrations has to be fixed. For the determination of the sets of concentrations used to generate the dense dataset of the test substance valproic acid (VPA) (see Supplement Data) three different design techniques were applied.

The here applied design techniques require pre-existing knowledge about the test substance VPA, which was provided in form of data from three former experiments. More precisely, the highest possible soluble concentration was identified by 46.4 mM, resulting in the concentration range $\mathcal{X} = [0, 46.4]$ mM. Moreover, potential regions of $EC_{10}$- and $EC_{50}$-values were given by [0.373, 0.756] mM and [6.0, 7.0] mM, respectively. Finally, it was assumed that the effect at the solvent control is equal to 100, whereas the response/effect converges to zero if the concentration tends to infinity.

The three different design techniques were initially used to determine designs, which consist of four different concentrations. Applying the log-equidistant design technique, the design was constructed by a dilution factor of ten starting with the highest solubility 46.4 mM$\approx 10^{5/3}$ mM. This procedure resulted in the set of concentrations given by $\{0, 0.464, 4.64, 46.4\}$ mM, which is called log-equidistant design in the following. Combining the log-equidistant technique with biological arguments led to the set of concentrations given by $\{0, 1, 10, 46.4\}$ mM, which is called IfADo design in the following. Note that the IfADo design also includes a second concentration that is greater than 5 mM, which was chosen based on previous knowledge on the relatively high $EC_{50}$. Finally, the Bayesian design technique (cf. "Statistical methods") was used. Hereby, it was assumed that the unknown concentration-response relationship of VPA could be described by a 4pLL-model. Additional necessary prior information about the parameters of the 4pLL-model was obtained by pre-existing knowledge: the asymptotes of the true concentration-response relationship were set to $c = 0$ and $d = 100$ and three potential values each were provided for the $EC_{10}$- and $EC_{50}$-levels, respectively. The corresponding potential parameter values $b$ and $e$ were calculated for the $3 \times 3$ combinations of the $EC_{10}$- and $EC_{50}$-values resulting in a prior distribution of 9 potential model parameters. The corresponding curves are presented in Fig. 1. The resulting Bayesian $D$-optimal design, denoted by Bayesian design in the following, is given by the set of concentrations $\{0, 1.229, 9.334, 46.4\}$ mM.

To compare designs with different numbers of concentrations, the initial designs (log-equidistant, IfADo and Bayesian) were extended progressively to sets of five, six, and seven concentrations (see Table 1). Hereby, the log-equidistant design was first extended by a dilution factor of ten, which resulted in adding the concentration 0.0464 mM. As a progressive dilution of ten for more than five concentrations was inappropriate, the log-equidistant design was extended using a dilution factor of $10^{1/3}$ for 4.64 mM and then for 46.4 mM instead, resulting in the concentrations $(10)^{1/3} = 2.154$ mM and then $(10)^{4/3} = 21.54$ mM. The IfADo design was extended by a dilution factor of ten to designs with five and six concentrations, respectively. Finally, the concentration $(10)^{4/3} = 21.54$ mM was also added to the IfADo design.

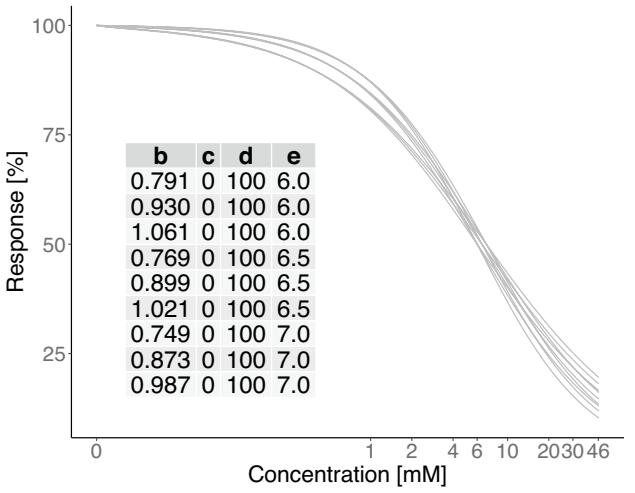



**Fig. 1** Model curves of prior distribution based on $EC_{10}$- and $EC_{50}$-values of the pre-experiments of VPA with information of the corresponding parameters of the 4pLL-models. Using different model curves as prior information, we can ensure robustness

The Bayesian design was extended to reduce the absolute distances between the different concentrations. Therefore, the gaps between the present concentrations of the Bayesian design were filled up with additional concentrations from largest to lowest determined by the available concentration closest to the midpoints of the present concentrations, in particular with 21.54 mM, 5.432 mM, and finally with 0.486 mM.

The log-equidistant, the IfADo, and the Bayesian designs were incorporated in the new experiments planned with sets of 50 concentrations. Those 50 concentrations further comprise the concentrations of locally $D$-optimal designs corresponding to the parameter combinations displayed in Fig. 1 (see Chernoff 1953 for details) and concentrations incorporated due to biological arguments. The data set, especially the values of the 50 different concentrations can be found in the Supplement (see Supplement Data).

### Definition of a reference curve for VPA data

In routine toxicity testing, concentrations are often defined by a log-equidistant design, which means that the concentrations are multiplied by the same dilution factor. In a first step, we studied whether the Bayesian design or the log-equidistant design leads to better model precision.

A precondition for the comparison of the precision of the different design techniques is a reference curve. The reference curve was obtained from cytotoxicity analyses of the test compound valproic acid (VPA) in six independent experiments with 50 concentrations each and all data points were fitted together to a 4pLL-model (Fig. 2). The purpose of using many concentrations and experiments was to

**Table 1** Designs under consideration with concentrations (in mM) included in the VPA data set

| Design | ♯Conc | Concentrations | | | | | | |
|---|---|---|---|---|---|---|---|---|
| IfADo | 4 | 0 | | | 1 | 10 | | 46.4 |
| IfADo | 5 | 0 | | 0.1 | 1 | 10 | | 46.4 |
| IfADo | 6 | 0 | 0.01 | 0.1 | 1 | 10 | | 46.4 |
| IfADo | 7 | 0 | 0.01 | 0.1 | 1 | 10 | 21.54 | 46.4 |
| Log-equidistant | 4 | 0 | | 0.464 | | 4.64 | | 46.4 |
| Log-equidistant | 5 | 0 | 0.0464 | 0.464 | | 4.64 | | 46.4 |
| Log-equidistant | 6 | 0 | 0.0464 | 0.464 | 2.154 | 4.64 | | 46.4 |
| Log-equidistant | 7 | 0 | 0.0464 | 0.464 | 2.154 | 4.64 | 21.54 | 46.4 |
| Bayesian | 4 | 0 | | 1.229 | | 9.334 | | 46.4 |
| Bayesian | 5 | 0 | | 1.229 | | 9.334 | 21.54 | 46.4 |
| Bayesian | 6 | 0 | | 1.229 | 5.432 | 9.334 | 21.54 | 46.4 |
| Bayesian | 7 | 0 | 0.486 | 1.229 | 5.432 | 9.334 | 21.54 | 46.4 |

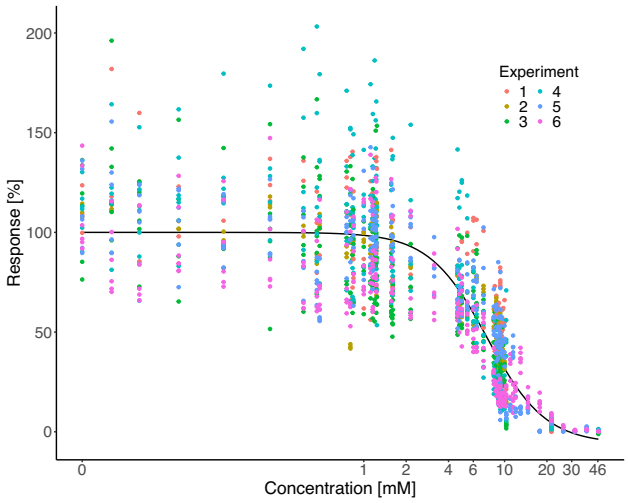



**Fig. 2** Cytotoxicity values for the six independent experiments of the VPA data set, 4pLL reference curve presented in black

approximate the true concentration–cytotoxicity relationship that theoretically would be approached when the number of experiments and concentrations grows to infinity.

### Bayesian design leads to a higher model precision compared to a log-equidistant design

In the following, the different presented design techniques to plan a cytotoxicity experiment are compared. First, the Bayesian design technique is compared to the frequently used log-equidistant design. To study the performance of both methods in the context of the different number of used concentrations, we consider designs with 4, 5, 6, or 7 concentrations based on the two design techniques, whereby the solvent control (where the VPA concentration was zero) is also counted as concentration (see Table 1). Note that for both methods, the highest soluble concentration (46.4 mM) was used as a starting point for dilutions.

To simulate specific scenarios that could represent individual experiments, three or six of the 30 observations (cytotoxicity) were randomly chosen for each selected concentration. With these specific scenarios, real concentration-response experiments were mimicked using actual data points, allowing for realistic variances and errors at different concentration levels to be accurately reflected. For each scenario, a 4pLL-model was fitted to the corresponding concentration-cytotoxicity data, further named 'scenario specific estimated curve' (SSEC). An example of the construction of a SSEC is shown in Fig. 3A. This simulation procedure was performed 3000 times for each design. To assess model precision in a practical manner, each SSEC was compared to the reference curve by calculating the Root Mean Squared Error (RMSE). The RMSE, represented by the area between the two curves (Fig. 3B), was used to make the differences easier to interpret. Additionally, we analyzed the maximal distance of the SSECs from the reference curve, as the *D*-optimality criterion is connected to the *G*-optimality criterion that seeks to minimize the maximal asymptotic variance (see "Modelling cytotoxic concentration-response data"). The smaller the RMSE/maximal distance, the higher the similarity of an SSEC to the reference curve. The resulting 3000 RMSEs for each design are summarized by box plots (Fig. 4), whereas the corresponding box plots of the maximal distances are depicted in Figure S1 in the Supplement Figures.

The results demonstrate consistently lower RMSEs (better) for the Bayesian design compared to the log-equidistant design strategy (Fig. 4). This is obtained for 4, 5, 6 as well as 7 concentrations and is consistent if three or six replicates were chosen. However, the difference in RMSEs between the Bayesian and the log-equidistant design strategy decreases if 7 concentrations are chosen, compared to the scenarios with 4–6 concentrations. This difference is explained by the fact that the critical concentration of

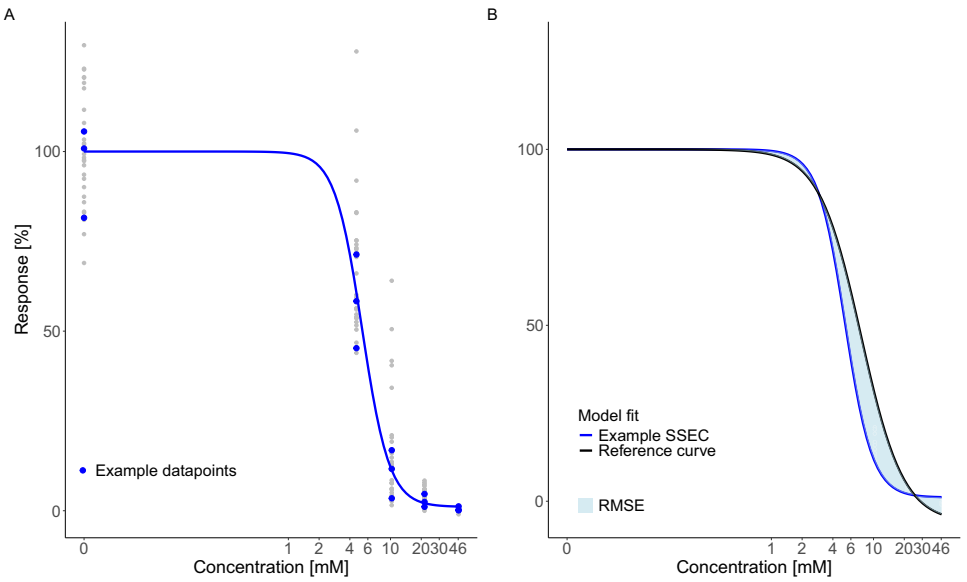



**Fig. 3** **A** Example illustrating the construction of a SSEC using the Bayesian design with five concentrations. The gray points depict all possible cytotoxicity values within the VPA dataset, while the blue data points represent three randomly selected cytotoxicity values, each representing a technical replicate. **B** Visualization of the RMSE of an exemplary SSEC compared to the reference curve. The greater the area between the two curves is, the higher is the RMSE, indicating a greater deviation from the reference curve

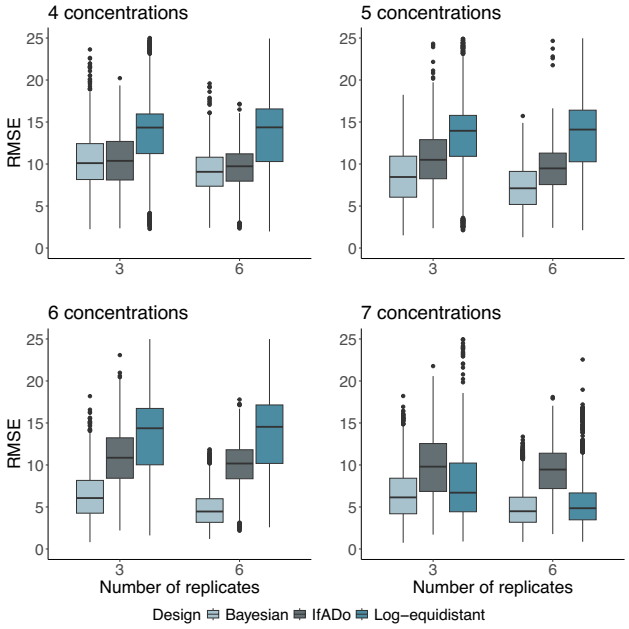



**Fig. 4** RMSE-values grouped by design, number of concentrations and used replicates for SSECs. Small RMSE values indicate higher precision of the design

21.54 mM becomes involved in both designs only when 7 concentrations are applied. Since the Bayesian design already includes higher concentrations when 4, 5, and 6 concentrations are chosen (explained in more detail in the discussion, see "Discussion and conclusion"), the improvement of model precision was higher for the log-equidistant design. Considering the maximum distance (see Figure S1 in Supplement Figures) similar results can be observed.

In laboratory routine, concentrations for cytotoxicity testing are often chosen based on previous experience or published data of a test substance, whereby smaller distances of concentrations are chosen in the concentration range of the expected $EC_{50}$. Usually, the choice of these concentrations is not based on a mathematical procedure but on previous knowledge, for example, former experiments showing in which concentration range a substance begins to be cytotoxic. Based on the experience of former experiments (see Albrecht et al. 2019), the concentrations of the so-called "IfADo design" were chosen in laboratory routine and also listed in Table 1.

The IfADo design technique is outperformed by the Bayesian design technique in terms of RMSE. While for 4 concentrations the mean RMSE of the IfADo design is only slightly higher (worse) than the RMSE of the Bayesian design, the latter is superior if 5, 6, or 7 concentrations are used (Fig. 4).

Moreover, the influence of the different design techniques on the precision of the $EC_{50}$-value (defined as the concentration that reduces vitality to 50% of the solvent control) was analyzed using the scenario-specific simulation. More precisely, we calculated the $EC_{50}$-values of the 3000 SSECs for each design and compared them with 7.147 mM, which is the $EC_{50}$-value of the reference curve. The results are presented in Fig. 5, where the red lines represent the $EC_{50}$-value of the reference curve. For the interpretation of the results, it is important to consider both the median and the interquantile ranges (IQR) of the $EC_{50}$-values.

In general, the Bayesian design approach leads to the most precise $EC_{50}$-estimations compared to the other design approaches although the Bayesian design is not constructed for that purpose. Considering the cases in which the number

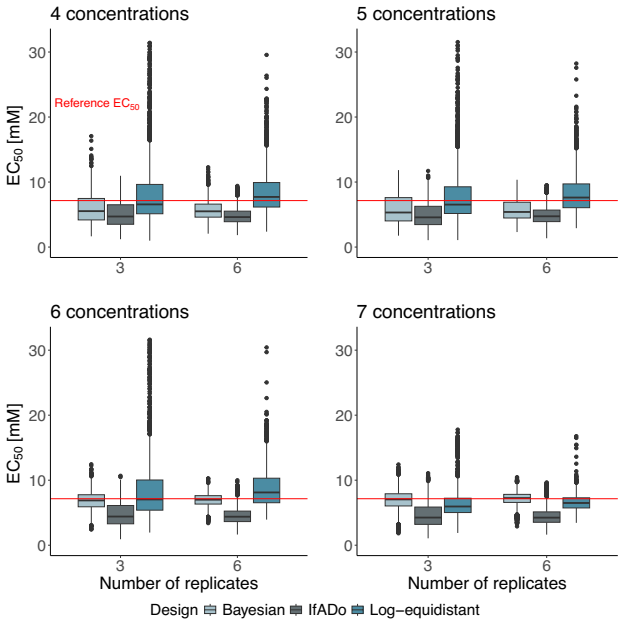



**Fig. 5** $EC_{50}$-values grouped by design, number of concentrations and used replicates for SSECs. The red line corresponds to the reference $EC_{50}$. The closer the $EC_{50}$-values of the SSECs are to the reference value, the more precisely the $EC_{50}$s are estimated

of different concentrations is 6 or 7, the Bayesian $EC_{50}$ median is closer to the reference than the one of the log-equidistant design. For 4 and 5 concentrations, the median $EC_{50}$ obtained by the Bayesian design shows a higher deviation from the reference than the $EC_{50}$ median obtained using the concentrations of the log-equidistant design (Fig. 5). Besides, the IQRs of the $EC_{50}$-values based on the Bayesian design technique are smaller than the ones based on the log-equidistant design technique in the situation of fewer than 7 concentrations. Comparing the designs consisting of 7 concentrations, the IQRs are of a similar size.

The precision of the $EC_{50}$ based on the IfADo design is worse than the one of the Bayesian and the log-equidistant design techniques. In particular, the IfADo design leads to a substantial deviation of the median $EC_{50}$ to the reference independent of the number of different concentrations involved.

The number of outliers produced, if the different design techniques are used, is also an indicator of the quality of the experiment. Considering both the RMSE-values (see Fig. 4) and the $EC_{50}$-values (see Fig. 5), the outliers for the Bayesian design technique are lower than for the log-equidistant design technique, independent of the number of concentrations used. Therefore, the Bayesian design leads to RMSEs and $EC_{50}$-values with substantially lower variability.

In toxicological experiments sometimes a sequential design of an experiment consisting of a pre-experiment and a main experiment is necessary. Therefore, a sequential design approach, involving a preliminary experiment followed by a main experiment, was evaluated to determine whether it could enhance the precision of concentration-dependent toxicity testing. The preliminary experiment is used to obtain knowledge about the $EC_{10}$-value and the $EC_{50}$-value of the test substance. This information is then utilized to refine the design of the main experiment, with the goal of obtaining more detailed insights into the cytotoxicity of the substance. It was investigated if the sequential procedure improves the performance of the different considered design techniques. The performance of the log-equidistant design and the IfADo design is improved by the sequential approach, whereas the performance of the Bayesian design is not affected. However, the sequential IfADo approach remains worse than the non-sequential Bayesian approach. The results indicate that a sequential experiment is not required when concentrations are initially selected using the Bayesian technique. For a comprehensive description of the analysis setup and detailed results, refer to the Supplement Sequential analysis.

### Bayesian design requires the smallest sample size

An important aspect of planning a cytotoxicity experiment is the choice of a sample size that ensures a pre-specified statistical quality of the experiment. An appropriate quality indicator is the length of the confidence interval (LCI) of the estimate of the $EC_{50}$-value. The LCI decreases if the sample size increases, while its value also depends on the variance of the data and the used design, i.e. the set of considered concentrations. Therefore, using the variance of the VPA data set (see Fig. 2), the dependence between the required sample size and the resulting LCI was investigated for all designs displayed in Table 1.

The Bayesian design technique requires a substantially smaller sample size than the other design techniques to achieve the same precision in terms of LCI (see Fig. 6 and Suppl. Figure S2). In Fig. 6, the whole relation between the sample size and the LCIs is visualized for the three different designs with seven concentrations, while the corresponding visualizations for the designs with smaller numbers of different concentrations can be found in Supplement Figure S2. The performance of the different designs can be compared by fixing a value for the LCI. For instance, if an LCI of 5 should be achieved and the Bayesian design with seven concentrations is used, a sample size of 25 is sufficient, while a five times larger sample size of 123 is necessary to achieve an LCI of 5 using the seven concentrations of the log-equidistant design. The corresponding required sample size of the IfADo design with seven concentrations is 63.

The LCI depends on the reference curve and the different concentrations in each considered design. This implies that increasing the number of concentrations does not necessarily

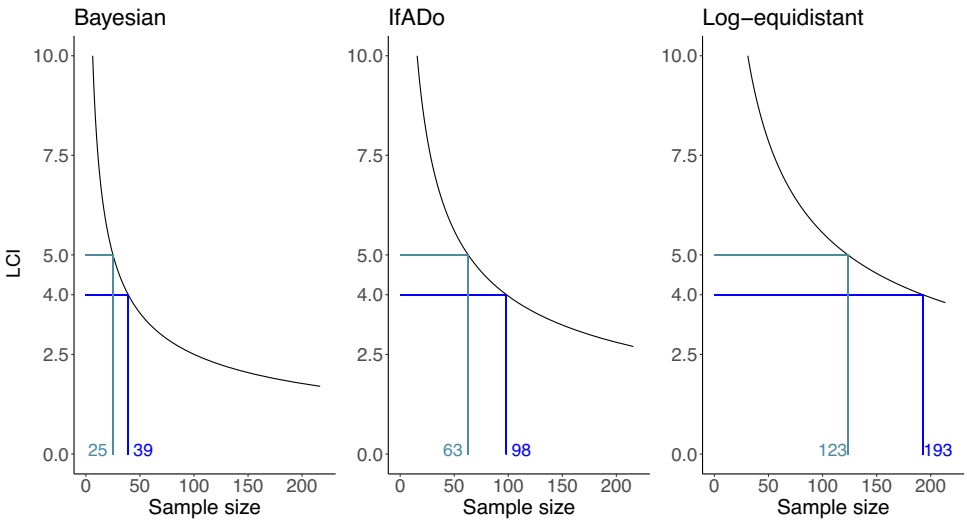



**Fig. 6** The relation between the sample size and the length of the confidence interval (LCI) of the $EC_{50}$-value is shown grouped by design. The blue lines indicate the number of observations required to achieve a specific level of precision for the $EC_{50}$ (LCI). To achieve the same precision of the $EC_{50}$ the Bayesian design requires considerably less observations than the other two design techniques

decrease the LCI when the concentrations have already been optimally selected for a smaller set of concentrations (see Suppl. Figure S2).

Summarizing, the Bayesian design technique leads to the same precision with less experimental effort compared to the two other design techniques.

## Design of an experiment without prior knowledge of a test compound

Often, an cytotoxicity experiment has to be planned for a test substance for which no direct prior information is available. In this situation, less precise information about the behavior of related test compounds can be used, e.g., based on historical data. The case in which no information about the new substance can be gained rarely occurs. Furthermore, the design of an experiment without any prior information involves different statistical methods, which are not included in this study for the sake of brevity. Thus, this study focuses on the optimal design of a cytotoxicity experiment of a new (unknown) test substance where information about related test compounds is available.

### Construction of a universal Bayesian design based on DILI data

Often, the cytotoxicity of a test substance has to be investigated without direct prior knowledge that can be used to design the experiment. In this situation, a sequential procedure that consists of a pre- and a main experiment is inevitable. Then, an important question is how the pre-experiment should be designed without direct prior information.

We investigated if the concentrations of the pre-experiment should be defined by a log-equidistant design, which is frequently used in routine cytotoxicity testing, or if the concentrations should be obtained by the Bayesian design technique using less informative prior knowledge. One option to generate this prior information is using available information about compounds that are assumed to be similar to the test compound under consideration.

For the substance under investigation, VPA, data of 103 other substances within a DILI data set (see Supplement Data) were used as (less informative) prior knowledge to construct a Bayesian design with seven concentrations. For each tested substance, determining the highest considered concentration depends on various limiting factors. The highest concentration is influenced by the highest solubility, as well as the pH value, and the requirement to include at least one concentration with no observable effect. As the concentration-response data of the different substances within the DILI data set vary substantially in the highest considered concentrations, the substance-specific data had to be normalized by dividing the set of originally used concentrations by the respective corresponding highest concentration. This normalization resulted in the same concentration range of all 103 substances, namely $\mathcal{X}_0 = [0, 1]$. The resulting data set is called the normalized DILI data set in the following. For each of the 103 substances within the normalized DILI data set the corresponding concentration-response data were fitted to a 4pLL-model. For 77 compounds the resulting model fits showed a goodness-of-fit value above 0.55 (see Albrecht et al. 2019), thus a convenient 4pLL-model fit was assumed here. The data of the remaining 26 substances were excluded from the further analysis, as they were inappropriate for the construction of suitable prior knowledge. Considering the distribution of the (normalized) $EC_{10}$ and $EC_{50}$, a major part of substances resulted in small values, respectively (Fig. 7). Only one substance, Troglitazone, had a normalized $EC_{50}$ -value greater than 1 corresponding to an original $EC_{50}$ value greater than its highest considered concentration used in the experiment. Therefore, Troglitazone was excluded from further analysis. Summarizing, parameter estimations of the

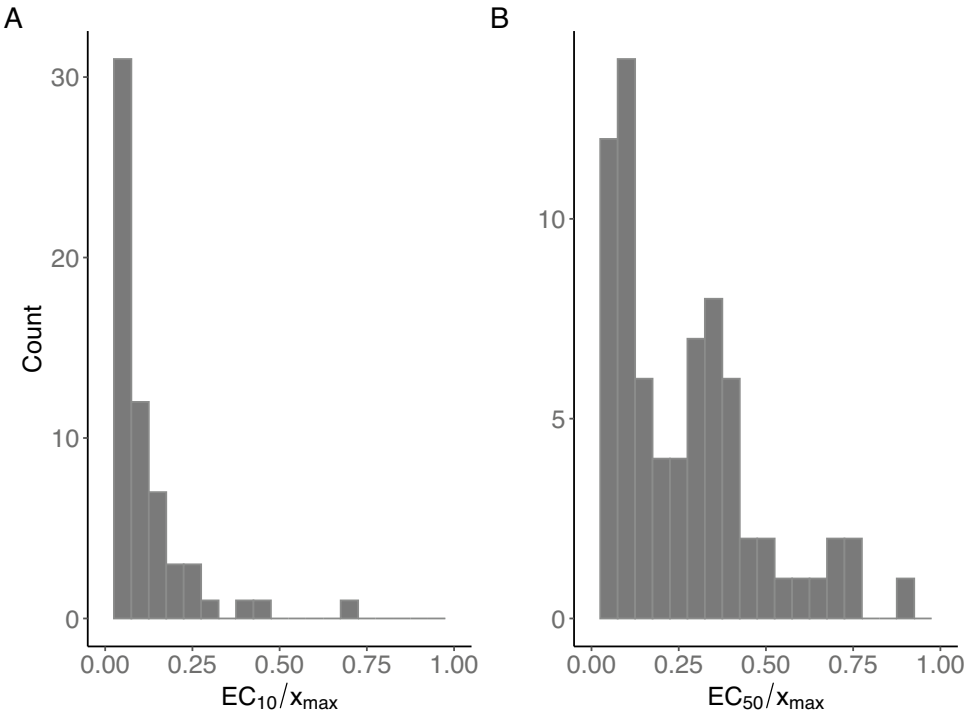



**Fig. 7** **A** Histogram of all $EC_{10}$-values of the considered substances in relation to the maximal concentration $x_{max}$, **B** Histogram of all $EC_{50}$-values of the considered substances in relation to the corresponding highest considered concentration $x_{max}$

4pLL-models fitted to 76 substances of the normalized DILI set served as prior distribution for the Bayesian design technique, whereby the corresponding 76 parameter combinations were equally weighted. The concentration range for the Bayesian design technique was also standardized to the concentration range $\mathcal{X}_0 = [0, 1]$.

The number of concentrations was fixed to seven to develop a design that shows a high model precision and can easily be managed in a laboratory routine experiment. The resulting Bayesian design with prior information of the DILI data set (short: BayesianDILI) consisting of seven concentrations is then given by the set: $\{0, 0.0300, 0.0623, 0.1351, 0.3206, 0.5177, 1\}$. Note that the maximal concentration of this design is 1 and therefore neither directly applicable for the different substances within the original DILI set nor for VPA. To obtain an appropriate design for the substance under consideration the design has to be multiplied by the specified maximal concentration. For the example of VPA with a highest considered concentration of 46.4 mM, the BayesianDILI is then transformed to the design consisting of the concentrations 0, 1.392, 2.891, 6.2 69, 14.876, 24.021, 46.4 mM.

### BayesianDILI design outperforms IfADo design for the example of VPA

If no prior knowledge about a test substance is available, an important question is how the corresponding experiment should be planned. The lack of prior information requires a sequential procedure consisting of a pre-experiment and a main experiment. The purpose of the pre-experiment is to gain some knowledge about the test substance, which then can be used for the design of the main experiment. The design of the pre-experiment itself is often chosen based on previous experience in laboratory routines. Besides a log-equidistant design technique is often applied, where the concentrations are chosen by a specified dilution factor. However, the application of a Bayesian design technique might be more suitable for the design of the pre-experiment.

In the following, it is investigated if a Bayesian technique based on indirect prior knowledge is superior to the currently established design strategies for planning a pre-experiment. More precisely, in the situation where the new substance is given by VPA, the BayesianDILI design given by the set $\{0, 1.392, 2.891, 6.269, 14.876, 24.021, 46.4\}$ mM (see previous Section) is compared to the log-equidistant, the IfADo and the Bayesian design with seven concentrations (see Table 1). For the comparison, simulations are performed based on the experimentally performed VPA dataset.

As this dataset does not include exactly the concentrations of the BayesianDILI design, the seven concentrations closest to the experimentally determined ("true") concentrations were selected. The resulting set of seven concentrations is given by $\{0, 1.229, 2.154, 6.331, 10.287, 21.54, 46.4\}$ mM, which is called BayesianDILI_VPA design in the further analysis. Similar to previous analyses, the SSECs for the BayesianDILI_VPA design were calculated and compared to the reference curve in terms of RMSE and $EC_{50}$-values in each simulation step. The resulting values were then compared to the RMSEs and the $EC_{50}$-values of the SSECs of

the log-equidistant, IfADo, and Bayesian design for VPA with seven concentrations (see Table 1). To prevent any ambiguity, the Bayesian design for VPA will be referred to as BayesianVPA in the following analysis. Note that the BayesianVPA design by construction has an advantage over the BayesianDILI_VPA design, because it is based on explicit prior knowledge about the substance VPA (see "Construction of the concentrations for VPA data"), whereas no knowledge about VPA itself was used to construct the BayesianDILI_VPA design (see "Construction of a universal Bayesian design based on DILI data").

The results demonstrate a good performance of the BayesianDILI_VPA design compared to the three other design strategies, which incorporated direct prior knowledge about the new substance (Fig. 8A and B). This is obtained both for the RMSEs and the $EC_{50}$-values. In particular, the BayesianDILI_VPA design leads to RMSEs which lie between the RMSEs of the IfADo design and the ones of the BayesianVPA/log-equidistant design. Furthermore, the variability of the RMSE-values of the BayesianDILI_VPA design is lower than the variability of the other designs, indicated by smaller interquartile ranges. Concerning the $EC_{50}$, the BayesianDILI_VPA design results in a median estimate which lies below the $EC_{50}$ of the reference curve, but which is still better than the ones based on the IfADo design and the log-equidistant design, respectively (see Fig. 8B). Finally, the variability of the estimated $EC_{50}$ based on the BayesianDILI_VPA design is low compared to the log-equidistant design and only minor outliers are present.

### BayesianDILI designs outperforms original and log-equidistant design for 76 substances of the DILI data set

In the following, the dependence of different design techniques on the quality of cytotoxicity experiments is investigated for substances of the DILI assay (except for VPA). More precisely, we considered the 76 substances of the DILI assay that served as prior information for the construction of the BayesianDILI design. Hereby, the BayesianDILI design was compared to the design originally used in the DILI assay for each substance (see Supplement Data) and to the frequently used log-equidistant design. For that purpose, the BayesianDILI design was rescaled based on the

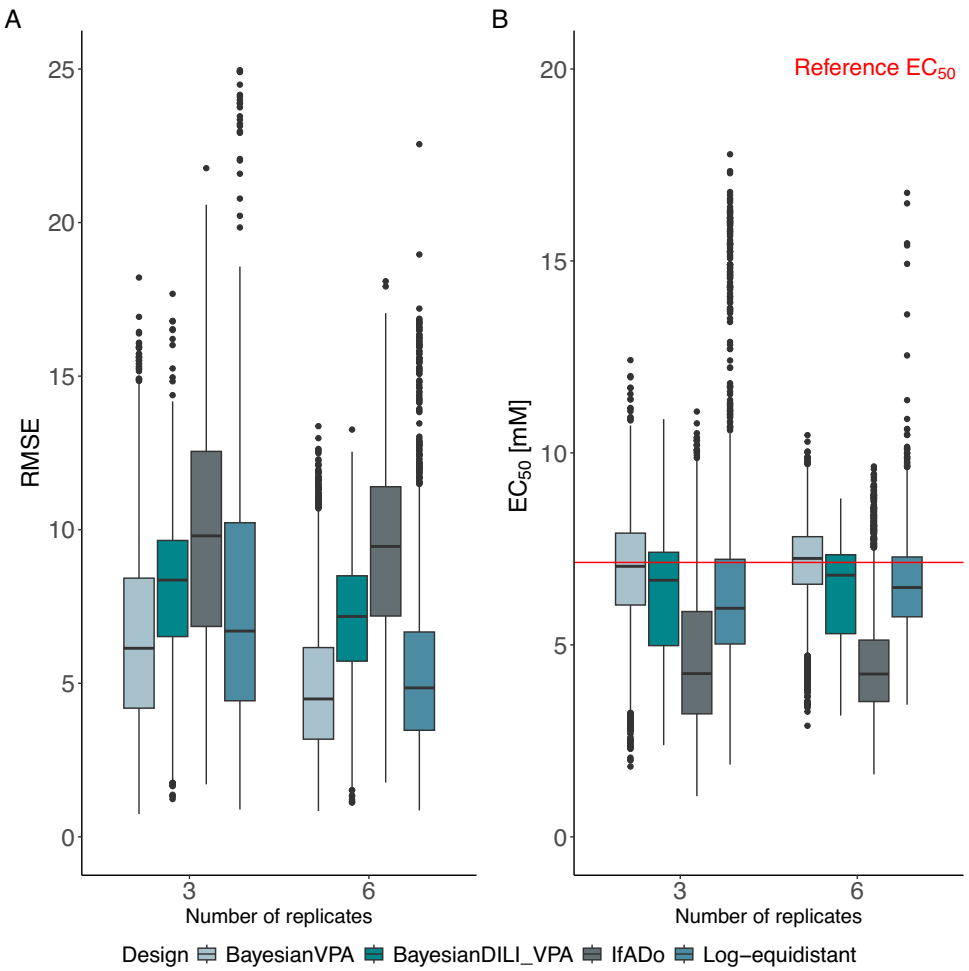



**Fig. 8** **A** RMSE-values grouped by design, number of concentrations and used replicates for SSECs. Small RMSE values indicate higher precision of the design. The BayesianDILI_VPA design performs better than the IfADo design although it does not incorporate prior knowledge about the test compound VPA itself. **B** $EC_{50}$-values grouped by design, number of concentrations and used replicates for SSECs. The red line corresponds to the reference $EC_{50}$. The closer the $EC_{50}$-values of the SSECs are to the reference value, the more precisely the $EC_{50}$-values are estimated. The BayesianDILI_VPA design outperforms the log-equidistant and IfADo design regarding the mean precision of the $EC_{50}$ -values, although it does not incorporate prior knowledge about the test compound VPA itself

substance-specific highest considered concentration (see " Construction of a universal Bayesian design based on DILI data" for details). Whereas the log-equidistant design with seven concentrations was constructed by using a dilution factor of 10 starting from the substance-specific highest considered concentration and is called log10 design in the following analysis. Since designs with different numbers of concentrations are not directly comparable, the sample size of all designs was fixed to the sample size of the original design using a rounding procedure (Pukelsheim and Rieder 1992).

For each substance, the performance of the different designs, i.e. the BayesianDILI, log10 and original design was analyzed using a simulation study. In contrast to the previous data-driven simulation, this analysis is based on computer-simulated data due to the limited availability of large datasets for the DILI substances. During the computer-based simulation, we assume normally distributed homoscedastic errors for all observations. This approach allows us to theoretically generate observations based solely on the model assumptions, eliminating the need for extensive experimental data. While this ensures greater comparability between simulations, it comes at the cost of losing the practical advantages of using real-world data, such as robustness and real-life applicability. However, the benefit of theoretical simulations lies in their flexibility and control over experimental conditions. This allows us to analyze the considered designs under idealized conditions and provides insights that might be challenging to obtain from empirical data due to variability issues. More precisely, this computer-based simulation is based on 1000 repetitions, which were conducted for each scenario, i.e. for each of the 76 substances using either the corresponding BayesianDILI, the original, or the log10 design. One simulation step consisted of the following intermediate steps: First, a 4pLL model was fitted to the original data of the considered substance and fixed as a reference curve for further analysis. Moreover, the corresponding substance-specific pooled variance was calculated based on the original data. Then new data was simulated using the responses of the reference curve at the concentrations of the considered design together with additive normally distributed random errors, whose variance coincided with the substance-specific pooled variance. Finally, a 4pLL model (SSEC) was fitted to the simulated data and compared to the substance-specific reference curve using the RMSE (see Fig. 3B for an illustration of this procedure).

The resulting mean RMSE-values over the 1000 simulation steps of all 76 substances are shown for the BayesianDILI design on the *y*-axis versus the log10 design on the *x*-axis, where each dot represents one substance (Fig. 9). Values below the bisecting dashed line indicate a better performance of the BayesianDILI design compared to the log-equidistant one. Similarly, values above the dashed line show a better performance of the log10 design.

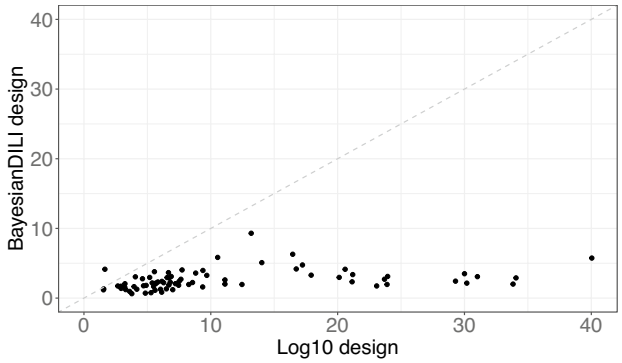



**Fig. 9** The Scatterplot shows RMSE-values corresponding to the simulation results of 76 substances regarding BayesianDILI and log10 design. Each dot represents one substance. The BayesianDILI design outperforms the log10 design for 75 of 76 substances

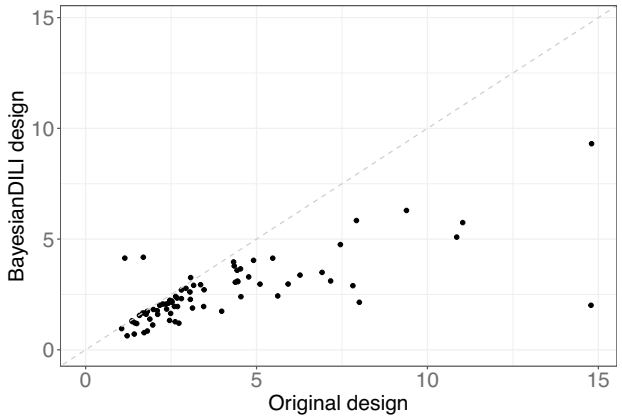



**Fig. 10** The Scatterplot shows RMSE-values corresponding to the simulation results of 76 substances regarding BayesianDILI and the original used design. Each dot represents one substance. The BayesianDILI design outperforms the original design for the majority of substances

The simulation results show significantly lower mean RMSEs for the BayesianDILI design compared to the log10 design for 75 of 76 substances, with the exception of one compound (Griseofulvin) (Fig. 9). While the mean RMSEs based on the BayesianDILI design vary between 0 and 10, the values of the mean RMSEs based on the log10 design are up to four times higher.

The mean RMSEs of the BayesianDILI design are also smaller than the RMSEs of the original design in the DILI assay (Fig. 10). More precisely, the original design outperforms the BayesianDILI design only for three of 76 substances. These three substances show a rather high toxicity due to their maximal concentration.

Based on the simulation results, the mean distance of the $EC_{50}$-values to the $EC_{50}$-values of its reference curve was calculated for each substance under consideration and design technique. For the sake of comparability between all 76 substances, the $EC_{50}$-values were normalized by the maximal concentration of each substance, respectively. Congenial to the analysis of the RMSE-values, the mean $EC_{50}$-values based on the BayesianDILI design are consistently smaller than the ones based on the original and the log10 design (Suppl. Figure S3 and S4).

## Guideline for the design of cytotoxicity experiments

In cytotoxicity testing, careful consideration must be given to the experimental design, as the concentrations used have a significant impact on the precision of the statistical analysis. In the following, a guideline for the design of cytotoxicity experiments is presented based on our findings (Fig. 11). The recommended procedure for an arbitrary cytotoxicity experiment is presented distinguishing between the scenarios with and without pre-existing direct knowledge about the test substance (Fig. 11). Furthermore, the proceeding based on the guideline is described for the example of the substance tolcapone (Fig. 12).

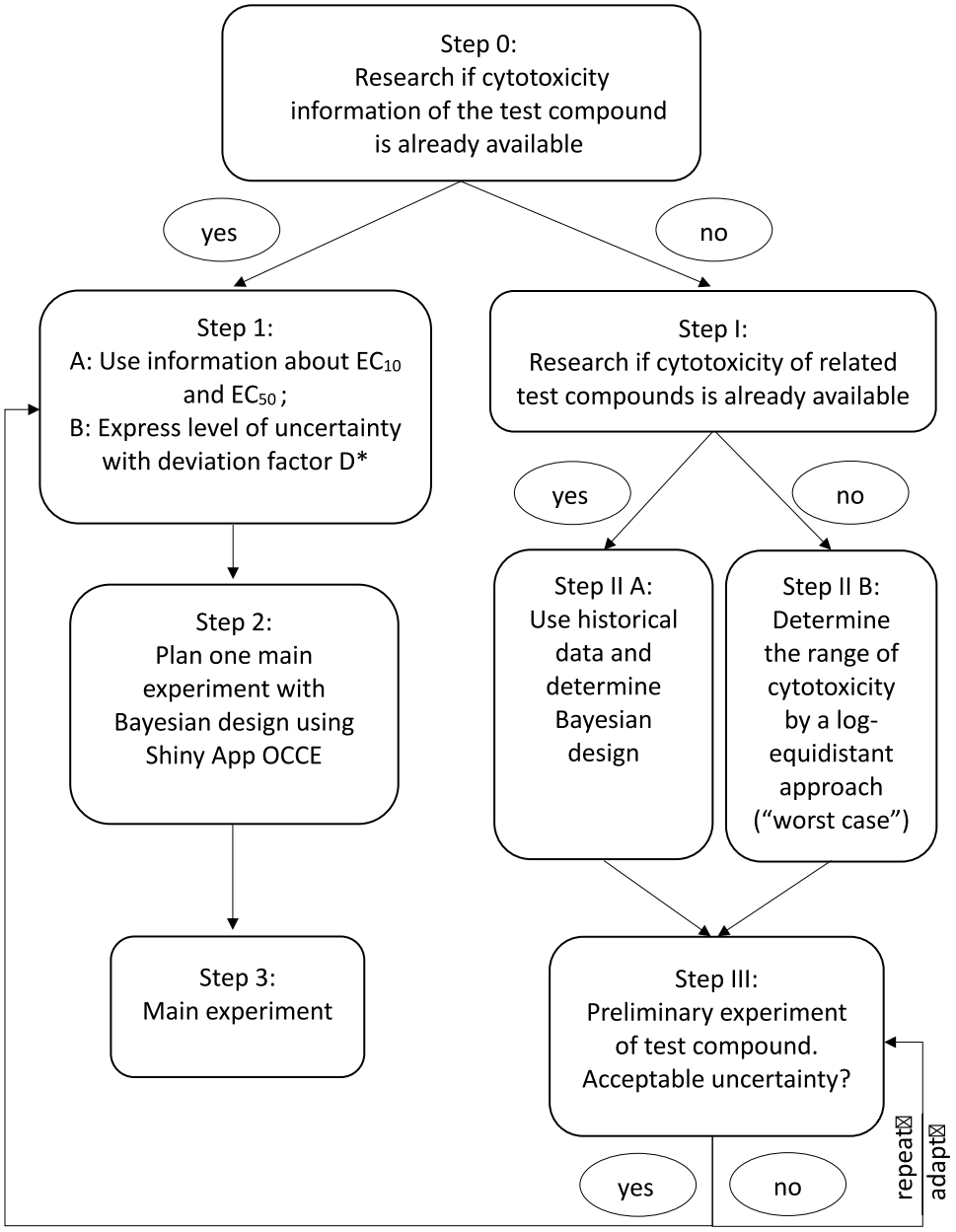



**Fig. 11** Recommendation for the planning of new cytotoxicity experiments based on the findings of this study. *To ensure a more robust design accounting for uncertainty a deviation factor *D* can be specified. Based on this an upward and downward deviation of the alert concentrations corresponding to exactly this factor can be assumed. If the supplied information of the substance is considered very safe no deviation is assumed, which would result in a deviation factor of 1. The less specific the information is, the higher the factor should be. The deviation factor *D* should not exceed $\frac{x_{max}}{EC_{50}}$. In the absence of prior knowledge of the test substance a statistician should be consulted for the calculation of Bayesian and adaptive designs

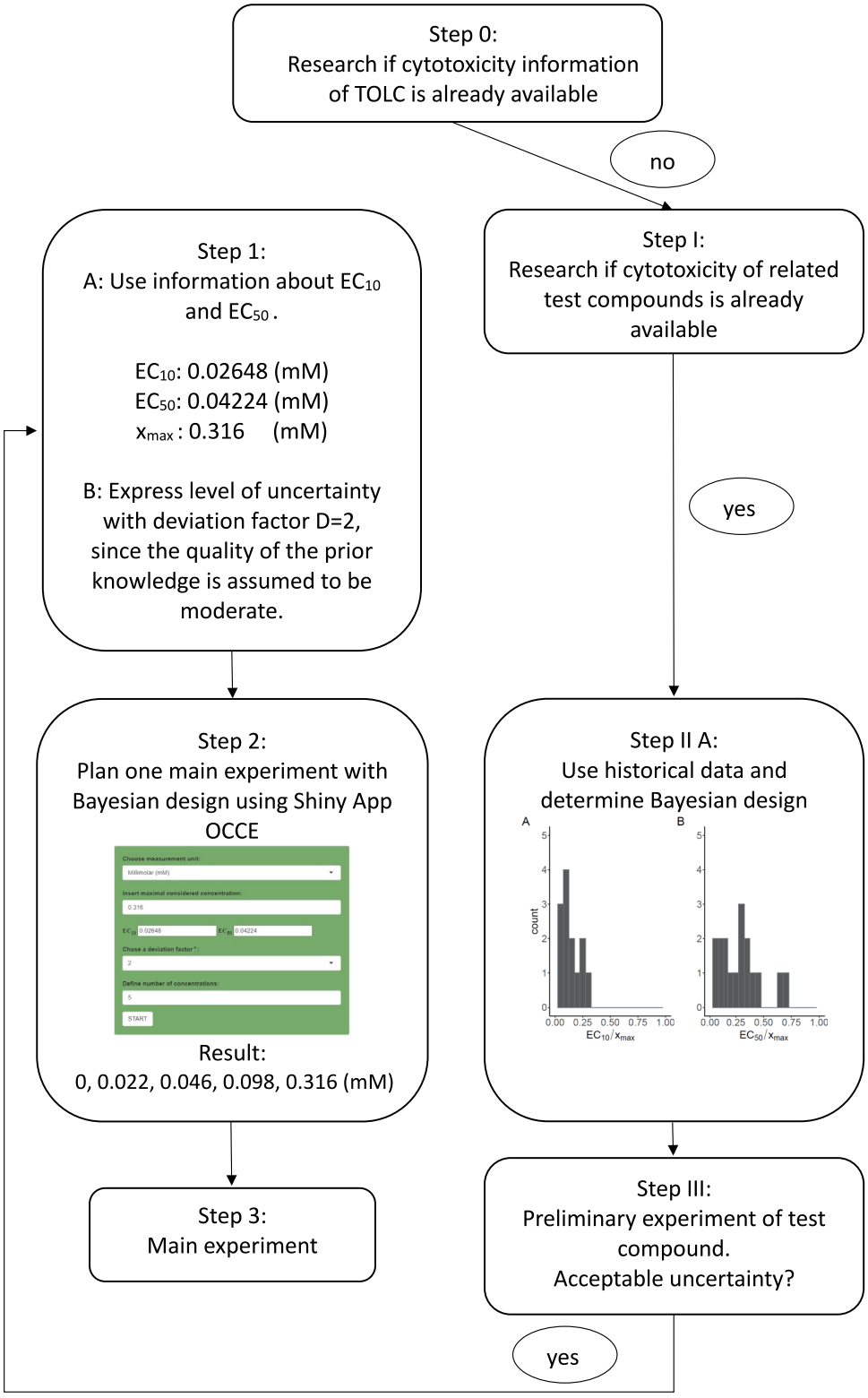



**Fig. 12** Recommended procedure for the example of tolcapone (TOLC) for the scenario without explicit prior knowledge of the test substance

In the following the individual steps of this guideline will be discussed.

*Step 0: Research if cytotoxicity information of the test compound is already available*. The aim of this initial step is to find out if the specific test compound has already been previously concentration-dependently tested for cytotoxicity. It is not necessary that the test conditions and the tested cell types are identical to the planned experiment.

However, the more similar the test conditions are, the better.

If cytotoxicity information about the test substance is known due to previous experiments of the test substance, one proceeds to *Step 1 A: Use information about* $EC_{10}$ *and* $EC_{50}$. In this step the alert concentrations can either be identified by reported values or calculated based on previous cytotoxicity data by, e.g. the 4pLL model (Ritz et al. 2015).

*Step 1 B: Express level of uncertainty with deviation factor D*. A deviation factor $D$ has to be fixed to reflect the possible range of $EC_{10}$- and $EC_{50}$-values. More specifically, all values in $[1/D \cdot EC_{10}, D \cdot EC_{10}]$ and $[1/D \cdot EC_{50}, D \cdot EC_{50}]$ are assumed to be possible values for the $EC_{10}$ and the $EC_{50}$, respectively. Hereby, the deviation factor $D$ can be chosen within $[1, \frac{x_{max}}{EC_{50}}]$, in which the highest deviation factor is limited by the maximally considered concentration ($x_{max}$) and the $EC_{50}$. In this case, it is assumed to be possible that the $EC_{50}$ value and the maximally considered concentration coincide. It should be considered that the deviation factor does not represent an exact natural parameter but to a certain degree includes the subjective assessment of the investigator.

If the information about the substance is precise, a deviation of $D = 1$ will be chosen. Setting $D = 1$ would represent an unusual situation, in which a) the substance has already been tested with the same cytotoxicity assay and cell line; and b) the $EC_{10}$- and $EC_{50}$-values do not vary between independent experiments (experiments performed on different days). As a consequence a higher deviation factor $D > 1$ will be justified in practice. If the previous experiments used the same cytotoxicity test and cell line, the investigator should quantify the deviation factor by the corresponding confidence intervals of independent experiments (for example the basis of the deviation factor will be 2, 3 or 5 if the confidence interval ranges 2, 3 or 5 fold for the individual experiments). This value will be multiplied by an additional factor, e.g., 2 if a different cell line or different experimental conditions (for example a deviating incubation time) were used. The additional factor increases the more the new assay design deviates from the original assay design used to obtain the EC -values. If the quality of previous data justifies a small deviation factor, the shape of the concentration-response curve will be determined more precisely.

In *Step 2: Plan one main experiment with Bayesian design using Shiny App OCCE* the provided Shiny app OCCE (see http://shiny.statistik.tu-dortmund.de:8080/app/occe) can be used. For this purpose $EC_{10}$, $EC_{50}$, the maximally considered concentration $x_{max}$, deviation factor $D$ (from Step 1) and the number of concentrations ($N$) that should be tested will be entered. The result will be a list of $N$ concentrations that should be used to conduct the main experiment (Step 3).

In the absence of previous knowledge about the test substance of interest, continue with *Step I: Research if cytotoxicity of related test compounds is already available*. Here related compounds can be substances which belong to a same chemical class and were tested by the similar cytotoxicity test.

*Step II A: Use historical data and determine Bayesian design*. The purpose of this step is to enable a higher density of concentrations around the range where $EC_{10}$- and $EC_{50}$ -values can be expected. For the calculation a statistician should be consulted.

*Step II B: Determine the range of cytotoxicity by a log-equidistant approach ("worst case")*. This "worst case" refers to a situation where almost no information concerning cytotoxicity is available. In this worst case we still know a very wide concentration range, e.g. $10^{-10}$ to 10 M. The cytotoxicity of almost all substances occurs in this range. This means that preliminary experiments have to be performed to narrow down the region of interest. For this purpose an equidistant approach for example with factors of 10 or 100 can represent a useful first step.

*Step III: Preliminary experiment of test compound. Acceptable uncertainty?* In principle, the preliminary experiment can lead to a scenario where a lower concentration will show no toxicity at all while the next higher concentration will show 100% toxicity. In this case the above described deviation factor ($D$) may support the decision whether to directly move to Step 1 or to perform additional preliminary experiments. The concept of the deviation factor can be used to quantify the level of uncertainty by using the range of the confidence intervals of the preliminary experiments. If the deviation factor leads to an exceedance of the maximally considered concentration an additional preliminary experiment should be considered. In general a statistician should be consulted for the calculation of the adaptive designs in this step.

In the following the procedure based on the guideline is described by the example of the substance tolcapone (TOLC) (Fig. 12). In this example it is assumed that no prior knowledge of the substance is available in Step 0. Accordingly, one proceeds with Step I. Since data of 20 related compounds to tolcapone, tested with the same cytotoxicity assay, are available in this example, the corresponding $EC_{10}$- and $EC_{50}$ -values are used to construct distributions for possible $EC_{10}$- and $EC_{50}$-values of the test substance TOLC. Based on this a Bayesian design is constructed for the preliminary experiment in Step III. The corresponding uncertainty is assumed to be acceptable since the preliminary experiment showed a confidence interval of the $EC_{50}$ of a nearly 2-fold range. Accordingly, the cytotoxicity information of the preliminary experiment ($EC_{10}$, $EC_{50}$, $D$, $x_{max}$) is used in Step 1. Entering these values in the OCCE app while aiming for five different concentrations results in the calculation of the following five

optimal concentrations: 0, 0.022, 0.046, 0.098, 0.316 (mM) (Step 2). The main experiment with at least three technical replicates can then be performed with the corresponding concentrations (Step 3).

### Shiny app OCCE (optimal concentrations for cytotoxicity experiments)

Often the calculation of optimal concentrations for an upcoming cytotoxicity experiment requires deeper knowledge in software programming. To make the calculations easily accessible to everyone, we have developed a Shiny app (see http://shiny.statistik.tu-dortmund.de:8080/app/occe). This Shiny app called **O**ptimal **C**oncentrations for **C**ytotoxicity **E**xperiments (short: OCCE) enables the calculation of optimal concentrations for user-specified situations. First the unit of measurements and the maximally considered concentration for the compound of interest needs to be supplied. Additionally a prior guess of the alert concentrations $EC_{10}$ and $EC_{50}$, as well as a corresponding deviation factor (for details see "Guideline for the design of cytotoxicity experiments") has to be entered (for an example see Fig. 13). Then the user can specify the desired number of concentrations $N$ for the calculation. Afterwards the resulting $N$ optimal concentrations based on the supplied knowledge are calculated and returned as a list. Furthermore, details of the parameters and the calculation methodology are given.

Apart from the calculator of optimal concentrations, the OCCE app includes a tool for comparing specific concentrations to the optimal concentrations. In particular, the efficiency of a user-specified design can be evaluated by this app. The higher the efficiency is, the better the chosen design is. This allows for the evaluation of the quality of other designs in comparison to the optimal concentrations. If, for instance, the exact calculated concentration of 2.821 mM is difficult to measure in practice, the user can assess the impact of adjusting the concentration to 3 mM (see Fig. 14). A design with an efficiency over 80% is classified as a good design, while an efficiency of 70 to 80% is sufficient and a design below 70% should not be used to plan a new experiment.

## Discussion and conclusion

This study introduces designs for cytotoxicity experiments for two scenarios: On the one hand, where pre-existing knowledge of the test substance is available and on the other hand, where this is not the case.

For the first scenario, the performance of three different design techniques (log-equidistant, IfADo, and Bayesian) was investigated by the example of valproic acid (VPA). Based on pre-existing knowledge of the $EC_{10}$ and $EC_{50}$, designs with 4–7 concentrations were constructed for each technique, respectively.

To ensure a comparable assessment of the different designs, a reference curve for VPA was determined based on experiments with an unusually high number of 50 different concentrations with 3–6 technical replicates. This reference curve was used to quantify the degree of deviation obtained for the different design techniques in terms of the RMSE and precision of the $EC_{50}$.

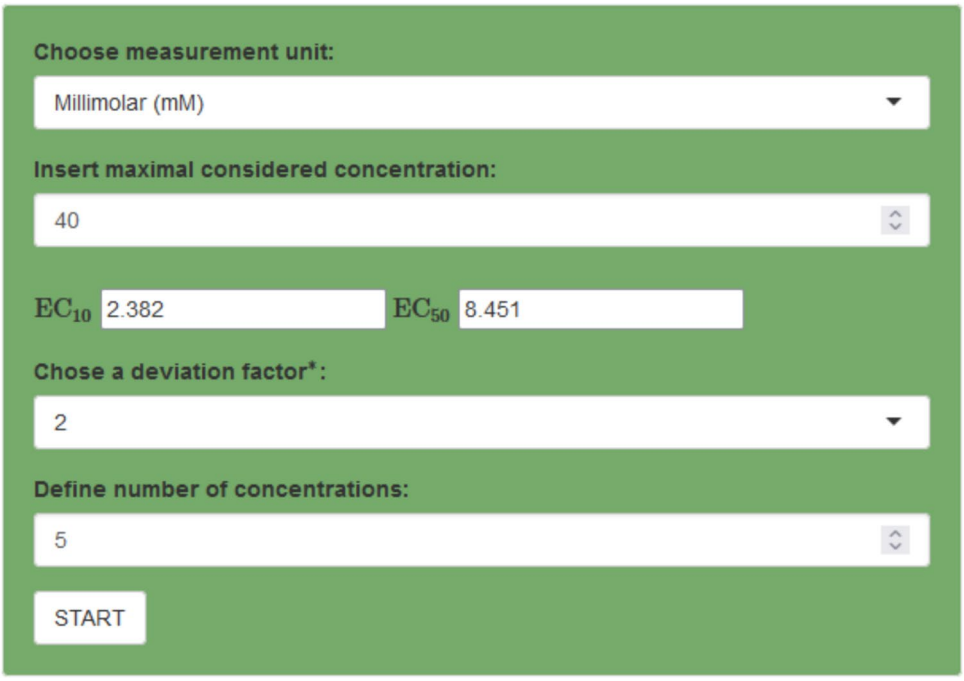



**Fig. 13** Example of supplied cytotoxicity information in Shiny app OCCE. The resulting optimal concentrations for this example are 0, 2.821, 6.099, 13.503, 40 (mM)

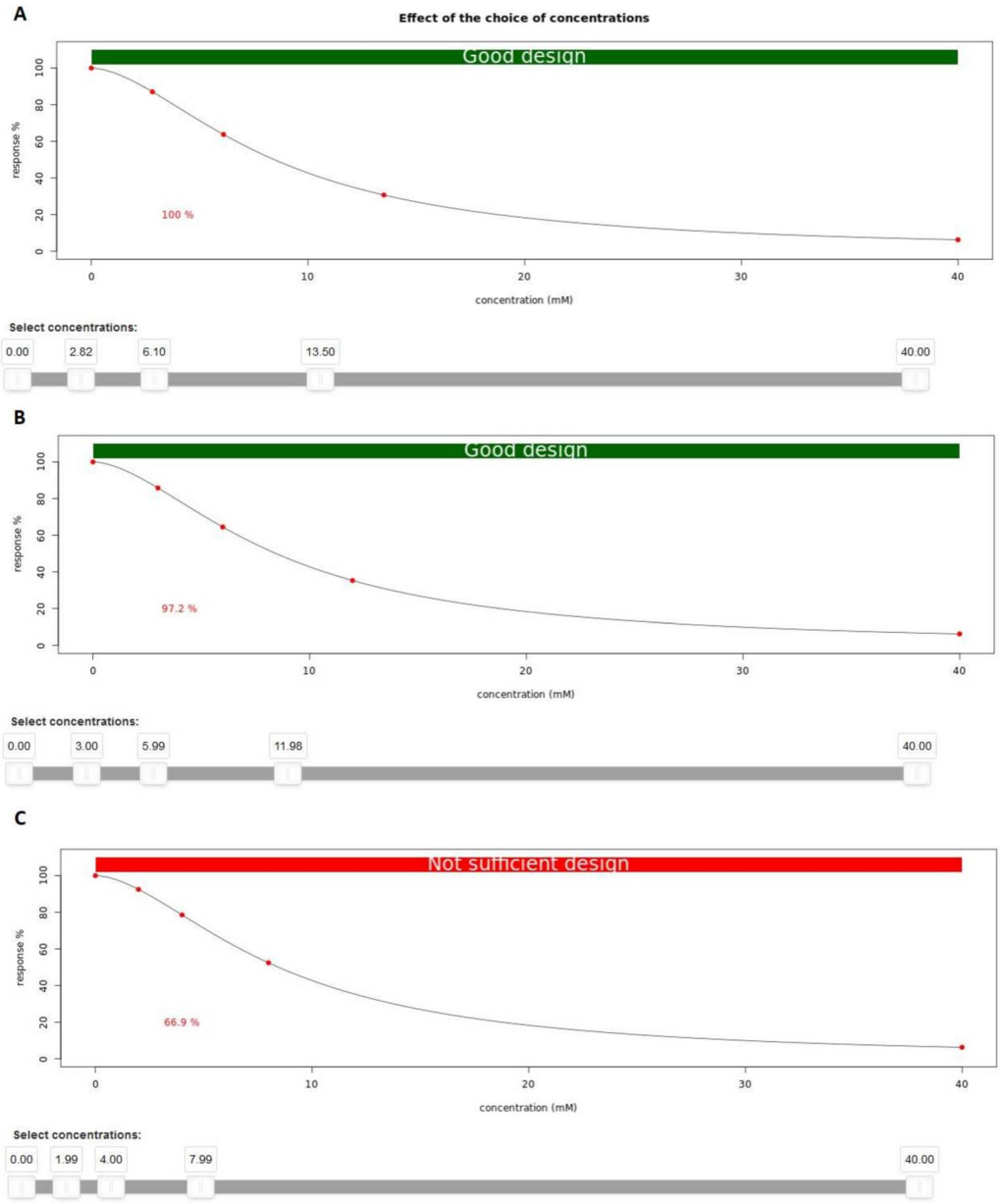



**Fig. 14** Visualization of the influence of the choice of concentrations based on the corresponding efficiency. **A** Optimal concentrations result in a corresponding efficiency of 100%. **B** A different set of concentrations result in a high efficiency of 97.2%. **C** Different set of concentrations result in a low efficiency of 66.9% and lead to a non sufficient design

The results show that the Bayesian design technique leads to a higher model precision compared to the other techniques for all numbers of concentrations. The Bayesian design leads to RMSEs and $EC_{50}$-values with substantially lower variability and therefore better reproducibility of experimental results. Besides, the log-equidistant design with four to six concentrations leads to high deviations from the reference curve. When incorporating the seventh concentration into the log-equidistant design, it performs considerably better. In contrast to this, no substantial improvement is visible for using more concentrations in the IfADo design. Furthermore, the Bayesian design technique requires by far the smallest sample size for achieving the same precision of the $EC_{50}$ estimate as the other designs.

In addition, the influence of a pre-experiment for the three different design techniques was considered. When the pre-experiment was designed with the Bayesian or log-equidistant design with seven concentrations, no sequential approach consisting of a pre- and a main experiment was necessary, because the gained information was already sufficient in the one-step procedure and not improved by splitting the experiments. For the IfADo design with seven

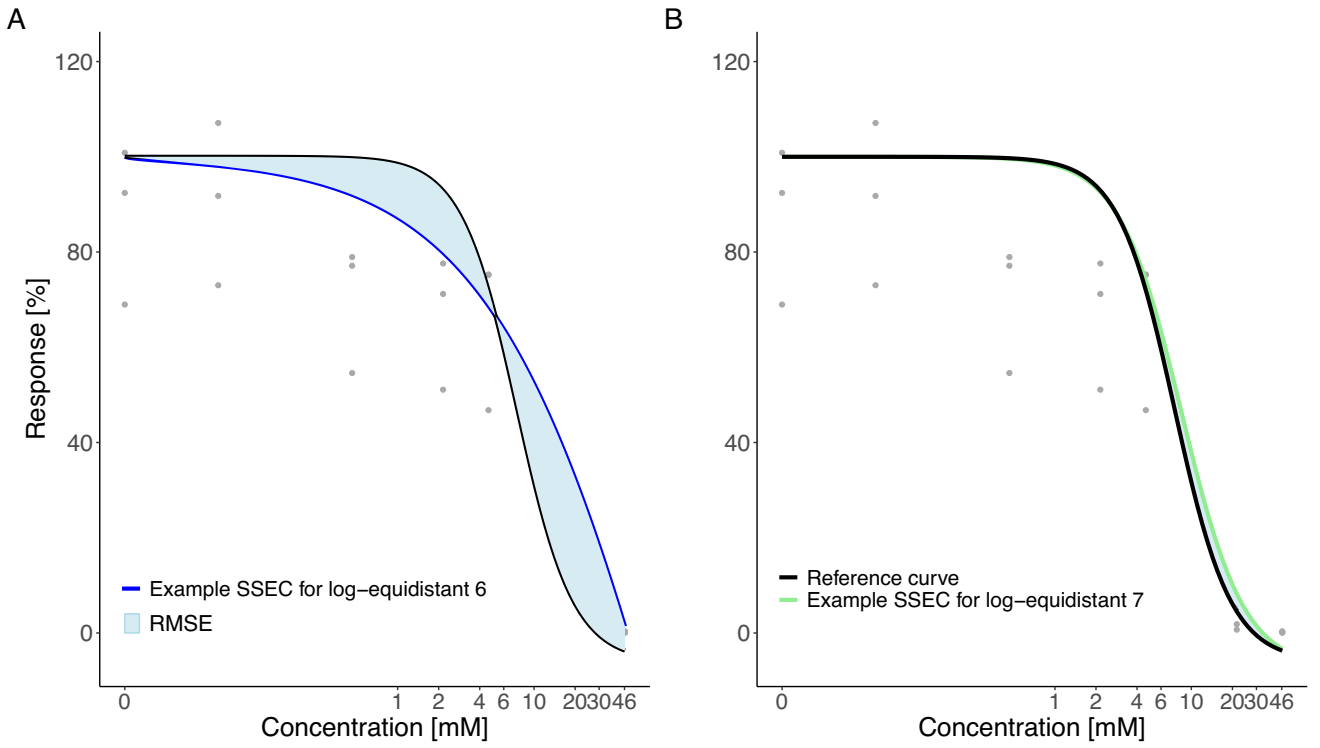



**Fig. 15** The precision of the log-equidistant design is largely influenced (improved) by inclusion of the specific concentration 21.45 mM. **A** Visualization of the RMSE of an example SSEC for the log-equidistant design with six concentrations compared to the reference curve. **B** Visualization of the RMSE of an example SSEC for the log-equidistant design with seven concentrations compared to the reference curve

concentrations in the pre-experiment, an improvement using the sequential approach was visible. In general, the sequential approach tends to be superior, if the design of the pre-experiment is suboptimal.

Whereas the pre-existing knowledge in this scenario is based on the $EC_{10}$- and $EC_{50}$-values of previous experiments identified with 0.373, 0.5645, 0.756 and 6, 6.5, 7 mM, respectively, the $EC_{10}$ and $EC_{50}$ observed in the VPA data set are identified with 2.58 and 7.15 mM, so that the previous knowledge deviates from the reference curve in this analysis. Therefore, the designs based on this previous knowledge can be confounded and improved by incorporating adjusted previous knowledge. Nevertheless, the Bayesian technique performed better compared to the other design techniques.

For the second scenario, a large data set consisting of 76 substances (not including VPA) was used to construct a design, called BayesianDILI design, which is based on less specific prior knowledge of these substances. The BayesianDILI design can be adjusted for each test compound in the DILI data set and related compounds by utilizing the corresponding highest considered concentration. The performance of this design was compared to the three former design techniques on VPA data. Here, the BayesianDILI design tailored to VPA, called BayesianDILI_VPA, performed considerably better than the IfADo design, although it did not use specific prior knowledge of the test substance. The BayesianVPA and log-equidistant design showed a higher model precision compared to the BayesianDILI_VPA design, due to their construction on more specific prior knowledge of VPA. In terms of the precision of the $EC_{50}$ the BayesianDILI_VPA design performed worse than the BayesianVPA design but even better than the log-equidistant or IfADo design. It should be considered that the $EC_{50}$ of VPA does not represent the median of the $EC_{50}$-values of all investigated substances, because VPA is comparatively less toxic. The BayesianDILI design could show better results for substances where the $EC_{50}$ is closer to the median.

Considering the performance of the BayesianDILI design for the 76 substances, that were used for its construction, a computer-based simulation study was performed. Hereby, the BayesianDILI design was compared to log-equidistant designs and the original designs used in the DILI data set in terms of RMSE for each substance, respectively. For the vast majority of substances the BayesianDILI design shows a higher model precision than the log-equidistant and originally used design in terms of the RMSE. The analysis of the precision of the $EC_{50}$ showed coherent results. Note that the BayesianDILI design was developed based on prior information of these 76 substances. However, the prior information of a particular substance entered only with weight 1/76 to

the prior distribution, so the results also indicate a good performance of the BayesianDILI design for the scenario of no pre-existing knowledge. It should be noted that a Bayesian design could have also been constructed with less prior information. However, the quantity and the quality of the prior knowledge increases the efficiency of the Bayesian design technique.

Apart from the Bayesian technique, the log-equidistant design with seven concentrations showed a high model precision, at least for VPA. Nevertheless, it has to be noted that this results from incorporating the specific concentration 21.45 mM. In Fig. 15 the improvement of the model precision when incorporating the seventh concentration in the log-equidistant design is shown exemplarily for one SSEC.

In this study, we considered only designs with equal numbers of replicates at each concentration for practical reasons. However, this limits the calculation of optimal designs, as they are not inherently equally weighted by definition.

Furthermore, we prioritized a user-friendly approach by extending the Bayesian design using a simple midpoint interpolation technique. While other methods, such as optimizing additional concentrations based on the criterion value, could have been applied, they would likely yield similar concentrations. Our approach ensures that the original Bayesian design concentrations are preserved while broadly covering the design space. This is advantageous in practice, as it facilitates the identification of the entire concentration-response curve.

The results clearly demonstrate the advantages of using techniques of optimal design theory in practice. The usage of the Bayesian technique with and without incorporating pre-existing knowledge of a specific test substance resulted in the most precise statistical inference of the corresponding experiments. Here, the approach is used for mostly hepatotoxic substances, but it can be easily adjusted to different substance sets.

To provide an easy calculation tool, we developed a Shiny app that allows users to easily determine optimal concentrations for future cytotoxicity experiments. Additionally, this Shiny app enables the comparison of various concentrations against the optimal values.

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s00204-024-03893-1.

**Acknowledgements** We thank Astra Zeneca for providing some of the test compounds for the DILI data set. Furthermore, we would like to thank the reviewer for their valuable comments, which have enhanced the quality of this manuscript.

**Author Contributions** All authors provided a concept of this work. CP, WA and TB performed the cytotoxicity experiments and provided the data set. LS, PB and KS analyzed the results. LS provided the Shiny App. LS, WA, JH and KS wrote the manuscript. LS, TB, PB, JH and KS reviewed the manuscript.

**Funding** Open Access funding enabled and organized by Projekt DEAL. This work has been supported by the Research Training Group "Biostatistical Methods for High-Dimensional Data in Toxicology" (RTG 2624, Project P5) funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation - Project Number 427806116).

**Availability of data and materials** The data set as it is used in this publication can be found in the supplementary material (see Supplement Data).

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethical approval** Not applicable.

**Consent to participate** Not applicable.

**Consent for publication** Not applicable.

**Code availability** The corresponding Shiny app can be found online openly accessible at http://shiny.statistik.tu-dortmund.de:8080/app/occe.



## References

Albrecht W, Kappenberg F, Brecklinghaus T et al (2019) Prediction of human drug-induced liver injury (DILI) in relation to oral doses and blood concentrations. Arch Toxicol 93:1609–1637

Bezanson J, Edelman A, Karpinski S et al (2017) Julia: a fresh approach to numerical computing. SIAM Rev 59(1):65–98. https://doi.org/10.1137/141000671

Brecklinghaus T, Albrecht W, Duda J et al (2022) In vitro/in silico prediction of drug induced steatosis in relation to oral doses and blood concentrations by the Nile Red assay. Toxicol Lett 368:33–46

Brecklinghaus T, Albrecht W, Kappenberg F et al (2022) Influence of bile acids on the cytotoxicity of chemicals in cultivated human hepatocytes. Toxicol In Vitro 81:105344

Chaloner K, Larntz K (1989) Optimal Bayesian design applied to logistic regression experiments. J Stat Plan Inference 21(2):191–208

Chernoff H (1953) Locally optimal designs for estimating parameters. Ann Math Stat 24:586–602

Ghallab A, Hassan R, Hofmann U et al (2022) Interruption of bile acid uptake by hepatocytes after acetaminophen overdose ameliorates hepatotoxicity. J Hepatol 77(1):71–83

Gu X, Albrecht W, Edlund K et al (2018) Relevance of the incubation period in cytotoxicity testing with primary human hepatocytes. Arch Toxicol 92:3505–3515

Holland-Letz T, Kopp-Schneider A (2021) An R-shiny application to calculate optimal designs for single substance and interaction trials in dose response experiments. Toxicol Lett 337:18–27

Jennrich RI (1969) Asymptotic properties of non-linear least squares estimators. Ann Math Stat 40(2):633–643

Kappenberg F, Brecklinghaus T, Albrecht W et al (2020) Handling deviating control values in concentration–response curves. Arch Toxicol 94(11):3787–3798

Kappenberg F, Duda JC, Schürmeyer L et al (2023) Guidance for statistical design and analysis of toxicological dose–response experiments, based on a comprehensive literature review. Arch Toxicol 97:2741–2761

Khetani SR, Kanchagar C, Ukairo O et al (2013) Use of micropatterned cocultures to detect compounds that cause drug-induced liver injury in humans. Toxicol Sci 132(1):107–117

Kiefer J (1974) General equivalence theory for optimum designs (approximate theory). Ann Stat 2(5):849–879

McGree JM, Eccleston JA, Duffull SB (2008) Compound optimal design criteria for nonlinear models. J Biopharm Stat 18(4):646–661. https://doi.org/10.1080/10543400802071352

Proctor WR, Foster AJ, Vogt J et al (2017) Utility of spherical human liver microtissues for prediction of clinical drug-induced liver injury. Arch Toxicol 91:2849–2863

Pukelsheim F (2006) Optimal design of experiments. SIAM, Philadelphia

Pukelsheim F, Rieder S (1992) Efficient rounding of approximate designs. Biometrika 79(4):763–770

R Core Team (2022) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Ritz C, Baty F, Streibig JC et al (2015) Dose-response analysis using R. PLoS One 10:e0146021

Sandig L (2024) Kirstine.jl: a julia package for Bayesian optimal design of experiments. J Open Source Softw 9(97):6424. https://doi.org/10.21105/joss.06424

Silvey SD (1980) Optimal design. Chapman and Hall, London

Vorrink SU, Zhou Y, Ingelman-Sundberg M et al (2018) Prediction of drug-induced hepatotoxicity using long-term stable primary hepatic 3D spheroid cultures in chemically defined conditions. Toxicol Sci 163(2):655–665

# Supplementary material

## Sequential approach improves IfADo design

Concentration-dependent toxicity testing might be improved by using a sequential procedure, in which the main experiment is done after the conduction of a pre-experiment. The pre-experiment is used to obtain knowledge about the $EC_{10}$-value and the $EC_{50}$-value of the test substance. Based on this information, the main experiment is performed to gain detailed knowledge about the cytotoxicity of the test substance.

To analyze the influence of the choice of the design conditions the design is varied in the pre-experiment with three possibilities (log-equidistant, IfADo and Bayesian) with seven concentrations each. Based on the VPA data set scenario specific estimated curves (SSECs) as described in Section Results were calculated 3000 times with three observations at each concentration. Based on each of those SSECs the confidence intervals of the corresponding $EC_{10}$ and $EC_{50}$ are calculated. With the help of this, parameter combinations were constructed to incorporate uncertainty and a prior distribution is established based on the parameter combinations. Afterwards a Bayesian $D$-optimal design with seven concentrations can be developed based on this prior distribution (which is unique for each SSEC) assuming equal weights. To make sure the resulting concentrations are available in the VPA data set, the concentrations of the SSEC specific Bayesian design were fixed to the seven different nearest available concentrations of the data set. The main experiment then is conducted with these specific concentrations resulting from the Bayesian technique. In a last step, three observations per concentration were drawn randomly from all data points available in the data set to imitate a main experiment. Additionally, a 4pLL-model is fitted to the new data points. This model then can be compared to the reference curve via RMSE and precision of the $EC_{50}$. Note that the other design techniques were not used for the design of the main experiment, as they performed worse than the Bayesian design technique with seven concentrations in the previous analyses.

The RMSEs and $EC_{50}$-value of the sequential approach were compared to the RMSEs and the $EC_{50}$-value of the non-sequential procedure, in which a main experiment with six replicates at each concentration was conducted based on either the Bayesian, log-equidistant or IfADo design with seven concentrations. Note that for the sake of comparability, the number of replicates was chosen such that both the non-sequential and the sequential procedures result in the same total sample size.

The performance of the log-equidistant design and the IfADo design is improved by the sequential approach, whereas the performance of the Bayesian design is not affected (see Fig. 1). More precisely, the results show that neither the median RMSE nor the median $EC_{50}$-value is improved by the sequential approach if the pre-experiment had already been based on the Bayesian technique. In contrast, the RMSEs and the $EC_{50}$-values of the IfADo design were improved substantially by the sequential procedure compared to the non-

sequential approach. However, the sequential IfADo approach remains worse than the non-sequential Bayesian approach. Focusing on the log-equidistant design, the sequential approach reduces the number and extension of outliers in terms of RMSEs and $EC_{50}$-values.

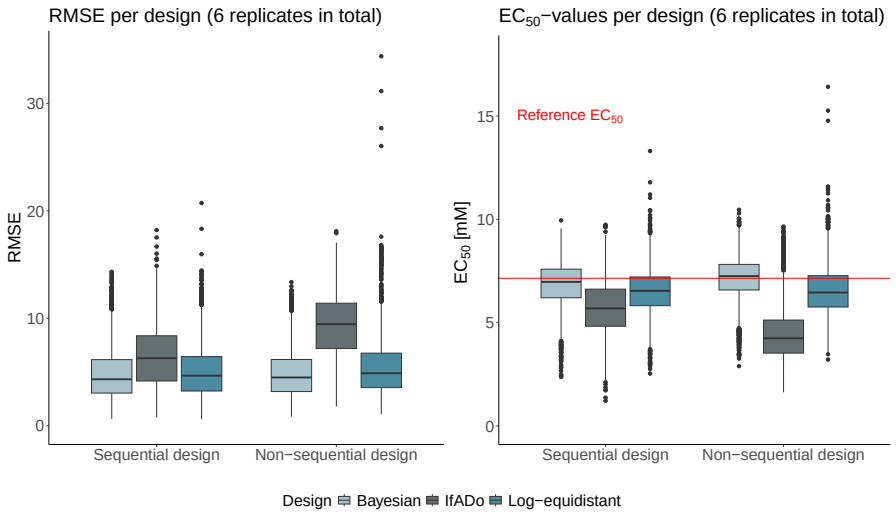



Figure 1: IfADo and log-equidistant design are improved by the sequential design procedure in terms of precision and reduction of outliers.
A: RMSE-values for sequential and non-sequential approach grouped by pre-experimental condition (different designs), B: $EC_{50}$-values for sequential and non-sequential approach grouped by pre-experimental condition (different designs). The red line corresponds to the reference $EC_{50}$.

# Supplementary material

## Figures

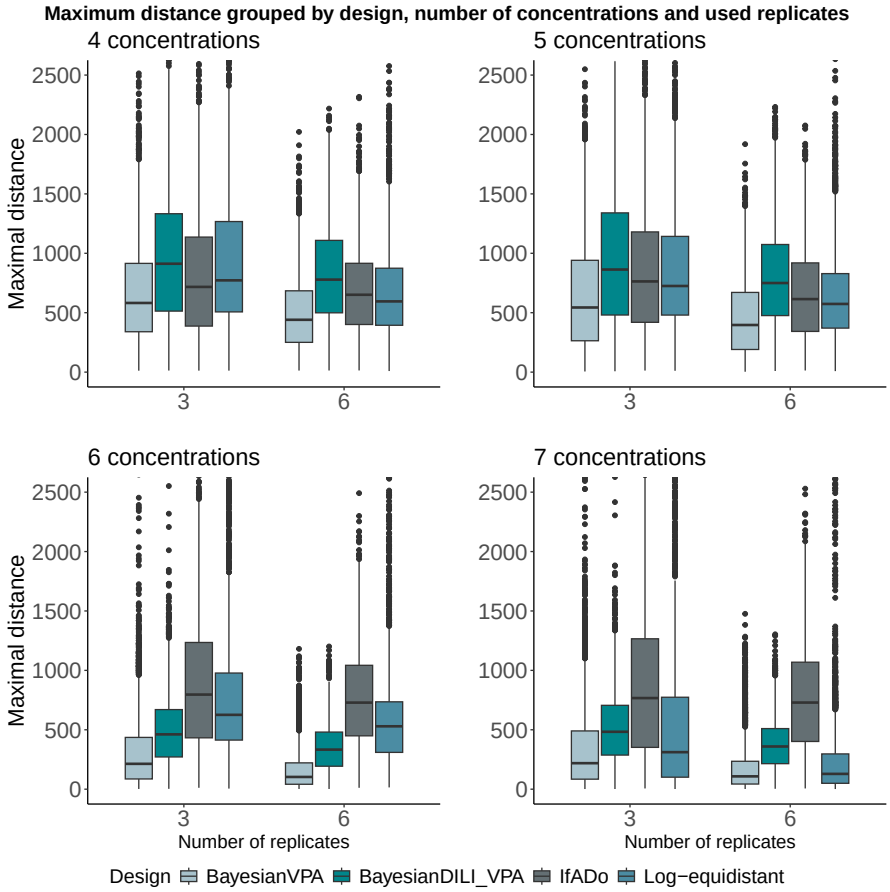



Figure S1: Maximum distance grouped by design, number of concentrations, and the use of replicates. This figure presents only an extract of the data, excluding extreme outliers across all designs.

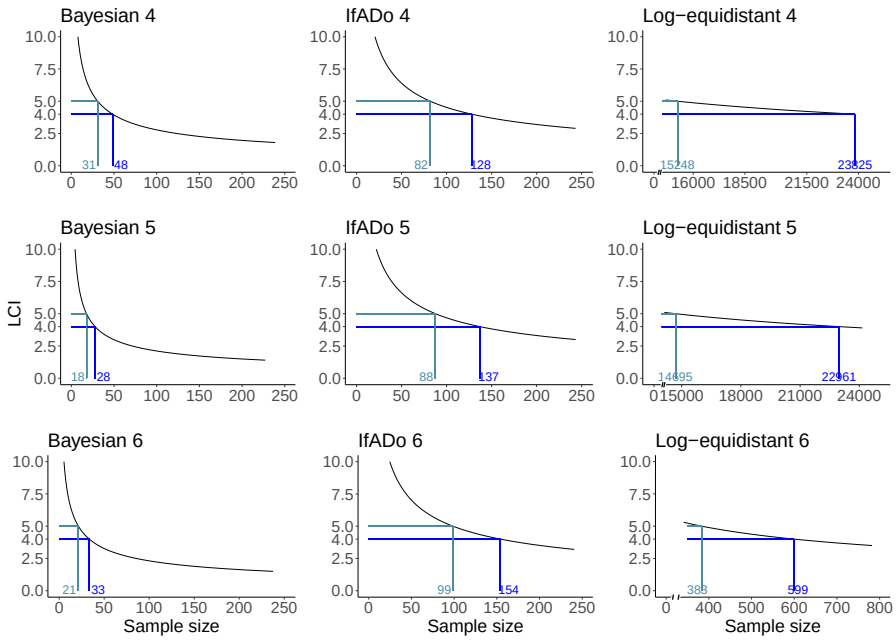



Figure S2: The relation between the sample size and the length of the confidence interval (LCI) of the $EC_{50}$-value is shown grouped by design and number of concentrations. The blue lines indicate the number of observations required to achieve a specific level of precision for the $EC_{50}$ (LCI). To achieve the same precision of the $EC_{50}$ the Bayesian design requires considerably less observations than the other two design techniques.

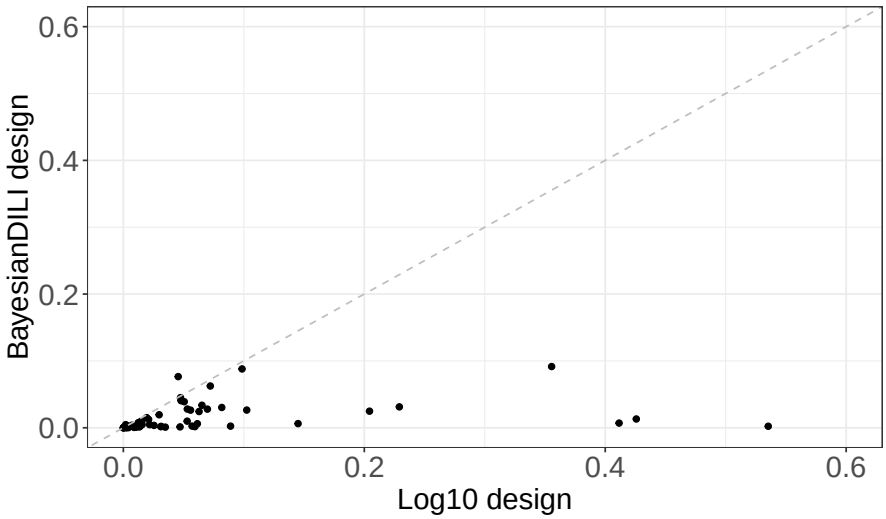



Figure S3: An extract of the simulation results from all 76 test compounds is shown regarding the precision of the $EC_{50}$ for the BayesianDILI compared to the log10 design. The precision is measured by the mean distance of the simulated $EC_{50}$-values compared to the reference $EC_{50}$ for each substance. Each dot represents one substance. For the sake of comparability, too extreme values are not shown in this plot. The BayesianDILI design outperforms the log10 design for the majority of substances.

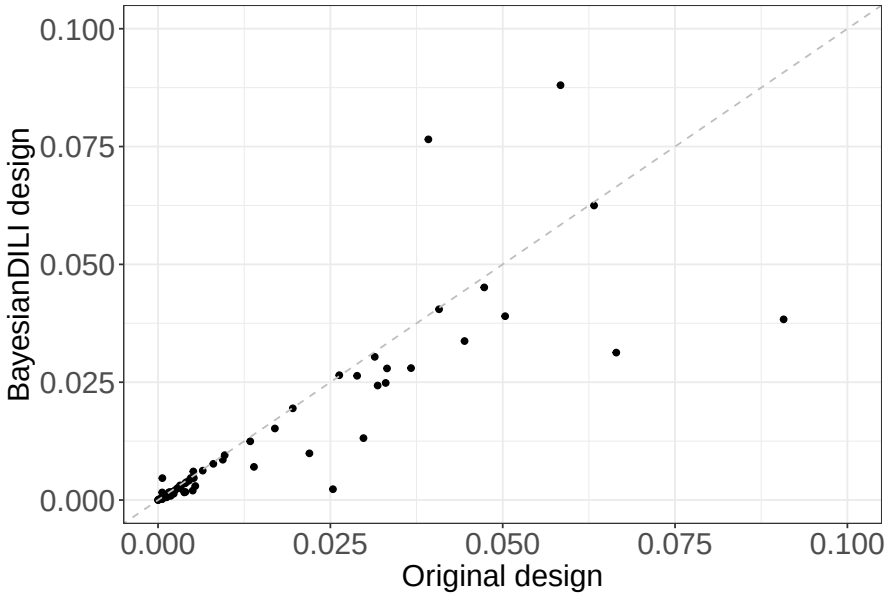



Figure S4: An extract of the simulation results from all 76 test compounds is shown regarding the precision of the $EC_{50}$ for the BayesianDILI compared to the original used design. The precision is measured by the mean distance of the simulated $EC_{50}$-values compared to the reference $EC_{50}$ for each substance. Each dot represents one substance. For the sake of comparability, too extreme values are not shown in this plot. The BayesianDILI design outperforms the original design for the majority of substances.

# 2. Article: Optimal designs for identifying effective doses in drug combination studies

# OPTIMAL DESIGNS FOR IDENTIFYING EFFECTIVE DOSES IN DRUG COMBINATION STUDIES

**Leonie Schürmeyer**
Department of Statistics
TU Dortmund University
Dortmund, 44227, Germany
schuermeyer@statistik.tu-dortmund.de

**Ludger Sandig**
Department of Statistics
TU Dortmund University
Dortmund, 44227, Germany

**Leonie Theresa Hezler**
Global Biostatistics and Data Sciences
Boehringer Ingelheim Pharma GmbH & Co. KG
Biberach an der Riß, 88397, Germany

**Bernd-Wolfgang Igl**
Global Biostatistics and Data Sciences
Boehringer Ingelheim Pharma GmbH & Co. KG
Biberach an der Riß, 88397, Germany

**Kirsten Schorning**
Department of Statistics
TU Dortmund University
Dortmund, 44227, Germany
schorning@statistik.tu-dortmund.de

June 9, 2025

**ABSTRACT**

We consider the optimal design problem for identifying effective dose combinations within drug combination studies where the effect of the combination of two drugs is investigated. Drug combination studies are becoming increasingly important as they investigate potential interaction effects rather than the individual impacts of the drugs. In this situation, identifying effective dose combinations that yield a prespecified effect is of special interest. If nonlinear surface models are used to describe the dose combination-response relationship, these effective dose combinations result in specific contour lines of the fitted response model.

We propose a novel design criterion that targets the precise estimation of these effective dose combinations. In particular, an optimal design minimizes the width of the confidence band of the contour lines of interest. Optimal design theory is developed for this problem, including equivalence theorems and efficiency bounds. The performance of the optimal design is illustrated in several examples modeling dose combination data by various nonlinear surface models. It is demonstrated that the proposed optimal design for identifying effective dose combinations yields a more precise estimation of the effective dose combinations than commonly used ray or factorial designs. This particularly holds true for a case study motivated by data from an oncological dose combination study.

***K*eywords** Confidence band; Drug combination; Multivariate effective doses; Nonlinear surface modeling; Optimal design

## 1 Introduction

In various disciplines of pharmaceutical development, analyzing potential drug-drug interactions is fundamental, and drug combination studies are becoming increasingly relevant (see e.g., Chou [2008] and Mokhtari et al. [2017] among

many others). In general, the detection of positive drug-drug interactions might result in an increased efficacy, reduced adverse events, and cost reduction. For this purpose, one focus of dose combination studies is on identifying dose combinations that achieve a prespecified effect.

However, little research is available on modeling dose combination-response data and designing experiments for dose combinations in practice [Holland-Letz and Kopp-Schneider, 2018]. Often, the individual dose-response relationships of the monotherapies are known, but their combination is of particular interest, especially when the mode of action is unknown. In this situation, there are different approaches to specify the type of interaction, i.e. whether there is synergism, additivity, or antagonism of the investigated drugs, which correspond to a positive interaction, no interaction, or a negative interaction, respectively (see e.g., Loewe [1927], Bliss [1939] and Lee et al. [2007]). Here, common strategies are, for example, the curve-shift analysis, isobolograms, combination indices, and universal surface response analysis (see e.g., Zhao et al. [2010] and Foucquier and Guedj [2015]). While modeling the drug-drug interaction based on the combination index or isobolograms results in a one-dimensional linkage not capturing the response effect itself at each possible dose combination, dose response surface modeling enables an estimation of the drug-drug interaction across the entire design space based on the response effect. Therefore, Zhou et al. [2024] proposes to model dose combination response data using specific nonlinear surface models. The two-dimensional setting of the surface models introduces new challenges. Similar to traditional one-dimensional dose-response experiments, researchers are interested not only in modeling the effects across the entire design space but also in evaluating specific effective doses, which are defined as the dose combinations that achieve a predetermined level of the effect. In the one-dimensional setting, the dose that results in a specified effect level is unique, in particular, if the response is strictly increasing in the dose. In the two-dimensional setting, multiple dose combinations yield a specific effect level, forming contour lines on the surface model, even if the response is strictly increasing in both doses. Therefore, a design that facilitates precise estimation of these effective dose combinations in drug combination studies is needed.
This paper aims to construct efficient designs for identifying effective dose combinations using nonlinear surface models. Although the theoretical and practical considerations of optimal designs for different kinds of classical one-dimensional dose-response models are well established (see, e.g., Bretz et al. [2010], Dette and Schorning [2016] and Bornkamp et al. [2009]), there is less research on the design of drug combination experiments. Ray designs, whose dose levels correspond to fixed proportions of the mixture of substances, are frequently used in drug combination studies (see e.g., Straetemans et al. [2005] and Rønneberg et al. [2021]). Therefore, Almohaimeed and Donev [2014] established locally D-optimal designs for specific ray designs. Holland-Letz and Kopp-Schneider [2018] consider optimal experimental designs for estimating the drug combination index based on Loewe additivity. Besides, Papathanasiou et al. [2019] consider nonlinear surface models to describe the dose combination response relationship across the whole design space but mainly address the design problem by using the classical D-optimality criterion instead of explicitly targeting the precise estimation of effective dose combinations. To address this gap, we propose a new and flexible design criterion that aims to estimate effective dose combinations in drug combination studies accurately.

The paper is structured as follows. First, a suitable definition of effective dose combinations will be given in Section 2, where we will also present surface modeling as used in drug combination studies. In Section 3, we will introduce the optimal design theory in the context of drug-combination studies aiming to estimate the effective dose combinations accurately. More precisely, an optimal design minimizes the $L_q$-norm of the variance of the prediction for the effect at the effective dose-combinations. Note that a similar approach is proposed by Miller et al. [2007] for the one dimensional set-up, but can not be directly extended to the present situation of dose combination studies. For the present context, we will provide equivalence theorems and a lower bound for the efficiencies that can even be used if the optimal design is unknown.
In practice, the derived optimal designs should result in more precise predictions of the effective dose-combinations. Therefore, Section 4 and Section 5 are devoted to illustrating the advantages of the derived optimal designs in specific scenarios of dose combination studies. Based on a real case study, we will compare the performance of the derived optimal designs against established designs such as factorial and ray designs and classical locally D-optimal design. The considered design approaches will be compared theoretically, based on efficiencies, and practically through a simulation study. In addition to the case study, further scenarios of dose combination studies that, in particular, reflect different types of interactions and surface models will be evaluated within a simulation framework.
Finally, we examine the robustness of the developed design approach. In drug combination studies, monotherapies are often known in advance, while the interaction effect may not be clearly defined before conducting the drug combination experiment. Therefore, we consider robust designs for various values of the interaction effect. Additionally, we investigate the capability of the resulting designs to provide precise estimates of the effective doses across different scenarios and assess their performance under these varying conditions.

Table 1: Possible choices of the one-dimensional regression functions for $\eta_C$ and $\eta_D$ in the dose combination model (1). Note that the placebo parameter of the depicted regression functions from Bornkamp et al. [2009] is removed, as it is already part of the dose combination model (1).

| Model | Formula | Parameters |
|---|---|---|
| Linear | $f(d, \theta_f) = \delta d$ | $\theta_f = \delta$ |
| Exponential | $f(d, \theta_f) = \mathrm{E}_1(\exp(\frac{d}{\delta}) - 1)$ | $\theta_f = (\mathrm{E}_1, \delta)$ |
| Emax | $f(d, \theta_f) = \mathrm{E}_{\max} \frac{d}{\mathrm{ED}_{50}+d}$ | $\theta_f = (\mathrm{E}_{\max}, \mathrm{ED}_{50})$ |
| sigmoid Emax | $f(d, \theta_f) = \mathrm{E}_{\max} \frac{d^h}{\mathrm{ED}_{50}^h+d^h}$ | $\theta_f = (\mathrm{E}_{\max}, \mathrm{ED}_{50}, h)$ |

## 2 Modeling dose combination data

We consider the effect of the combination of two different substances, denoted by $C$ and $D$, by modeling it with the nonlinear regression model

$$Y_{ij} = \eta\big((c_i, d_i), \theta\big) + \varepsilon_{ij} \quad \text{for } i = 1, \ldots, n; j = 1, \ldots, r_i,$$

where $\varepsilon_{ij}$ are independent random variables such that $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$, $\sigma^2 > 0$. This means that observations are taken at $n$ different dose combinations $(c_1, d_1), \ldots, (c_n, d_n)$, which vary in the design space (i.e., the dose combination space) $\mathcal{Z} = \mathcal{X}_C \times \mathcal{X}_D = [0, c_{\max}] \times [0, d_{\max}] \subset \mathbb{R}^2$ and $r_i$ observations are taken at each $(c_i, d_i)$, $i = 1, \ldots, n$. Let $N = \sum_{i=1}^n r_i$ denote the total sample size. The regression model $\eta$ and the $m$-dimensional parameter $\theta$ are used to describe the relationship between the response and the dose combination. We assume that the function $((c, d), \theta) \mapsto \eta((c, d), \theta)$ is continuously differentiable both in $(c, d)$ and in $\theta$.

Natural choices for the regression function $\eta$ include the regression functions used to describe the dose-specific effect of the individual substances. More precisely, let $\eta_C(\,\cdot\,, \theta_C)$, $\theta_C \in \Theta_C \subset \mathbb{R}^{m_C}$, and $\eta_D(\,\cdot\,, \theta_D)$, $\theta_D \in \Theta_D \subset \mathbb{R}^{m_D}$ be the parts of the individual regression functions describing the dose-specific effect of the substances $C$ and $D$, respectively. Then we define the regression function for the dose combination

$$\eta\big((c, d), \theta\big) = \theta_0 + \eta_C(c, \theta_C) + \eta_D(d, \theta_D) + \gamma \eta_C(c, \theta_C)\eta_D(d, \theta_D)\,, \tag{1}$$

where the parameter $\theta = (\theta_0, \theta_C, \theta_D, \gamma)$ consists of the individual parameters $\theta_C$ and $\theta_D$ of the regression functions $\eta_C$ and $\eta_C$, a joint placebo parameter $\theta_0$ and the parameter $\gamma \in \mathbb{R}$ that describes potential dose-independent interaction effects between the two substances. Note, that due to identifiability reasons the parameters $\theta_C$ and $\theta_D$ do not contain a substance specific parameter for the placebo-effect. The interaction effect is positive if $\gamma > 0$, whereas $\gamma < 0$ indicates a negative interaction. If $\gamma = 0$, there is no interaction between the two substances. Summarizing, for the regression function $\eta$ in (1) it holds for the dimension of the parameter $\theta$ that $m = 2 + m_C + m_D$. In clinical and toxicological applications popular choices for the function $\eta_C$ and $\eta_D$ are the linear, exponential, Emax, or sigmoid Emax model, as described by Bornkamp et al. [2009] and presented in Table 1. Note that the functions depicted in Table 1 are continuously differentiable, hence also $\eta$ is differentiable.

For asymptotic arguments we assume that $\lim_{N\to\infty} \frac{r_i}{N} = \omega_i \in (0, 1)$ and collect this information in the matrix

$$\xi = \begin{pmatrix} (c_1, d_1) & \cdots & (c_n, d_n) \\ \omega_1 & \cdots & \omega_n \end{pmatrix}.$$

Following Kiefer [1974], we refer to $\xi$ as an approximate design on the design space $\mathcal{Z}$. This means that the dose combinations $(c_i, d_i)$ define the different experimental conditions where observations are to be taken, and the weights $\omega_i$ represent the relative proportion of observations at the corresponding dose combination $(c_i, d_i)$. If an approximate design is given and $N$ observations can be taken, a rounding procedure proposed by Pukelsheim and Rieder [1992] is applied to obtain integers $r_i$ from the not necessarily integer valued quantities $N\omega_i$. Assume that observations are taken according to an approximate design $\xi$. Under certain assumptions of regularity, the distribution of $\hat{\theta}$ is asymptotically normal [Jennrich, 1969]. Moreover, the distribution of the predicted effect at a specific dose combination $(c_0, d_0)$ is also asymptotically normal, with

$$\sqrt{N}\Big(\eta\big((c_0, d_0), \hat{\theta}\big) - \eta\big((c_0, d_0), \theta\big)\Big) \xrightarrow{\mathcal{D}} \mathcal{N}\Big(0, \varphi\big((c_0, d_0), \xi, \theta\big)\Big).$$

Here, $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution and the function $\varphi$ is defined by

$$\varphi\big((c_0,d_0),\xi,\theta\big) = \left(\frac{\partial}{\partial\theta}\eta\big((c_0,d_0),\theta\big)\right)^T M^-(\xi,\theta)\left(\frac{\partial}{\partial\theta}\eta\big((c_0,d_0),\theta\big)\right), \tag{2}$$

$$\text{where} \quad M(\xi,\theta) = \int_Z \frac{\partial}{\partial\theta}\eta\big((c,d),\theta\big)\left(\frac{\partial}{\partial\theta}\eta\big((c,d),\theta\big)\right)^T d\xi\big((c,d)\big) \tag{3}$$

is the information matrix corresponding to the design $\xi$, and $\frac{\partial}{\partial\theta}\eta((c,d),\theta)$ is the gradient of $\eta$ with respect to the parameter $\theta \in \mathbb{R}^m$. The matrix $M^-(\xi,\theta)$ is the generalized inverse of the information matrix, whereby the design $\xi$ has to satisfy $\frac{\partial}{\partial\theta}\eta((c_0,d_0),\theta) \in \text{range}(M(\xi,\theta))$.

Therefore, the asymptotic variance of the prediction $\eta((c_0,d_0),\hat{\theta})$ at a prespecified dose combination is given by $\varphi((c_0,d_0),\xi,\theta)$. The asymptotic behavior in (2) can be used to construct an asymptotic confidence interval for the prediction of the effect at the dose combination $(c_0,d_0)$. More precisely, the confidence interval for the level $(1-\alpha)$ is given by

$$\eta\big((c_0,d_0),\hat{\theta}\big) \pm z_{1-\alpha/2}\frac{\hat{\sigma}}{\sqrt{N}}\varphi^{1/2}\big((c_0,d_0),\hat{\theta}\big), \tag{4}$$

where $\hat{\sigma}$ and $\hat{\theta}$ denote the maximum likelihood estimates of the parameters $\sigma$ and $\theta$, $z_{1-\alpha/2}$ denotes the $(1-\alpha/2)$-quantile of the standard normal distribution and the function $\varphi$ is given in (2).

### 2.1 Effective doses in dose combination studies

In classical dose-response experiments where the effect of one substance is investigated, special interest lies on effective doses (ED) that result in a certain percentage of the maximal effect. In particular, following Bretz et al. [2010], the $\text{ED}_p$-dose at which $p\%$ $(0 < p < 100)$ of the maximal effect is achieved is defined by

$$\text{ED}_p = \min\left\{x \in (a,b] \;\middle|\; \frac{h(x,\theta)}{h(b,\theta)} = \frac{p}{100}\right\}, \tag{5}$$

where $h(x,\theta) = f(x,\theta) - f(0,\theta)$ and $f(x,\theta)$ is the considered regression function, which depends on both the dose $x$ and the unknown parameter $\theta$. Note that using the definition in (5) results in a unique $\text{ED}_p$. If the regression function $f(x,\theta)$ is additionally strictly increasing, the effect level $p$ is attained at exactly one dose.

However, in the two-dimensional setting of dose combination studies, there is no unique $\text{ED}_p$-dose combination, even if the regression function $\eta((c,d),\theta)$ is strictly increasing in both doses $c$ and $d$. Instead, several different dose combinations $(c,d) \in \mathcal{Z}$ yield the same effect, resulting in contour lines on the surface model. Extending the one-dimensional definition of the $\text{ED}_p$ in (5), the set of multivariate effective doses for achieving $p\%$ $(0 < p < 100)$ of the maximum effect in the considered design range $\mathcal{Z} = [0, c_{\max}] \times [0, d_{\max}]$ can be defined by

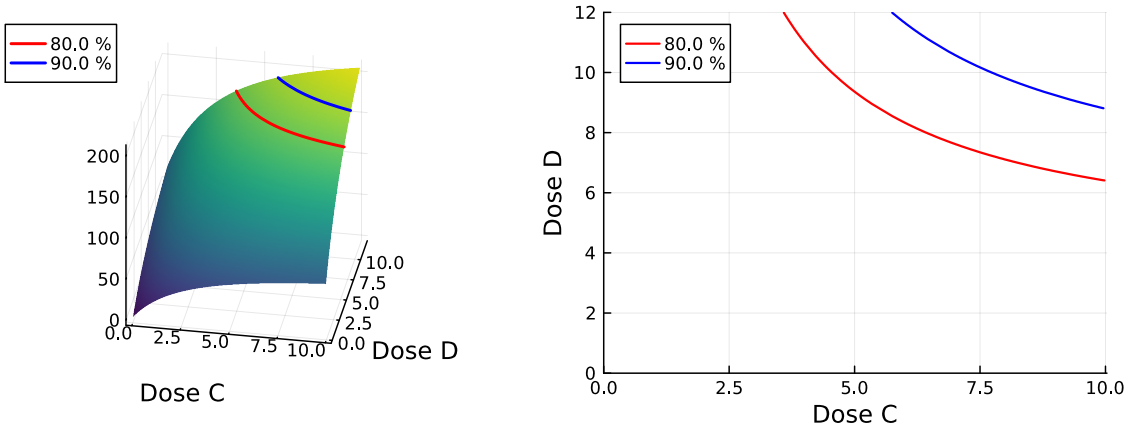
$$\text{MED}_p(\theta) = \left\{(c,d) \in \mathcal{Z} \;\middle|\; \frac{\eta\big((c,d),\theta\big) - \min_{(c_0,d_0)\in\mathcal{Z}}\eta\big((c_0,d_0),\theta\big)}{R_{\max}} = \frac{p}{100}\right\}, \tag{6}$$

where the $R_{\max}$ denotes the maximal effect of the response in the design space and is given by

$$R_{\max} = \max_{(c,d)\in\mathcal{Z}}\eta\big((c,d),\theta\big) - \min_{(c,d)\in\mathcal{Z}}\eta\big((c,d),\theta\big).$$

Note that the $\text{MED}_p(\theta)$ defined in (6) depends on the parameter $\theta$. For the sake of brevity, we denote the multivariate effective doses for $p$ % of the maximal effect with $\text{MED}_p$.

**Example 1.** Consider the situation where dose combinations of the substances $C$ and $D$ that achieve $p_1 = 80\%$ and $p_2 = 90\%$ of the maximal effect are of interest in a dose combination study. Further, assume that the dose combination response relationship can be described by an interaction model (1), where the individual regression functions are given by Emax models with parameter $\theta_C = (80, 3)$ and $\theta_D = (120, 10)$, respectively. The interaction effect is given by $\gamma = 0.02$. The corresponding combination response surface model is shown in Figure 1 (a). The dose combinations that result in $80\%$ and $90\%$ of the maximal effect, thus the sets $\text{MED}_{80}$ and $\text{MED}_{90}$, are marked by the red and orange (contour) line, respectively (Figure 1 (b)).




(a) Surface model with contour lines $p_1 = 80$, $p_2 = 90$.

(b) Visualization of the $\text{MED}_{80}$ and $\text{MED}_{90}$.

Figure 1: Response surface and multivariate effective doses for the model from Example 1.

## 3 Optimal design theory for identifying sets of effective doses in drug combination studies

Dose combination studies aim to identify the dose combinations $(c, d) \in \text{MED}_p$ for prespecified values for $p\%$. Using (4) the pointwise confidence band of the prediction of effects at the dose combinations contained in the $\text{MED}_p$ is given by

$$\left\{ \eta\big((c,d),\hat{\theta}\big) \pm z_{1-\alpha/2} \frac{\hat{\sigma}}{\sqrt{N}} \varphi^{1/2}\big((c,d),\xi,\hat{\theta}\big) \;\middle|\; (c,d) \in \text{MED}_p \right\}, \tag{7}$$

where the asymptotic variance function $\varphi((c,d),\xi,\hat{\theta})$ is defined in (2). The smaller the width of this confidence band is, the more precise the prediction of the effect at dose combinations $(c, d)$ contained in $\text{MED}_p$ becomes, which in turn leads to a precise identification of the $\text{MED}_p$. Consequently, from a design perspective, a good design $\xi$ should minimize the width of the confidence band in (7) at each dose combination $(c, d) \in \text{MED}_p$. This corresponds to a minimization of the asymptotic variance in (2) concerning the design $\xi$. Unfortunately, a simultaneous minimization is only possible in rare and probably unrealistic settings. We therefore propose a design criterion to minimize a $L_q$-norm of the function $\varphi$, $q \in [1, \infty)$. More precisely, let $p_1, \ldots, p_k \in (0, 100)$ be $k \geq 1$ prespecified percentages and $\mathcal{C}(\theta) = \bigcup_{i=1}^{k} \text{MED}_{p_i} \neq \emptyset$ the joint set of the corresponding MED-sets. Then, we can use the $L_q$-norm

$$\phi_{\text{MED}_q}(\xi,\theta) = \left( \frac{1}{l_{\mathcal{C}(\theta)}} \int_{\mathcal{C}(\theta)} \varphi^q\big((c,d),\xi,\theta\big) d\mu(c,d) \right)^{\frac{1}{q}} \tag{8}$$

of the function $\varphi$ defined in (2), where $\mu$ denotes an appropriate measure on the design space $\mathcal{Z}$ and $l_{\mathcal{C}(\theta)} = \mu\big(C(\theta)\big) > 0$ the corresponding value of the measure $\mu$ evaluated at the set $C(\theta)$. Typical choices for the measure $\mu$ are the uniform distribution on the set $C(\theta)$ or discrete measures on the design space $\mathcal{Z}$ (as later chosen in the simulation study in Section 5.2). Note that the function in (8) is similar to the I-criterion that is discussed in [Fedorov and Leonov, 2013, p. 58].

**Definition 2.** Let $p_1, \ldots, p_k \in (0, 100)$, $k \geq 1$, $q \in [1, \infty)$. Then a design $\xi^*$ is called locally $\text{MED}_q$-optimal for the $k$ effective dose combinations $\text{MED}_{p_1}, \ldots, \text{MED}_{p_k}$ and the parameter $\theta$, if it satisfies $\frac{\partial}{\partial\theta}\eta\big((c,d),\theta\big) \in \text{range}\big(M(\xi^*,\theta)\big)$ for all $(c, d) \in \bigcup_{i=1}^{k} \text{MED}_{p_i}$ and if it minimizes the function $\phi_{\text{MED}_q}(\xi,\theta)$ in (8) over the space of all approximate designs $\xi$ on $\mathcal{Z}$ with $\frac{\partial}{\partial\theta}\eta\big((c,d),\theta\big) \in \text{range}\big(M(\xi,\theta)\big)$ for all $(c, d) \in \bigcup_{i=1}^{k} \text{MED}_{p_i}$.

A central tool of optimal design theory is the equivalence theorem, which is frequently used to check the optimality of candidate designs determined numerically by meta-heuristic algorithms whose convergence is not guaranteed (see e.g. Kennedy and Eberhart [1995], Masoudi et al. [2019], Sandig [2024] among many others). Moreover, equivalence theorems can be used to reduce the infinite dimensional optimization problems arising in optimal design theory to finite dimensional ones. Due to the convexity of the $L_q$-norm, the criterion derived in (8) is convex with respect to the design $\xi$. Thus, we can derive the corresponding equivalence theorem (Theorem 3) that can be used to check the $\text{MED}_q$-optimality of a given design $\xi^*$. Note that Theorem 3 consists of two parts. The first part is derived for the

case where the information matrix $M(\xi^*, \theta)$ of the design $\xi^*$ is non-singular. However, there might be cases where the information matrix $M(\xi^*, \theta)$ of the $\mathrm{MED}_q$-optimal design $\xi^*$ is singular. For instance, this situation can arise if only $k = 1$ set of effective concentrations, denoted as $\mathrm{MED}_p$, is of interest. In Section 4, the statement of the theorem is used to check the optimality of the numerically determined designs. The proof of Theorem 3 can be found in the Appendix.

**Theorem 3.** *1. Let $\xi^*$ be an approximate design on $\mathcal{Z}$ such that the corresponding information matrix $M(\xi^*, \theta)$ is non-singular. The design $\xi^*$ is locally $\mathrm{MED}_q$-optimal if and only if the inequality*

$$\int_{\mathcal{C}(\theta)} \big(\varphi((c,d), \xi^*, \theta)\big)^{q-1} \alpha^2\big((c_0,d_0),(c,d),\xi^*,\theta\big)\, d\mu(c,d) - \phi^q_{\mathrm{MED}_q}(\xi^*,\theta) \leq 0 \tag{9}$$

$$\text{with } \alpha\big((c_0,d_0),(c,d),\xi,\theta\big) = \left(\frac{\partial}{\partial\theta}\eta\big((c,d),\theta\big)\right)^T M^{-1}(\xi,\theta)\left(\frac{\partial}{\partial\theta}\eta\big((c_0,d_0),\theta\big)\right),$$

*holds for all $(c_0, d_0) \in \mathcal{Z}$. Moreover, equality is achieved in* (9) *for all $(c, d)$ in the support of the design $\xi^*$.*

*2. Let $\xi^*$ be an approximate design on $\mathcal{Z}$ such that the corresponding information matrix $M(\xi^*, \theta)$ is singular. The design $\xi^*$ is locally $\mathrm{MED}_q$-optimal if and only if there exists a generalized inverse $G \in \mathbb{R}^{m\times m}$ of $M(\xi^*, \theta)$ such that the inequality*

$$\int_{\mathcal{C}(\theta)} \big(\varphi((c,d), \xi^*, \theta)\big)^{q-1} \alpha^2\big((c_0,d_0),(c,d),\xi^*,\theta\big)\, d\mu(c,d) - \phi^q_{\mathrm{MED}_q}(\xi^*,\theta) \leq 0 \tag{10}$$

$$\text{with } \alpha\big((c_0,d_0),(c,d),\xi,\theta\big) = \left(\frac{\partial}{\partial\theta}\eta\big((c,d),\theta\big)\right)^T G^T M(\xi,\theta) G \left(\frac{\partial}{\partial\theta}\eta\big((c_0,d_0),\theta\big)\right),$$

*holds for all $(c_0, d_0) \in \mathcal{Z}$. Moreover, equality is achieved in* (10) *for all $(c, d)$ in the support of the design $\xi^*$.*

**Remark 4.** If the $L_1$-norm is used in (8), the criterion, $\phi_{\mathrm{MED}_1}$, simplifies to an A-optimality criterion (see e.g. [Pukelsheim, 2006, p. 137]). More precisely, $\phi_{\mathrm{MED}_1}$ can be rewritten by

$$\begin{aligned}
&\int_{\mathcal{C}(\theta)} \left(\frac{\partial}{\partial\theta}\eta\big((c,d),\theta\big)\right)^T M^-(\xi,\theta) \left(\frac{\partial}{\partial\theta}\eta\big((c,d),\theta\big)\right) d\mu(c,d) \\
&= \int_{\mathcal{C}(\theta)} \mathrm{tr}\left[M^-(\xi,\theta)\left(\frac{\partial}{\partial\theta}\eta\big((c,d),\theta\big)\right)\left(\frac{\partial}{\partial\theta}\eta\big((c,d),\theta\big)\right)^T\right] d\mu(c,d) \\
&= \mathrm{tr}\Bigg[M^-(\xi,\theta) \underbrace{\int_{\mathcal{C}(\theta)} \left(\frac{\partial}{\partial\theta}\eta\big((c,d),\theta\big)\right)\left(\frac{\partial}{\partial\theta}\eta\big((c,d),\theta\big)\right)^T d\mu(c,d)}_{=:B}\Bigg] = \mathrm{tr}\left[M^-(\xi,\theta)B\right],
\end{aligned}$$

where the design $\xi$ has to satisfy $\mathrm{range}(B) \subset \mathrm{range}\big(M(\xi,\theta)\big)$.

**Remark 5.** In order to investigate the quality of a (not necessarily optimal) design $\xi$ for the purpose of identifying effective dose combinations, we consider the $\phi_{\mathrm{MED}_q}$-efficiency

$$\mathrm{eff}_{\phi_{\mathrm{MED}_q}}(\xi) := \frac{\phi_{\mathrm{MED}_q}(\xi^*,\theta)}{\phi_{\mathrm{MED}_q}(\xi,\theta)} \in (0,1]\,. \tag{11}$$

The quantity $\mathrm{eff}_{\phi_{\mathrm{MED}_q}}$ has an intuitive interpretation. Consider an experiment with $N$ observations that will be performed according to the approximate design $\xi^*$. Then $\frac{1}{N}\phi_{\mathrm{MED}_q}(\xi^*,\theta)$ can be interpreted as the approximate value of the mean predictive variance based on the experiment. Now, suppose $\xi$ is a different design, for which we could take $M$ observations instead. Both approaches will result in the same mean predictive variance if, and only if

$$\frac{\phi_{\mathrm{MED}_q}(\xi^*,\theta)}{\phi_{\mathrm{MED}_q}(\xi,\theta)} = \frac{N}{M}.$$

In other words, when $\xi$ has an efficiency of $e$, we would need to take a factor of $1/e$ more observations to obtain the same precision as under the locally $\mathrm{MED}_q$-optimal design $\xi^*$.

**Corollary 6.** *For an arbitrary design $\xi$ we have*

$$\mathrm{eff}_{\phi_{\mathrm{MED}_q}}(\xi) \geq 1 - \frac{\max\limits_{(c_0,d_0)\in\mathcal{Z}} \Psi\big(\xi,(c_0,d_0)\big)}{\phi_{\mathrm{MED}_q}(\xi)} =: \mathrm{elb}(\xi), \tag{12}$$

*where* $\Psi(\xi, (c_0, d_0))$ *corresponds to the left hand side of inequality* (9) *if the information matrix is non-singular, or respectively of inequality* (10) *in Theorem 3 for the singular case. Consequently,* $\mathrm{elb}(\xi)$ *is a lower bound for the efficiency of a design* $\xi$ *in terms of* $\Psi$*. Notably, the lower bound does not depend on the optimal design* $\xi^*$*.*

The statement in (6) follows, for example, from rearranging Equation (2.63) in [Fedorov and Leonov, 2013, p. 67], which is applicable here because the $\mathrm{MED}_q$-criterion in (8) is convex and Theorem 3 holds.

So far, the optimality criterion for identifying an effective dose combination depends on the unknown, but true parameter $\theta$. Thus, the corresponding locally optimal designs require a-priori information about the parameter vector. In dose combination studies, preliminary knowledge regarding the individual dose-response relationships and the corresponding parameters might be available from earlier dose-finding studies. Moreover, locally optimal designs can be applied as benchmarks for commonly used designs and can serve as starting points for constructing more robust designs with respect to model assumptions. A robust version of the $\phi_{\mathrm{MED}_q}$-criterion can be developed following a Pseudo-Bayesian approach (see e.g. Pronzato and Walter [1985], Chaloner [1989] and Chaloner and Larntz [1989]). More precisely, let $\pi$ be a prior distribution of the unknown parameter $\theta \in \Theta$, then a design $\xi^*$ is called Bayesian $\mathrm{MED}_q$-optimal, if it minimizes the function

$$\psi(\xi) = \int_{\Theta} \phi_{\mathrm{MED}_q}(\xi, \theta) d\pi(\theta) \tag{13}$$

for all possible designs $\xi \in \Xi$, where $\Xi$ denotes the set of all possible approximate designs on the design space $\mathcal{Z}$. For the Bayesian version of the $\mathrm{MED}_q$-criterion, analogous versions of the equivalence theorem in Theorem 3 and efficiency bounds such as in Corollary 6 can be derived.

## 4 Case study

In oncological research, test substances are examined to investigate whether they can reduce the growth of tumors. For this reason, the Tumor Growth Inhibition (TGI) [%] is often used as the primary endpoint. When $\mathrm{TV}_{t,i,j}$ denotes the tumor volume of animal $j = 1, \ldots, n_i$ in treatment group $i = 1, \ldots, d$ at day $t = 0, \ldots, T$, then $\mathrm{TGI}_{t,i,j}$ is defined as

$$\mathrm{TGI}_{t,i,j} = \left(1 - \frac{\mathrm{TV}_{t,i,j} - \mathrm{TV}_{0,i,j}}{\mathrm{med}_j(\mathrm{TV}_{t,0,j}) - \mathrm{med}_j(\mathrm{TV}_{0,0,j})}\right) \cdot 100\,\%.$$

Here, $t = 0$ refers to measurements at baseline and $i = 0$ indicates the negative control and placebo group.

Hence, a TGI value of 100% denotes a stasis with no observable tumor growth, whereas a TGI value greater than 100% signifies a reduction (regression) in the tumor volume. The following example of a mouse experiment consists of two test drugs $C$ and $D$, with 3 and 4 dose levels and corresponding negative control observations within the design region $\mathcal{Z} = [0, 20] \times [0, 7]$. In addition to these data from the monotherapies, some combinations are used and the TGI is measured. The corresponding design, denoted as original design, is displayed in Appendix Table 4.

The data set for this example is derived from 4 pooled experiments of a tumor model, maintaining important parameters such as cell line, strain, and route of administration consistent across all experiments. TGI is evaluated on day 21 and normalized against the control group in each of the experiments. By pooling these experiments, we achieve a relatively high total sample size for preclinical experiments with unbalanced sample sizes in the individual dose groups.

The corresponding data set was utilized to fit a dose combination model as defined in (1). First, an investigation was conducted to determine which dose-response model best describes the monotherapies. This was achieved by applying the multiple comparison and modeling approach from Bretz et al. [2005] exclusively to the data points representing each monotherapy. Based on the test statistic, the optimal model selected for both substances was the sigmoid Emax model. Then, all data points were used to fit a model via maximum likelihood estimation for the corresponding dose combination model, where the corresponding maximum likelihood estimate was determined with $\theta_0 = 19.05$, $\mathrm{E}_{\max,C} = 111.10$, $\mathrm{ED}_{50,C} = 5.83$, $h_C = 2.86$, $\mathrm{E}_{\max,D} = 410.82$, $\mathrm{ED}_{50,D} = 20.00$, $h_D = 0.78$, $\gamma = -0.0075$. A visualization of the model and the corresponding data points is given in Figure 2. Please note that the interaction effect $\gamma$ is slightly negative ($\gamma = -0.0075$), such that the negative interaction would be visible in the plotted response structure only for much higher dose combinations outside the design space $\mathcal{Z}$. This may be attributed to the predominance of data points from the monotherapies in the case study, with only a limited number of measurements for dose combinations $(c, d) \in \mathcal{Z}$, where the individual doses $c, d$ are not the placebo doses.

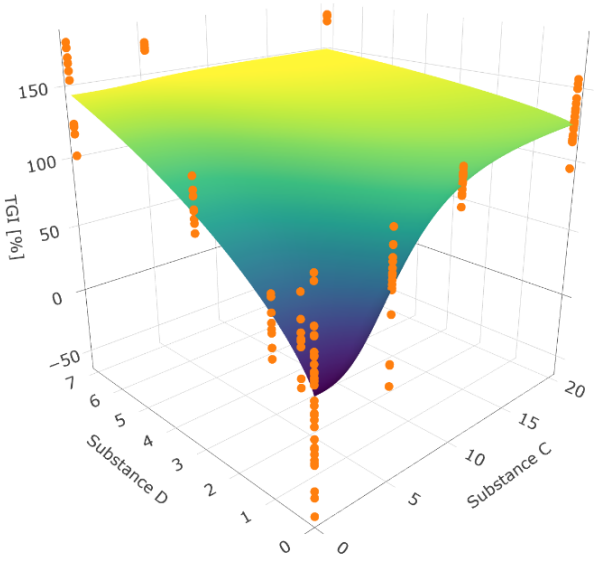



Figure 2: Drug combination surface model for the TGI in % for case study data with sigmoid Emax model for both substances. Additionally, the corresponding data points are displayed in orange.

## 5 Numerical results

In the following sections, we will compare the performance of various design approaches from both theoretical and practical perspectives. The theoretical comparison focuses on efficiencies of the different design approaches, while the practical aspect is explored through simulation studies.

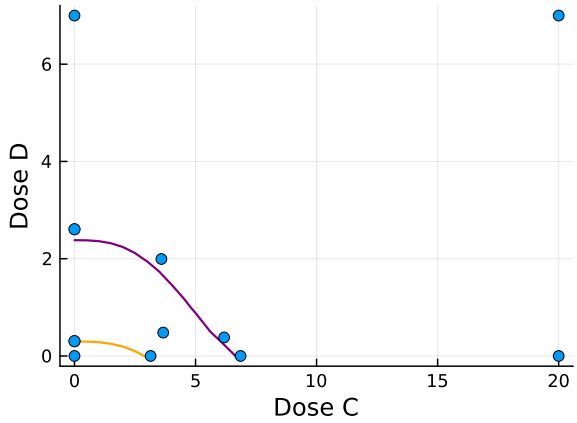
We will compare conventional designs, such as factorial and ray designs, against the derived $\mathrm{MED}_q$-optimal designs in terms of their precision in estimating effective dose combinations. For the comparison, we restrict ourselves to the $L_2$-norm resulting in locally $\mathrm{MED}_2$-optimal designs to which we refer as MED-optimal designs for the sake of brevity. The considered factorial or ray designs are visualized in the Appendix Figure 11. Additionally, we will compare these designs to the corresponding locally D-optimal designs. The initial comparison will be based on the real case study introduced in Section 4, followed by evaluations across further simulation scenarios in which we consider various dose combination response relationships with different types of interaction.

To find MED-optimal designs in practice, we use the Julia package Kirstine.jl [Sandig, 2024], where we implement (8) and (9) via a custom design criterion. For the measure $\mu$ we choose a uniform discrete distribution on a finite number of points in $\mathcal{C}(\theta)$, which we obtain by first evaluating the response function on a grid, and then interpolating the contour lines with the marching squares algorithm [Lorensen and Cline, 1987]. To verify that a numerically obtained designs is (almost) optimal, we compute the efficiency lower bound (12) obtained in Corollary 6 via the maximum of $\Psi$ over a $101 \times 101$-point grid on $\mathcal{Z}$. For all considered MED-optimal designs, the corresponding efficiency bound is at least $0.99$ (see the tabulated designs in the Appendix). The subsequent simulations and analyses were performed in R [R Core Team, 2024]. Our code and the computed designs can be found in Zenodo at [Schürmeyer and Sandig, 2025].

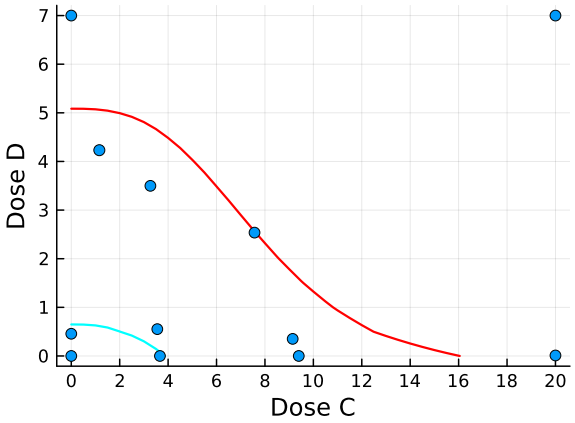
### 5.1 Efficiency-based comparison for the case study data

It is well established that the selection of dose combinations has a direct impact on the estimation precision. However, it is crucial to investigate the extent of this influence. The quality of different designs in terms of their ability to estimate the effective dose combinations can be theoretically assessed based on their associated $\phi_{\mathrm{MED}}$-efficiency in (11) (see Remark 5 for details). In this context, a higher $\phi_{\mathrm{MED}}$-efficiency indicates a superior design with respect to the corresponding MED-criterion.

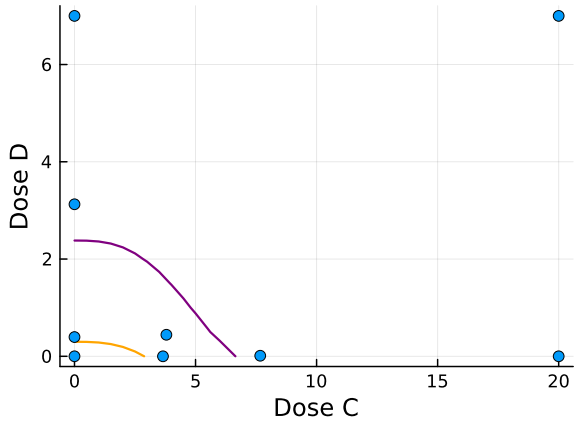
In the following, the performance of the different design approaches is compared theoretically based on their corresponding $\phi_{\mathrm{MED}}$-efficiencies for the parameter estimate of the case study data (see Section 4 for details). In this context a precise estimation of the effective dose combinations leading to 10% and 50% of the maximal effect in terms of TGI is of interest, to investigate a beginning TGI and a bisection of the TGI in terms of the maximal effect within the design



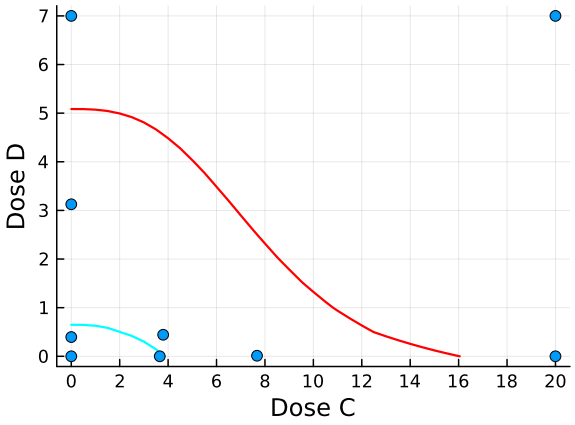

(a) Locally MED-optimal design with $\mathrm{MED}_{10}$ and $\mathrm{MED}_{50}$ sets.




(b) Locally MED-optimal design with $\mathrm{MED}_{20}$ and $\mathrm{MED}_{80}$ sets.




(c) D-optimal design with $\mathrm{MED}_{10}$ and $\mathrm{MED}_{50}$ sets.




(d) D-optimal design with $\mathrm{MED}_{20}$ and $\mathrm{MED}_{80}$ sets.

Figure 3: Visualization of the considered locally MED-optimal designs and D-optimal design for the case study, which includes their support points represented by dots, with their areas corresponding to their weights.

space. Furthermore, a second set of effective concentrations is of interest to define the magnitude of the maximal effect of the TGI, i.e. where 20% and 80% of the maximal effect are achieved.

More precisely, we consider conventional designs as the factorial $4 \times 4$ design or the ray designs and the locally MED-optimal design for the effective doses $\mathrm{MED}_{10}$ and $\mathrm{MED}_{50}$ using the parameter estimate of the case study. Furthermore, we compare the designs to the corresponding locally D-optimal design. Note that two sigmoid Emax models describe the case study data the best. In this case a factorial $3 \times 3$ is no appropriate choice, since it is not able to capture the dose combination-response effect based on this design. Especially, the monotherapies cannot be captured with only three dose combinations, when a four-parametric model such as the sigmoid Emax model is assumed. Consequently, we do not include the factorial $3 \times 3$ design in the following analysis. Additionally, the considered locally MED-optimal designs and the D-optimal design are illustrated in Figure 3, along with the corresponding contour lines of interest. The locally MED-optimal design results in dose combinations with higher weight near the contour lines of interest, while the D-optimal design aims to capture the overall response-relationship.

The efficiencies of all considered designs, evaluated according to the multivariate effective dose criterion on different sets of contour lines, are presented in Table 2. By definition, the highest efficiencies across the different criteria are achieved by their corresponding optimal designs. For a precise estimation of effective doses at the 10% and 50% levels, the D-optimal design demonstrates the next highest efficiency, followed by the locally MED-optimal design for the 20% and 80% contour levels, and then the original design. However, all three designs exhibit efficiencies that are at most 60%, indicating an insufficient performance compared to the optimal one. Furthermore, an efficiency of at

Table 2: $\phi_{\mathrm{MED}_2}$-efficiencies of the considered designs for different effect levels based on the case study.

| Design | Levels | |
|---|---|---|
| | $(10, 50)$ | $(20, 80)$ |
| Ray $4/2$ | 0.07 | 0.09 |
| Ray $4/4$ | 0.06 | 0.08 |
| Factorial $4 \times 4$ | 0.04 | 0.06 |
| Original | 0.34 | 0.47 |
| D-optimal | 0.6 | 0.8 |
| MED $(10, 50)$ | 1.00 | 0.19 |
| MED $(20, 80)$ | 0.52 | 1.00 |

most 60% for all other designs demonstrates, that theoretically more than 65% more observations are necessary for all considered designs to achieve the same precision as the locally $\mathrm{MED}(10, 50)$-optimal design. The factorial design and both ray designs show very low efficiencies, demonstrating inadequate capability to estimate those contour lines and thus determine the corresponding effective concentrations accurately.

For the $\mathrm{MED}$-criterion on other contour lines, i.e. 20% and 80%, similar results can be seen. Notably, the D-optimal design shows an efficiency of 80%, reflecting a good performance in this setting. Additionally, the original design shows a higher efficiency than in the first setting, indicating a better ability to identify these contour lines. However, its efficiency remains below 50%, indicating an insufficient performance. For the original design, twice as many observations are required to achieve theoretically the same precision as the corresponding locally $\mathrm{MED}$-optimal design. The locally $\mathrm{MED}$-optimal design for the first set of contour lines (i.e. 10% and 50%) exhibits an efficiency below 20% demonstrating poor performance. This indicates that a misspecified design performs poorly for contour lines or other settings. Note that the robustness of locally $\mathrm{MED}$-optimal designs is considered in Section 5.3. However, the commonly used factorial and ray designs perform even worse.

### 5.2 Simulation study

We perform a simulation study to compare the different design approaches for different types of interaction and different dose-combination response relationships. In particular, we investigate how precisely the contour lines of the effective dose combinations can be estimated for the different design approaches with the existing variability of the data points.

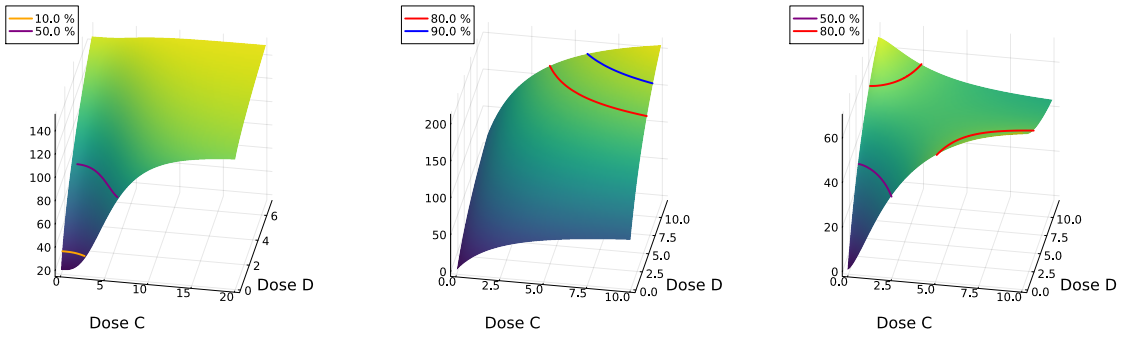
The considered parameters and designs of the scenarios are shown in Table 3. Additionally, the scenario specific response surface relationships are visualized in Figure 4. The scenarios exhibit distinct differences in their response surfaces. Scenario 1 is motivated by the case study (see Section 4) and shows no visible interaction of the two considered substances within the design space due to the slightly negative interaction effect of $\gamma = -0.0075$. Here, the contour levels of 10% and 50% can be achieved by doses from each monotherapy. Conversely, Scenario 2 demonstrates a clear positive interaction, where the contour levels of 80% and 90% can only be achieved by dose combinations that are both genuinely greater than dose combinations including placebo doses. Scenario 3 shows a clear negative interaction within the design space, resulting in a visibly non-monotonic surface. The corresponding contour lines for Scenario 3 yield distinct sets due to the structure of its response surface.

We consider different total sample sizes $N \in \{27, 36, 45, 90, 152\}$ for all scenarios. For some designs, e.g., the factorial $3 \times 3$ design, this leads to an equal number of observations at each dose combination for every choice of $N$. For all the remaining designs, an efficient rounding procedure of Pukelsheim and Rieder [1992] is used to obtain integer numbers at each dose combination. For all scenarios under consideration, $s = 1000$ simulation iterations are performed, where the different design approaches are compared in terms of their precision.

The general structure of the simulation setup remains the same for all scenarios and consists of the following steps:
At first, the initial model and parameter setting is specified for each scenario, respectively (see Table 3). In detail, this corresponds to the assumed combination model $\eta((c, d), \theta_t)$ and the parameter setting $\theta_t$. For each simulation truth $N$ new data points with normally distributed errors $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ of the given variance $\sigma^2$ are sampled at the dose combination levels of the considered design. Then, the same dose combination model assumed is fitted to the new data points in simulation step $s$, resulting in the model fit $\eta((c, d), \hat{\theta}_s)$. For this model the $k$ effective dose combinations sets $\mathrm{MED}_{p_i}$ are calculated for the scenario's levels $p_i$, $i = 1, \ldots, k$. In the next step, the Root Mean Squared Error (RMSE)

Table 3: Parameter specification of different simulation scenarios.

| | Scenario 1 | Scenario 2 | Scenario 3 |
|---|---|---|---|
| Model basis | Case study | Example Choice | Example Choice |
| Model C | sigmoid Emax | Emax | sigmoid Emax |
| Model D | sigmoid Emax | Emax | Emax |
| Design space | $[0, 20] \times [0, 7]$ | $[0, 10] \times [0, 12]$ | $[0, 10] \times [0, 12]$ |
| True $\theta_0$ | 19.05 | 0 | 0 |
| True $\theta_C$ | $(111.10, 5.83, 2.86)$ | $(80, 3)$ | $(80, 3, 1.5)$ |
| True $\theta_D$ | $(410.82, 20.00, 0.78)$ | $(120, 10)$ | $(120, 10)$ |
| True $\gamma$ | $-0.0075$ | 0.02 | $-0.02$ |
| Contour levels $(p_1, p_2)$ | $(10, 50)$ | $(80, 90)$ | $(50, 80)$ |
| Error sd | $\sigma_{\text{CS}} = 24$ | $\sigma = 30$ | $\sigma = 11$ |
| Factorial $3 \times 3$ | – | ✓ | – |
| Factorial $4 \times 4$ | ✓ | ✓ | ✓ |
| Ray 3/2 | – | ✓ | – |
| Ray 4/2 | ✓ | – | ✓ |
| Ray 3/3 | – | ✓ | – |
| Ray 4/4 | ✓ | – | ✓ |
| D-optimal | ✓ | ✓ | ✓ |
| MED-optimal | ✓ | ✓ | ✓ |
| Misspecified MED-optimal | ✓ | – | – |




(a) Scenario 1 (b) Scenario 2 (c) Scenario 3

Figure 4: Response-surface models for the three considered scenarios with considered contour lines.

is calculated for all contour lines of interest via

$$\text{RMSE} = \frac{1}{k} \sum_{i=1}^{k} \sqrt{\frac{1}{l_i} \sum_{j=1}^{l_i} \Big( p_i - \eta\big((c_j, d_j)_i, \theta_t\big) \Big)^2}$$

$$\text{with } (c_j, d_j)_i \in \text{MED}_{p_i}(\hat{\theta}_s) \quad \text{for } i = 1, \ldots, k \text{ and } j = 1, \ldots, l_i \,.$$

where $l_i$ denotes the length of the set $\text{MED}_{p_i}(\hat{\theta}_s)$ that is calculated based on a grid, $i = 1, \ldots, k$. By means of this, the precision of the estimated effects at the MED sets is measured for each simulation step in comparison to the initial "true" model.

It is important to note that due to the complex model setup, there may be computational issues when fitting the model or computing the MED sets for the new simulated data points during the simulation step. These cases were excluded from further analysis. If such cases occurred in the analysis, this will be stated.

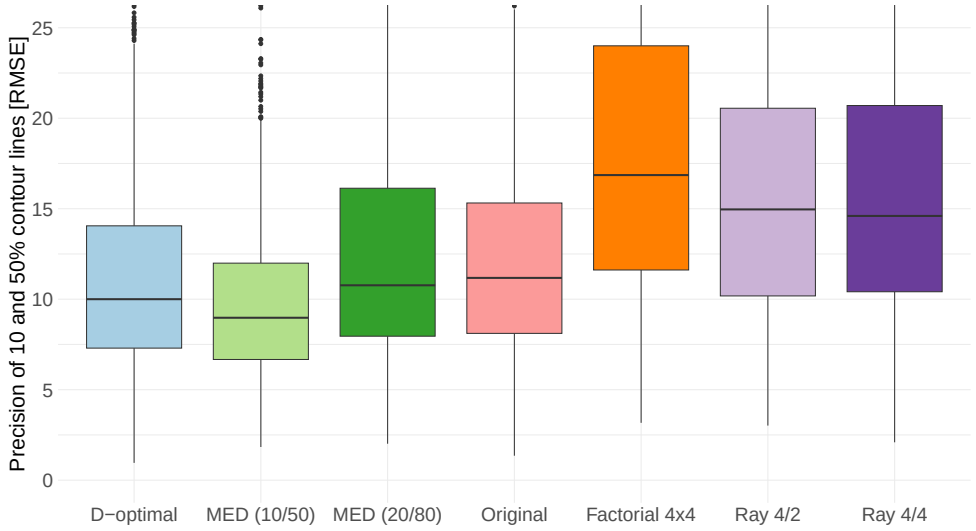



Figure 5: Extract of simulation RMSE values in Scenario 1 (case study) regarding the 10% and 50% contour levels for $N = 27$ measurements. The locally $\mathrm{MED}(10, 50)$-optimal design outperforms all other designs, especially the traditionally used factorial design and both ray designs. The locally $\mathrm{MED}(10, 50)$-optimal design shows the smallest RMSE-values and therefore the highest precision for prediction at the set of effective dose combinations. Additionally, the smallest variability can be seen by the locally $\mathrm{MED}(10, 50)$-optimal design.

Analyzing Scenario 1 shows how well the different designs perform in the practically motivated setting of the case study (see again Section 4). Since the case study serves as the source for this scenario, the initial parameter setting corresponds to the model fit of the case study. The corresponding results for the simulation study of Scenario 1 can be seen exemplarily for 27 measurements in Figure 5, where the precision of estimating the 10% and 50% contour lines is shown via the corresponding RMSE values.

Both locally $\mathrm{MED}$-optimal designs demonstrate strong performance, particularly when compared to the traditional factorial design or the ray designs. The locally $\mathrm{MED}$-optimal design, specifically tailored for estimating the corresponding 10% and 50% contour lines, exhibits the smallest median RMSE value and variability, outperforming all other design approaches. Furthermore, the locally D-optimal design shows the next best performance. The locally $\mathrm{MED}$-optimal design for the 20% and 80% contour lines also performs well and outperforms the factorial design and the ray designs. Overall, the factorial design and the ray designs are clearly inferior to the other design approaches in this context, as indicated by their higher RMSE values and variability. However, it is important to note that both ray designs achieve higher precision than the factorial $4 \times 4$ design, despite utilizing fewer dose combination levels.

A similar pattern is observed for the other sample sizes. The locally $\mathrm{MED}$-optimal design for estimating the 10% and 50% contour line outperforms the other designs for all numbers of observations. As the number of measurements increases, both the RMSE values and the variability of all designs decrease, indicating higher precision for all designs (see Appendix Figure 14). Nevertheless, the factorial or ray designs remain inferior to the other design approaches. The misspecified locally $\mathrm{MED}$-optimal design for the contour lines of 20% and 80% also outperforms the original design for higher sample sizes in terms of the median RMSE values and ranks third best in this case. Additionally, it is important to note that the factorial $4 \times 4$ design and both ray designs result in very large outliers, particularly for small sample sizes.

Furthermore, it should be noted that issues may arise when fitting the models during the simulation due to the model's complexity. Across all $N$ and all replications, one model could not be estimated for all simulations if the locally D-optimal design, the original design, or a ray design was used. Additionally, four problematic simulation steps were observed in the factorial design. Given that 1000 simulation steps are performed for each different sample size and only a few problematic steps were encountered in total, these can be considered negligible. The misspecified design showed more problematic simulation iterations with 3 to 64 issues for estimating the model. Here, the number decreased with increasing the sample size. The locally $\mathrm{MED}$-optimal design tailored to the contour lines 10% and 50% did not exhibit any problems when estimating the models during the simulation steps.

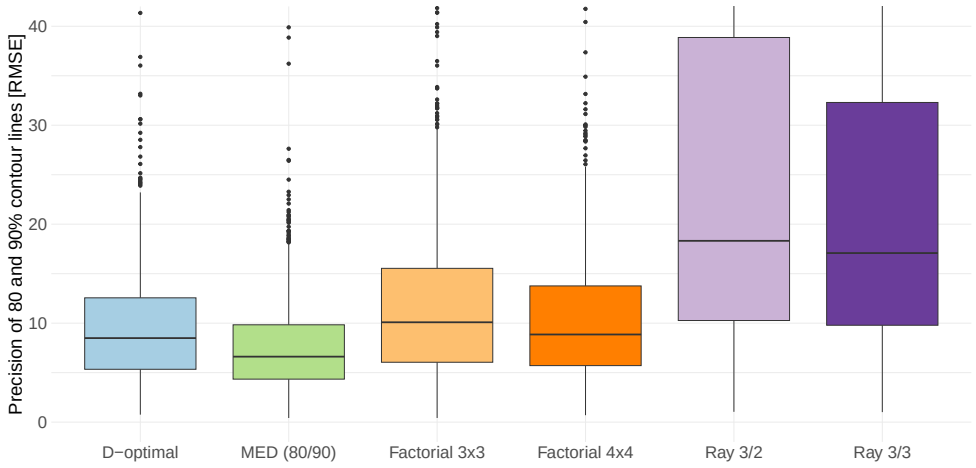



Figure 6: Extract of simulation RMSE values in Scenario 2 (two Emax models) regarding the 80% and 90% contour levels for $N = 27$ measurements. The locally MED-optimal design outperforms all other designs. It shows the smallest RMSE-values and therefore the highest precision of the set of effective dose combinations.

In the second scenario a combination of two Emax models with a clear positive interaction effect is considered (see Figure 4). Here, the precise estimation of high effective doses leading to 80% and 90% of the maximal effect is of interest. Similar to the first scenario, the different design approaches (see Table 3 and Appendix Table 5) are compared. Additionally, the locally MED-optimal and D-optimal designs are displayed in Appendix Figure 12.

The precision of the considered design approaches regarding the contour lines of interest in scenario 2 is shown in Figure 6 for the case of 27 measurements. Clearly, the locally MED-optimal design outperforms all other designs, seen by the lowest RMSE values, indicating the highest precision. The next highest precision is achieved by the D-optimal and the factorial $4 \times 4$ design. Both designs show a similar performance, with a slightly lower variability of the D-optimal design. Furthermore, the factorial $3 \times 3$ design ranks fourth in this scenario, but is considerably less effective than the locally MED-optimal design. In fact, the median RMSE value for the factorial $3 \times 3$ design exceeds the 75% quartile of the locally MED-optimal design. Both ray designs exhibit much higher RMSE values and greater variability than all other designs, indicating that they are not suitable for estimating MED curves in this scenario.

While precision generally improves with an increasing number of measurements used in the simulation, both ray designs then still perform poorly relative to other approaches (see Appendix Figure 15). Specifically, the Ray 3/2 design shows RMSE values more than three times higher than those of the locally MED-optimal design. Although the Ray 3/3 design yields smaller RMSE values than Ray 3/2, it remains clearly inferior to both factorial designs and especially to the locally MED-optimal design.

In this scenario, no issues regarding model fit were encountered for any of the design choices in the simulation study.

The third scenario differs from the first two in the structure of the combination model. On the one hand, here a sigmoid Emax model and a Emax model are considered, and on the other hand, a clearly negative interaction is visible in the plot of the response surface (see Figure 4).

For this scenario, the considered designs are displayed with their support points and corresponding weights in Appendix Table 6. Additionally, a visualization of the considered locally MED-optimal and D-optimal design can be found in Appendix Figure 13.

The corresponding comparison for estimating the contour lines 50% and 80% of the maximal effect is shown for 27 measurements in Figure 7. In this scenario, the mean RMSE values of all designs exhibit much more similar values than in the previous scenarios. However, the locally MED-optimal design still shows the smallest RMSE values. The D-optimal design is the only approach that shows a high variability in this scenario, indicating its inability to estimate the contour lines of interest precisely. The corresponding 75% quartile of the D-optimal design reaches up to twice the value of the 75% quartile of the other design approaches. The factorial $4 \times 4$ design and the ray designs perform very similarly in this setting, although they differ in the number of dose combination levels used per design. In particular, the factorial design uses 16 dose combinations, whereas the different ray designs use 8 and 10.

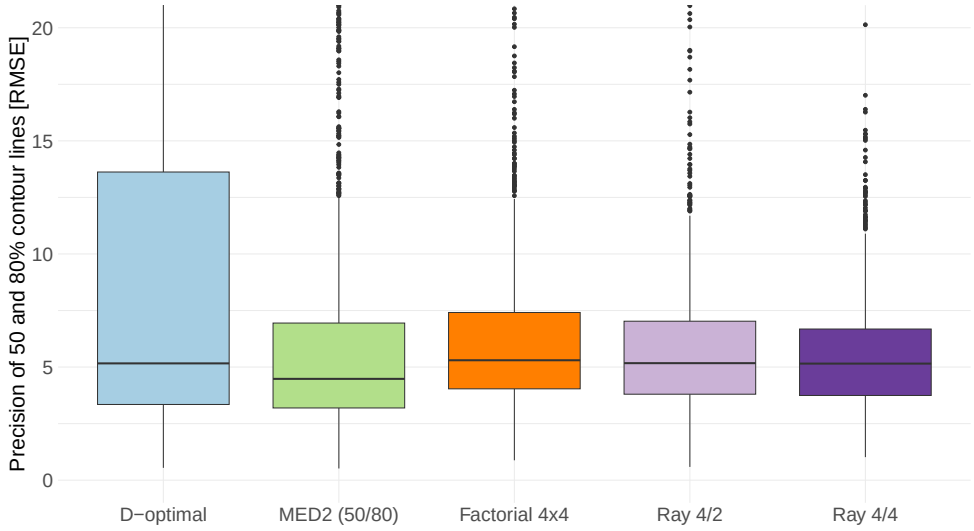



Figure 7: Extract of simulation RMSE values in Scenario 3 (sigmoid Emax and Emax model) regarding the 50% and 80% contour levels for $N = 27$ measurements. The locally MED-optimal design shows the lowest median RMSE value.

As the number of measurements $N$ increases (see Appendix Figure 16), the D-optimal design demonstrates an improvement in mean RMSE values, accompanied by a reduction in variability, as indicated by the smaller whiskers of the boxes. However, it still exhibits the highest variability among all design approaches. The differences of the locally MED-optimal design to the other design approaches becomes even more clear with increasing $N$, highlighting its superior performance. Although the factorial design and both ray designs exhibit higher precision with considerably lower RMSE values, their results remain very similar to one another, even at larger sample sizes.

Two problematic replications were identified for the factorial design out of all the simulation iterations considered. The other designs resulted in successfully estimated the models and contour lines of interest in each step.

### 5.3 Robustness analysis

In dose combination experiments, the dose-response relationships of the single substances are often well-known from former experiments, but the interaction effect is unknown. Thus, it is reasonable to account for different possible values of the interaction effect $\gamma$ when planning the drug combination experiments. To ensure robustness, Bayesian designs can be used to reflect the uncertainty about the interaction effect. In this section, the performance of the locally MED-optimal design itself and its corresponding Bayesian version defined in (13) is investigated in terms of their ability to estimate effective dose combinations most precisely.

First, we investigate a setting where the interaction effect is not precisely known but is assumed to be non-negative. We constructed a locally MED-optimal design for the simulation truth $\gamma = 0.02$ and a locally MED-optimal design for the misspecified value $\gamma = 0.05$. Moreover, the Bayesian MED-optimal design for different non-negative values of $\gamma$ was calculated using the definition in (13).

In detail, we investigate the parameter setting of the second scenario (Table 3). Here, the considered design approaches are evaluated in terms of their ability to estimate the 80% and 90% contour lines most precisely when the true parameter setting corresponds to $\theta_t$ with an interaction effect $\gamma = 0.02$.

To imitate the case of not knowing the real interaction effect before conducting a dose combination experiment, but assuming the interaction effect to be greater or equal to zero, we consider a uniform prior on $\{0, 0.01, 0.02\}$ for the interaction effect $\gamma$. The corresponding Bayesian MED-optimal design is displayed in Appendix Figure 18 and Table 7. Additionally, the resulting contour lines for the different values of $\gamma$ are shown in Appendix Figure 17. Although the response-surface is substantially different for the different choices of the interaction effect, the contour lines in this case look quite similar. Due to this fact, the Bayesian MED-optimal design has very similar support points to both locally MED-optimal designs in this case.

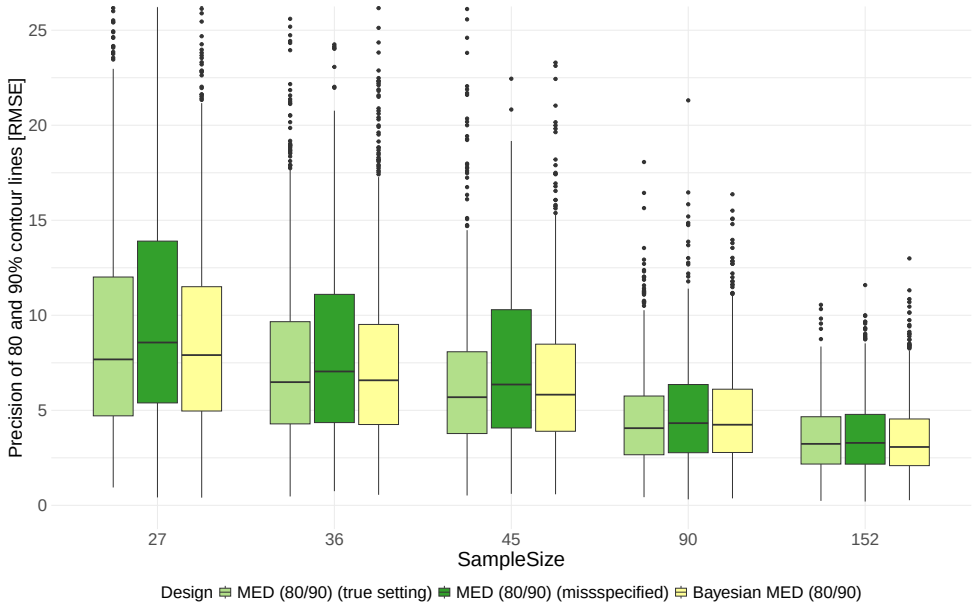



Figure 8: Extract of RMSE values for the robustness analysis of Scenario 2 (two Emax models) grouped by design and total sample size for the robustness analysis of MED designs. All designs show a similar ability to estimate the contour lines for $p_1 = 80$ and $p_2 = 90$ in this setting.

In a simulation study we compare the three different designs (Appendix Table 7) in terms of their ability to estimate the 80% and 90% contour lines precisely. Figure 8 shows the RMSE values of the 80% and 90% curves for different sample sizes. All designs seem to perform quite similar, as the median RMSE value as well as their corresponding variability seen by the corresponding box show similar values across the designs. The misspecified locally MED-optimal design, which assumed an interaction effect of $\gamma = 0.005$ instead of $\gamma = 0.02$ performs slightly worse for small sample sizes, but shows a similar performance to the other approaches in case of higher sample sizes. This can be explained by the fact, that the MEDs also look similar in this setting for different values of $\gamma$, although the corresponding surfaces differ.

Besides, there were no numerical difficulties to estimate the Bayesian $\mathrm{MED}$ or the correct specified locally $\mathrm{MED}$-optimal design, respectively. The misspecified locally $\mathrm{MED}$-optimal design instead showed a few problematic simulation iterations ranging from 12 to 46, while the number of problematic iteration steps decreased with increasing sample size.

To further explore the capabilities of a Bayesian $\mathrm{MED}$ optimal design in settings where the contour lines exhibit substantially different structures, we examined the previously considered parameter settings with alternative contour lines of interest, specifically at 10% and 30%. In this context, the contour lines distinctly vary based on different choices of the interaction effect (see Appendix Figure 19 and Appendix Figure 20).

The ability of the correctly specified locally MED-optimal design, the misspecified version with $\gamma = 0.005$, and the corresponding Bayesian design (see Appendix Table 7) to accurately estimate the true 10% and 30% contour lines is examined in another simulation study structured similarly to the previous one (see Section 5.2). Figure 9 illustrates the precision of the 10% and 30% contour lines for the different design approaches. Unlike in the previous setting, the misspecified locally MED-optimal design exhibits noticeable higher RMSE values compared to the other designs, and this trend persists even with larger sample sizes. The locally MED-optimal design demonstrates the highest precision, closely followed by its Bayesian version.

Furthermore, there were no difficulties to estimate the Bayesian $\mathrm{MED}$ or both locally $\mathrm{MED}$-optimal designs.

In a second setting, we explore the capabilities of the Bayesian MED design when the interaction effect remains uncertain and a broader range of values for $\gamma$ is considered. We choose a uniform prior on $\{-0.02, 0.01, 0, 0.01, 0.02\}$ for $\gamma$.

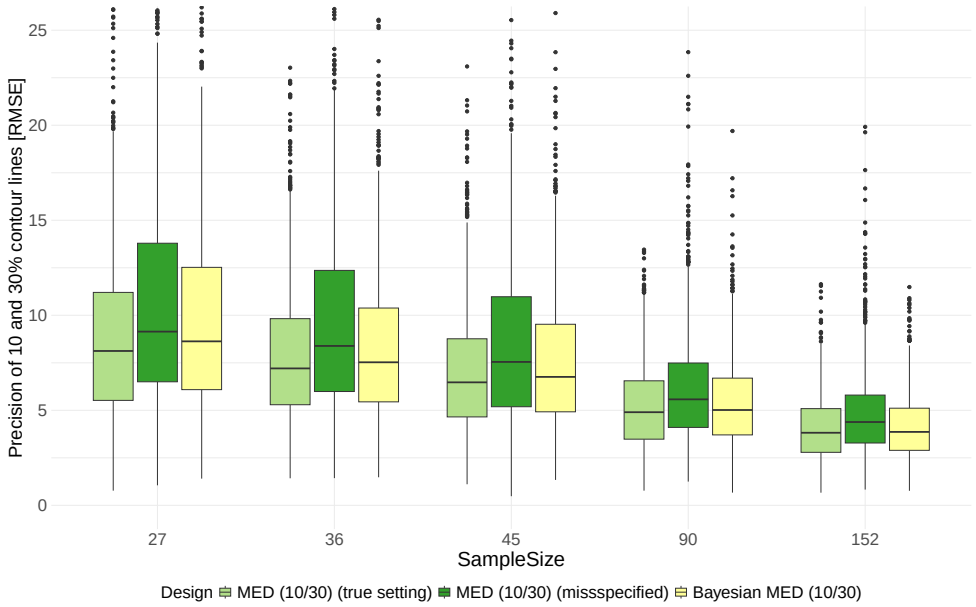



Figure 9: Extract of RMSE values for the robustness analysis of Scenario 2 (two Emax models) with simulation truth $\gamma = 0.02$ grouped by design and total sample size for the robustness analysis of MED designs. The locally MED-optimal design and the Bayesian MED design show a similar ability to estimate the contour lines for $p_1 = 10$ and $p_2 = 30$ in this setting, while the misspecified design shows higher RMSE values.

In this scenario, the true (not known before) dose combination relationship is assumed to exhibit a negative interaction effect with $\gamma_t = -0.01$. The corresponding contour lines are displayed in Appendix Figure 21 and exhibit widely different shapes. A locally MED-optimal design is calculated for the simulated truth of $\gamma_t = -0.01$ and a misspecified $\gamma = 0.02$ as well as the Bayesian MED design for the considered prior. The designs are displayed in Appendix Figure 22 and Table 8. Note that the missspecified design was originally constructed for the simulated truth in Scenario 2.

Figure 10 illustrates the precision of the 80% and 90% contour lines for the second setting in this robustness analysis. In this case, the locally MED-optimal design, which was tailored to the true parameter setting, performs best as expected and can be treated as the gold standard. The misspecified locally MED-optimal design exhibits considerably higher RMSE values, indicating a lower ability to estimate the contour lines accurately. Conversely, the Bayesian MED-optimal design, which accounts for a range of different values of the interaction effect $\gamma$, demonstrates excellent performance in this context, nearly on par with the locally optimal design.

Besides, there were no difficulties to estimate the Bayesian MED or both locally MED-optimal designs.

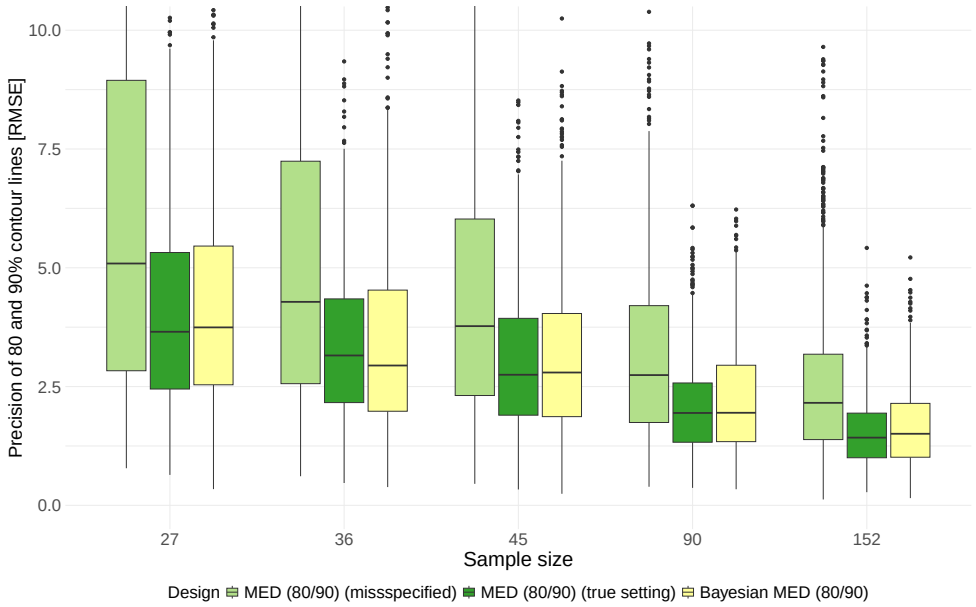



Figure 10: Extract of RMSE values for the robustness analysis with simulation truth $\gamma = -0.01$ (two Emax models) grouped by design and total sample size for the robustness analysis of MED designs. The locally MED-optimal design and the Bayesian MED design show a similar ability to estimate the contour lines for $p_1 = 80$ and $p_2 = 90$ in this setting, while the misspecified design shows clearly higher RMSE values.

## 6 Discussion and conclusion

Identifying effective dose combinations that achieve a prespecified percentage of the maximal effect has recently become more relevant in drug-combination studies. However, so far, the optimal design of drug-combination experiments is based on the established classical D-optimality criterion (see Papathanasiou et al. [2019]) or criteria related to therapeutic indices, such as the drug combination index (see Holland-Letz and Kopp-Schneider [2018]). The disadvantage of these designs is that they may result in an imprecise estimation of effective dose combinations, as given by contour lines, if nonlinear surface models are used to describe the dose combination response relationship. Therefore, we propose a novel design approach for drug-combination experiments that directly addresses the precise estimation of effective dose combinations. Here, $\mathrm{MED}_p$ denotes the set of dose combinations resulting in $p\%$ of the maximum effect, i.e. the contour line of a nonlinear surface model at $p\%$. We define a design as optimal for identifying effective dose combinations if it minimizes the $L_q$-norm of the variances of the confidence band describing the effect at the dose combinations contained in $\mathrm{MED}_p$.

For this novel criterion, we provide optimal design theory via equivalence theorems and some further analytic results. In an extensive simulation study, we compare the performance of the $\mathrm{MED}$-optimal design with commonly used designs, including factorial and ray designs. The analysis also incorporates D-optimal designs. One of the simulation scenarios is based on a real case study. Here, a data set of four pooled mouse experiments was used to investigate the dose-response relationship of two oncological substances.

Additionally, we explore further simulation scenarios that examine situations where the interaction between the two drugs is either negative or positive. These variations lead to distinct response surface relationships and different shapes of the resulting contour lines.

In all scenarios, the optimal designs for identifying effective dose combinations demonstrate strong performance, often outperforming the other designs considerably. In some scenarios, the locally D-optimal design also leads to effectively estimating the contour lines. However, it struggles if the scope is the identification of effective dose combinations for higher effect levels $p$. In this case, the contour lines of the nonlinear surface model are primarily located within the inner part of the design space. This inner part corresponds to combinations with both doses different from placebo, often not captured well by the dose combinations included in the locally D-optimal design. Furthermore, in monotonic settings, the ray designs capture the effects less effectively if higher contour lines are of interest. However, for drug combinations with adverse interaction effects and, therefore, for non-monotonic surface models, the factorial design and

both ray designs perform pretty well. Nevertheless, the optimal design for identifying dose combinations still shows the highest precision in estimating the contour lines of interest.

Given that prior knowledge of single-substance dose-response relationships often exists before conducting dose combination studies, while the interaction effects are unknown, we propose a robust version of the developed criterion using a (pseudo) Bayesian approach. More precisely, instead of assuming a specific value for the interaction effect, several possible values for the interaction effect are incorporated via a prior distribution. Performance comparisons in simulation studies reveal that the Bayesian approach enhances robustness, particularly if the prior knowledge about the interaction effect is less informative. In this case, the contour lines of the surface model vary considerably, and a potentially misspecified locally optimal design for identifying effective dose combinations is inferior to the corresponding Bayesian optimal design. However, if the prior distribution is informative, reflecting a specific direction of the type of interaction, the differences between the locally optimal and the Bayesian optimal design are minor. Here, the contour lines of interest show slight variation with different values of the interaction effect $\gamma$, and both the locally optimal designs (under slight misspecification) and the corresponding Bayesian design perform similarly. In situations where the effect is uncertain before conducting the experiment, and with varying contour lines depending on the interaction effect $\gamma$, the Bayesian optimal designs should be preferred over a potentially misspecified version of the locally optimal design for identifying effective dose combinations.

We emphasize that, even in the one-dimensional case, the confidence interval for the effective doses can be quite large in practical settings. Due to various uncontrollable factors in the laboratory, the effective doses can vary, even between two experiments with similar setups [Jiang and Kopp-Schneider, 2014]. This underscores the inherent variability of the ED itself. This issue can also arise in two-dimensional settings and is a persistent challenge. However, the developed criterion focuses on minimizing the asymptotic variance of the effective dose combinations, which also results in the minimal variability of the estimation of the effective doses.

The proposed design approach is very flexible, allowing to adapt to any nonlinear surface model, including non-monotonic regression functions for the individual drugs under consideration. We focused on identifying multivariate effective dose combinations that lead to $p\%$ of the maximal effect. However, the approach can be easily be extended to any contour line of interest. To facilitate the application of the developed optimality criterion by practitioners, code examples are provided in Zenodo at [Schürmeyer and Sandig, 2025].

The criterion is based on an $L_q$-norm where $q \in [1, \infty)$ has to be fixed in advance. We analyzed the performance of the optimal designs for various selected values of $q$, finding that, for example, $q = 2$ outperformed $q = 1$. However, we did not investigate the performance of higher values of $q$. We leave the quantification of the benefits associated with using higher values of $q$, as well as the extension of $q = \infty$ aligning with the supremum norm $L_\infty$, and an exploration of their advantages, for future research.

**Funding**
*This work has been supported by the Research Training Group "Biostatistical Methods for High-Dimensional Data in Toxicology" (RTG 2624, Project P5) funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation - Project Number 427806116).*

**Conflict of Interest**

*The authors have declared no conflict of interest.*

**Data Availability Statement**

The code that supports the findings of this study is openly available in Zenodo at [Schürmeyer and Sandig, 2025].

# Appendix

### Proof of Theorem 3

We follow the principal idea of the proof of Dette and Schorning [2016]. First, we consider the case where $M(\xi^*, \theta)$ is non-singular. Consequently, it holds that $\frac{\partial}{\partial\theta}\eta\big((c,d),\theta\big) \in \operatorname{range}\big(M(\xi^*,\theta)\big)$. Let

$$\mathcal{M} = \{M(\xi,\theta) \mid \xi \text{ is a design on } \mathcal{Z}\}$$

be the set of information matrices, where $M(\xi,\theta)$ is defined in (3). Note that $\mathcal{M}$ is a subset of all non-negative definite $(p \times p)$-matrices. For a combination $(c,d) \in \mathcal{Z}$, we introduce the function

$$\tilde{\varphi}(\,\cdot\,,(c,d)) : \mathcal{M} \to [0,\infty), \quad \tilde{\varphi}(M,(c,d)) = \frac{\partial}{\partial\theta}\eta\big((c,d),\theta\big)^T M^- \frac{\partial}{\partial\theta}\eta\big((c,d),\theta\big), \tag{14}$$

and the function

$$\tilde{\Phi}_{\mathrm{MED}_q} : \mathcal{M} \to [0,\infty), \quad \tilde{\phi}_{\mathrm{MED}_q}(M) = \left(\int_{\mathcal{C}(\theta)} \big(\tilde{\varphi}(M,(c,d))\big)^q \, d\mu((c,d))\right)^{\frac{1}{q}}. \tag{15}$$

Note that the functions in (14) and (15) are obtained by adapting the functions defined in (2) and (8). Because of the convexity of $\phi_{\mathrm{MED}_q}$ (with respect to $\xi$) and the convexity of the set $\mathcal{M}$, the function $\tilde{\phi}_{\mathrm{MED}_q}$ is also a convex function. Moreover, a design is $\xi^*$ is locally $\mathrm{MED}_q$-optimal if and only if the corresponding information matrix $M^* = M(\xi^*,\theta)$ minimizes the function $\tilde{\phi}$ on the set $\mathcal{M}$. Because of the convexity of $\tilde{\phi}_{\mathrm{MED}_q}$, the design $\xi^*$ is locally $\mathrm{MED}_q$-optimal if and only if the directional derivative of $\tilde{\phi}_{\mathrm{MED}_q}$ evaluated in $M^*$ is nonnegative for all directions $E_0 = E - M^*$, where $E \in \mathcal{M}$. Given that $\mathcal{M}$ is the convex hull of the set

$$\mathcal{D} = \big\{M(\delta_{(c_0,d_0)},\theta) \mid (c_0,d_0) \in \mathcal{Z}\big\},$$

where $\delta_{(c_0,d_0)}$ is the Dirac measure at the combination $(c_0,d_0)$, it is sufficient to prove the inequality for all $E \in \mathcal{D}$.

We now provide a step-by-step proof where we calculate the directional derivative $\partial\tilde{\phi}_{\mathrm{MED}_q}(M^*,E_0)$ of $\tilde{\phi}_{\mathrm{MED}_q}(M^*)$ in direction $E_0 \in \mathcal{M}$. If $q \geq 1$, the $L_q(\mu)$-norm (of a non-negative function) is differentiable. Combined with the chain rule, we obtain

$$\begin{aligned}\partial\tilde{\phi}_{\mathrm{MED}_q}(M^*,E_0) = &\left(\int_{\mathcal{C}(\theta)} \big(\tilde{\varphi}(M^*,(c,d))\big)^q \, d\mu((c,d))\right)^{\frac{1}{q}-1} \\ &\cdot \int_{\mathcal{C}(\theta)} \big(\tilde{\varphi}(M^*,(c,d))\big)^{q-1} \partial\tilde{\varphi}\big(M^*,E_0,(c,d)\big)\, d\mu((c,d)),\end{aligned} \tag{16}$$

where $\partial\tilde{\varphi}\big(M^*,E_0,(c,d)\big)$ denotes the directional derivative of $\tilde{\varphi}\big(M^*,(c,d)\big)$ in direction $E_0$ for a specific combination $(c,d) \in \mathcal{C} \subset \mathcal{Z}$. The next step depends on whether the information matrix is singular.

1. If the matrix $M^* = M(\xi^*,\theta)$ is non-singular, the function $\tilde{\varphi}\big(\,\cdot\,,(c,d)\big)$ is differentiable in $M^*$ and the directional derivative $\partial\tilde{\varphi}\big(M^*,E_0,(c,d)\big)$ is well-defined. To calculate $\partial\tilde{\varphi}\big(M^*,E_0,(c,d)\big)$, we follow the definition in Silvey [2013]

$$\begin{aligned}\partial\tilde{\varphi}\big(M^*,E_0,(c,d)\big) &= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\big(\tilde{\varphi}(M^*+\varepsilon E_0) - \tilde{\varphi}(M^*)\big) \\ &= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\frac{\partial}{\partial\theta}\eta\big((c,d),\theta\big)^T\left((M^*+\varepsilon E_0)^{-1} - (M^*)^{-1}\right)\frac{\partial}{\partial\theta}\eta\big((c,d),\theta\big) \\ &= g'(0)\end{aligned}$$

for $g(\varepsilon) := \frac{\partial}{\partial\theta}\eta((c,d),\theta)^T(M^*+\varepsilon E_0)^{-1}\frac{\partial}{\partial\theta}\eta((c,d),\theta)$. With the rules from matrix differential calculus [Magnus and Neudecker, 1999, pp. 148, 151, 176] we compute

$$\begin{aligned}\mathrm{d}[g] &= \frac{\partial}{\partial\theta}\eta((c,d),\theta)^T\,\mathrm{d}\big[(M^*+\varepsilon E_0)^{-1}\big]\frac{\partial}{\partial\theta}\eta((c,d),\theta) \\ &= -\frac{\partial}{\partial\theta}\eta((c,d),\theta)^T(M^*+\varepsilon E_0)^{-1}\,\mathrm{d}\big[M^*+\varepsilon E_0\big](M^*+\varepsilon E_0)^{-1}\frac{\partial}{\partial\theta}\eta((c,d),\theta) \\ &= -\frac{\partial}{\partial\theta}\eta((c,d),\theta)^T(M^*+\varepsilon E_0)^{-1}E_0(M^*+\varepsilon E_0)^{-1}\frac{\partial}{\partial\theta}\eta((c,d),\theta)\,\mathrm{d}[\varepsilon],\end{aligned}$$

hence we have

$$g'(0) = -\frac{\partial}{\partial\theta}\eta((c,d),\theta)^T (M^*)^{-1} E_0 (M^*)^{-1} \frac{\partial}{\partial\theta}\eta((c,d),\theta).$$

We continue with the original derivative,

$$\begin{aligned}
\partial\tilde{\varphi}((M^*, E_0, (c,d)) &= \frac{\partial}{\partial\theta}\eta\big((c,d),\theta\big)^T \left(-(M^*)^{-1} E_0 (M^*)^{-1}\right) \frac{\partial}{\partial\theta}\eta\big((c,d),\theta\big) \\
&= \frac{\partial}{\partial\theta}\eta\big((c,d),\theta\big)^T \left(-(M^*)^{-1}(M^* - E)(M^*)^{-1}\right) \frac{\partial}{\partial\theta}\eta\big((c,d),\theta\big) \\
&= \frac{\partial}{\partial\theta}\eta\big((c,d),\theta\big)^T \left((M^*)^{-1} E (M^*)^{-1}\right) \frac{\partial}{\partial\theta}\eta\big((c,d),\theta\big) \\
&\quad - \frac{\partial}{\partial\theta}\eta\big((c,d),\theta\big)^T (M^*)^{-1} \frac{\partial}{\partial\theta}\eta\big((c,d),\theta\big).
\end{aligned}$$

If $E_0 = E - M^*$, where $E = M(\delta_{c_0,d_0}, \theta) \in \mathcal{D}$, the partial derivative further reduces to

$$\partial\tilde{\varphi}\big((M^*, E_0, (c,d)\big) = \tilde{\alpha}^2\big((c_0,d_0),(c,d),M^*\big) - \tilde{\varphi}\big(M,(c,d)\big), \tag{17}$$

where the function $\tilde{\alpha}$ is given by

$$\tilde{\alpha}\big((c_0,d_0),(c,d),M,\theta\big) = \frac{\partial}{\partial\theta}\eta\big((c,d),\theta\big)^T M^{-1} \frac{\partial}{\partial\theta}\eta\big((c_0,d_0),\theta\big).$$

Using (16) in combination with (17), we obtain:

$$\begin{aligned}
&\partial\tilde{\phi}_{\mathrm{MED}_q}(M^*, E_0) \\
&= \left(\int_{\mathcal{C}(\theta)} \big(\tilde{\varphi}(M^*,(c,d))\big)^q d\mu((c,d))\right)^{\frac{1}{q}-1} \int_{\mathcal{C}(\theta)} \big(\tilde{\varphi}(M^*,(c,d))\big)^{q-1} \tilde{\varphi}(M^*,(c,d))\, d\mu((c,d)) \\
&\quad - \left(\int_{\mathcal{C}(\theta)} \big(\tilde{\varphi}(M^*,(c,d))\big)^q \, d\mu((c,d))\right)^{\frac{1}{q}-1} \\
&\quad \cdot \int_{\mathcal{C}(\theta)} \big(\tilde{\varphi}(M^*,(c,d))\big)^{q-1} \tilde{\alpha}^2\big(M^*,(c_0,d_0),(c,d)\big)\, d\mu((c,d)) \\
&= \left(\int_{\mathcal{C}(\theta)} \big(\tilde{\varphi}(M^*,(c,d))\big)^q \, d\mu((c,d))\right)^{\frac{1}{q}} \\
&\quad - \left(\int_{\mathcal{C}(\theta)} \big(\tilde{\varphi}(M^*,(c,d))\big)^q \, d\mu((c,d))\right)^{\frac{1}{q}} \left(\int_{\mathcal{C}(\theta)} \tilde{\varphi}(M^*,(c,d))\, d\mu((c,d))\right)^{-1} \\
&\quad \cdot \int_{\mathcal{C}(\theta)} \big(\tilde{\varphi}(M^*,(c,d),\theta)\big)^{q-1} \tilde{\alpha}^2\big((c_0,d_0),(c,d),M^*,\theta\big)\, d\mu((c,d)) \\
&= \tilde{\phi}_{\mathrm{MED}_q}(M^*) \left[1 - \big(\tilde{\phi}_{\mathrm{MED}_q}(M^*)\big)^{-q} \int_{\mathcal{C}(\theta)} \beta\big((c_0,d_0),(c,d),M^*\big)\, d\mu((c,d))\right],
\end{aligned}$$

where the function $\beta$ is given by

$$\beta\big((c_0,d_0),(c,d),M\big) = \big(\tilde{\varphi}(M,(c,d),\theta)\big)^{q-1} \tilde{\alpha}^2\big((c_0,d_0),(c,d),M,\theta\big).$$

Consequently, a design $\xi^*$ is $\mathrm{MED}_q$-optimal if and only if for the corresponding information matrix $M(\xi^*,\theta)$ the inequality

$$\int_{\mathcal{C}(\theta)} \beta\big((c_0,d_0),(c,d),M^*\big)\, d\mu(c,d) - \big(\tilde{\phi}_{\mathrm{MED}_q}(M^*)\big)^q \le 0.$$

holds for all $(c_0,d_0) \in \mathcal{Z}$ or in terms of the design if and only if the inequality

$$\int_{\mathcal{C}(\theta)} \big(\varphi((c,d),\xi^*,\theta)\big)^{q-1} \alpha^2\big((c_0,d_0),(c,d),\xi^*,\theta\big)\, d\mu(c,d) - \big(\phi_{\mathrm{MED}_q}(\xi^*,\theta)\big)^q \le 0 \tag{18}$$

holds for all $(c_0, d_0) \in \mathcal{Z}$. This proves the first part of the first assertion. It needs to be proven that equality holds for any support point $(c_0, d_0) \in \operatorname{supp}(\xi^*)$. Therefore, we assume the opposite, that there is strict inequality in (18). This leads to

$$\int_{\mathcal{Z}} \int_{\mathcal{C}(\theta)} \big(\varphi((c,d),\xi^*,\theta)\big)^{q-1} \alpha^2\big((c_0,d_0),(c,d),\xi^*,\theta\big)\, d\mu(c,d)\, d\xi((c_0,d_0)) < \big(\phi_{\mathrm{MED}_q}(\xi^*,\theta)\big)^q. \tag{19}$$

Using that

$$\begin{aligned}
&\int_{\mathcal{Z}} \alpha^2\big((c_0,d_0),(c,d),\xi^*,\theta\big)\, d\xi\big((c_0,d_0)\big) \\
&= \eta\big((c,d),\theta\big)^T M^{-1}(\xi,\theta)\eta\big((c_0,d_0),\theta\big)\eta\big((c_0,d_0),\theta\big)^T M^{-1}(\xi,\theta)\eta\big((c,d),\theta\big)\; d\xi\big((c_0,d_0)\big) \\
&= \eta\big((c,d),\theta\big)^T M^{-1}(\xi,\theta) M(\xi,\theta) M^{-1}(\xi,\theta)\eta\big((c,d),\theta\big) \\
&= \eta\big((c,d),\theta\big)^T M^{-1}(\xi,\theta)\eta\big((c,d),\theta\big),
\end{aligned}$$

the left-hand side of inequality (19) can be rewritten by:

$$\int_{\mathcal{Z}} \int_{\mathcal{C}(\theta)} \big(\varphi((c,d),\xi^*,\theta)\big)^{q-1} \alpha^2\big((c_0,d_0),(c,d),\xi^*,\theta\big)\, d\mu(c,d)\, d\xi((c_0,d_0)) = \big(\phi_{\mathrm{MED}_q}(\xi^*,\theta)\big)^q.$$

Consequently, equality must hold in (18) for $(c_0, d_0) \in \operatorname{supp}(\xi^*)$.

2. If $M^* = M(\xi^*, \theta)$ is singular and satisfies the range inclusion

$$\frac{\partial}{\partial\theta}\eta\big((c,d),\theta\big) \in \operatorname{range}\big(M(\xi^*,\theta)\big) \quad \text{for all } (c,d) \in \mathcal{C}(\theta),$$

the function $\tilde{\varphi}\big(\,\cdot\,,(c_0,d_0)\big)$ is not differentiable in $M^*$ resulting in the non-differentiability of $\tilde{\phi}_{\mathrm{MED}_q}$. However, in this case, a similar result like in the first part of the theorem is obtained using the concept of convex analysis and subgradients. More precisely, the matrix $M^*$ minimizes the function $\tilde{\phi}_{\mathrm{MED}_q}$ over all $M \in \mathcal{M}$ if and only if there exists a subgradient $N$ of $\tilde{\phi}_{\mathrm{MED}_q}$ such that

$$\operatorname{tr}[NE_0] \geq 0 \tag{20}$$

for all directions $E_0 = E - M^*$, where $E \in \mathcal{M}$. Following [Pukelsheim, 2006, p. 170], the set of all subgradients of the function $\tilde{\varphi}\big(\,\cdot\,,(c_0,d_0)\big)$ in $M^*$ is given by

$$D\tilde{\varphi}\big(M_0,(c,d)\big) = \left\{ \frac{\partial}{\partial\theta}\eta\big((c,d),\theta\big) G \frac{\partial}{\partial\theta}\eta\big((c,d),\theta\big)^T G^T \;\middle|\; G \in (M^*)^- \right\},$$

where $(M^*)^-$ is the set of all generalized inverses of $M^*$. Due to the differentiability of the $L_q$-norm the set of all subgradients of $\tilde{\phi}_{\mathrm{MED}_q}$ in $M^*$ is given by

$$\begin{aligned}
D\tilde{\phi}_{\mathrm{MED}_q}(M^*) = \Bigg\{ &\left( \int_{\mathcal{C}(\theta)} \Big(\tilde{\varphi}\big(M^*,(c,d)\big)\Big)^q d\mu\big((c,d)\big) \right)^{\frac{1}{q}-1} \\
&\cdot \int_{\mathcal{C}(\theta)} \big(\tilde{\varphi}(M^*,(c,d))\big)^{q-1} N\, d\mu\big((c,d)\big) \;\Bigg|\; N \in D\tilde{\varphi}\big(M^*,(c,d)\big) \Bigg\}.
\end{aligned} \tag{21}$$

The first assertion of the second part of the theorem follows by using (20) with (21). The proof of the second assertion is analogous to the proof given in the first part. □

**Tables**

Table 4: Considered designs with information of used dose combination levels and corresponding weights based on case study data (see Section 4). $K$ denotes the number of distinct dose combinations, and for the numerically obtained designs, elb is the lower bound on their efficiency from (12) in Corollary 6.

| Design | $K$ | $1-\text{elb}$ | Doses and Corresponding Weights | | | | | |
|---|---|---|---|---|---|---|---|---|
| Ray $4/2$ | 8 | | $(0.0, 0.0)$ $1/8$ $(20.0, 0.0)$ $1/8$ | $(0.0, 2.33)$ $1/8$ $(20.0, 7.0)$ $1/8$ | $(0.0, 4.67)$ $1/8$ | $(0.0, 7.0)$ $1/8$ | $(6.67, 0.0)$ $1/8$ | $(13.33, 0.0)$ $1/8$ |
| Ray $4/4$ | 10 | | $(0.0, 0.0)$ $1/10$ $(13.33, 0.0)$ $1/10$ | $(0.0, 2.33)$ $1/10$ $(13.33, 4.67)$ $1/10$ | $(0.0, 4.67)$ $1/10$ $(20.0, 0.0)$ $1/10$ | $(0.0, 7.0)$ $1/10$ $(20.0, 7.0)$ $1/10$ | $(6.67, 0.0)$ $1/10$ | $(6.67, 2.33)$ $1/10$ |
| Factorial $4 \times 4$ | 16 | | $(0.0, 0.0)$ $1/16$ $(6.67, 4.67)$ $1/16$ $(20.0, 0.0)$ $1/16$ | $(0.0, 2.33)$ $1/16$ $(6.67, 7.0)$ $1/16$ $(20.0, 2.33)$ $1/16$ | $(0.0, 4.67)$ $1/16$ $(13.33, 0.0)$ $1/16$ $(20.0, 4.67)$ $1/16$ | $(0.0, 7.0)$ $1/16$ $(13.33, 2.33)$ $1/16$ $(20.0, 7.0)$ $1/16$ | $(6.67, 0.0)$ $1/16$ $(13.33, 4.67)$ $1/16$ | $(6.67, 2.33)$ $1/16$ $(13.33, 7.0)$ $1/16$ |
| Original | 12 | | $(0.0, 0.0)$ $0.22$ $(0.0, 3.0)$ $0.05$ | $(5.0, 0.0)$ $0.11$ $(10.0, 3.0)$ $0.05$ | $(10.0, 0.0)$ $0.11$ $(20.0, 3.0)$ $0.05$ | $(20.0, 0.0)$ $0.16$ $(0.0, 7.0)$ $0.09$ | $(0.0, 0.3)$ $0.05$ $(5.0, 7.0)$ $0.03$ | $(0.0, 1.0)$ $0.05$ $(20.0, 7.0)$ $0.03$ |
| D-optimal | 9 | | $(0.0, 0.0)$ $0.105$ $(7.67, 0.01)$ $0.126$ | $(0.0, 0.4)$ $0.099$ $(20.0, 0.0)$ $0.125$ | $(0.0, 7.0)$ $0.125$ $(20.0, 7.0)$ $0.125$ | $(0.0, 3.13)$ $0.125$ | $(3.66, 0.0)$ $0.097$ | $(3.8, 0.44)$ $0.073$ |
| MED $(10, 50)$ | 11 | $6.0 \times 10^{-5}$ | $(0.0, 0.0)$ $0.049$ $(3.67, 0.48)$ $0.107$ | $(0.0, 0.3)$ $0.178$ $(6.18, 0.38)$ $0.109$ | $(0.0, 7.0)$ $0.012$ $(6.86, 0.0)$ $0.115$ | $(0.0, 2.61)$ $0.175$ $(20.0, 0.0)$ $0.013$ | $(3.14, 0.0)$ $0.107$ $(20.0, 7.0)$ $0.01$ | $(3.59, 1.99)$ $0.125$ |
| MED $(20, 80)$ | 12 | $3.3 \times 10^{-7}$ | $(0.0, 0.0)$ $0.011$ $(3.66, 0.0)$ $0.097$ | $(0.0, 7.0)$ $0.051$ $(7.57, 2.54)$ $0.151$ | $(0.0, 0.46)$ $0.144$ $(9.15, 0.35)$ $0.045$ | $(1.16, 4.23)$ $0.168$ $(9.4, 0.0)$ $0.057$ | $(3.27, 3.5)$ $0.005$ $(20.0, 0.01)$ $0.137$ | $(3.55, 0.55)$ $0.099$ $(20.0, 7.0)$ $0.035$ |

Table 5: Considered designs with information of used dose combination levels and corresponding weights for Scenario 2. $K$ denotes the number of distinct dose combinations, and for the numerically obtained designs, elb is the lower bound on their efficiency from (12) in Corollary 6.

| Design | $K$ | $1-\text{elb}$ | Doses and Corresponding Weights | | | | | |
|---|---|---|---|---|---|---|---|---|
| Ray $3/2$ | 6 | | $(0.0, 0.0)$ $1/6$ | $(0.0, 6.0)$ $1/6$ | $(0.0, 12.0)$ $1/6$ | $(5.0, 0.0)$ $1/6$ | $(10.0, 0.0)$ $1/6$ | $(10.0, 12.0)$ $1/6$ |
| Ray $3/3$ | 7 | | $(0.0, 0.0)$ $1/7$ $(10.0, 12.0)$ $1/7$ | $(0.0, 6.0)$ $1/7$ | $(0.0, 12.0)$ $1/7$ | $(5.0, 0.0)$ $1/7$ | $(5.0, 6.0)$ $1/7$ | $(10.0, 0.0)$ $1/7$ |
| Factorial $3 \times 3$ | 9 | | $(0.0, 0.0)$ $1/9$ $(10.0, 0.0)$ $1/9$ | $(0.0, 6.0)$ $1/9$ $(10.0, 6.0)$ $1/9$ | $(0.0, 12.0)$ $1/9$ $(10.0, 12.0)$ $1/9$ | $(5.0, 0.0)$ $1/9$ | $(5.0, 6.0)$ $1/9$ | $(5.0, 12.0)$ $1/9$ |
| Factorial $4 \times 4$ | 16 | | $(0.0, 0.0)$ $1/16$ $(3.33, 8.0)$ $1/16$ $(10.0, 0.0)$ $1/16$ | $(0.0, 4.0)$ $1/16$ $(3.33, 12.0)$ $1/16$ $(10.0, 4.0)$ $1/16$ | $(0.0, 8.0)$ $1/16$ $(6.67, 0.0)$ $1/16$ $(10.0, 8.0)$ $1/16$ | $(0.0, 12.0)$ $1/16$ $(6.67, 4.0)$ $1/16$ $(10.0, 12.0)$ $1/16$ | $(3.33, 0.0)$ $1/16$ $(6.67, 8.0)$ $1/16$ | $(3.33, 4.0)$ $1/16$ $(6.67, 12.0)$ $1/16$ |
| MED $(80, 90)$ | 7 | $5.5 \times 10^{-7}$ | $(0.0, 0.0)$ $0.004$ $(10.0, 12.0)$ $0.223$ | $(0.0, 12.0)$ $0.019$ | $(3.27, 12.0)$ $0.313$ | $(3.29, 5.92)$ $0.084$ | $(10.0, 0.0)$ $0.02$ | $(10.0, 5.88)$ $0.337$ |
| D-optimal | 13 | | $(0.0, 0.0)$ $0.164$ $(2.15, 4.22)$ $0.003$ $(10.0, 12.0)$ $0.157$ | $(0.0, 12.0)$ $0.166$ $(2.2, 4.2)$ $0.001$ | $(1.94, 12.0)$ $0.057$ $(10.0, 3.85)$ $0.003$ | $(1.94, 12.0)$ $0.085$ $(10.0, 0.0)$ $0.165$ | $(1.94, 12.0)$ $0.001$ $(10.0, 3.85)$ $0.03$ | $(2.15, 4.22)$ $0.059$ $(10.0, 3.85)$ $0.109$ |

Table 6: Considered designs with information of used dose combination levels and corresponding weights for Scenario 3. $K$ denotes the number of distinct dose combinations, and for the numerically obtained designs, elb is the lower bound on their efficiency from (12) in Corollary 6.

| Design | $K$ | $1-\text{elb}$ | Doses and Corresponding Weights | | | | | |
|---|---|---|---|---|---|---|---|---|
| Ray $4/2$ | 8 | | $(0.0, 0.0)$ $1/8$ | $(0.0, 4.0)$ $1/8$ | $(0.0, 8.0)$ $1/8$ | $(0.0, 12.0)$ $1/8$ | $(3.33, 0.0)$ $1/8$ | $(6.67, 0.0)$ $1/8$ |
| | | | $(10.0, 0.0)$ $1/8$ | $(10.0, 12.0)$ $1/8$ | | | | |
| Ray $4/4$ | 10 | | $(0.0, 0.0)$ $1/10$ | $(0.0, 4.0)$ $1/10$ | $(0.0, 8.0)$ $1/10$ | $(0.0, 12.0)$ $1/10$ | $(3.33, 0.0)$ $1/10$ | $(3.33, 4.0)$ $1/10$ |
| | | | $(6.67, 0.0)$ $1/10$ | $(6.67, 8.0)$ $1/10$ | $(10.0, 0.0)$ $1/10$ | $(10.0, 12.0)$ $1/10$ | | |
| Factorial $3 \times 3$ | 9 | | $(0.0, 0.0)$ $1/9$ | $(0.0, 6.0)$ $1/9$ | $(0.0, 12.0)$ $1/9$ | $(5.0, 0.0)$ $1/9$ | $(5.0, 6.0)$ $1/9$ | $(5.0, 12.0)$ $1/9$ |
| | | | $(10.0, 0.0)$ $1/9$ | $(10.0, 6.0)$ $1/9$ | $(10.0, 12.0)$ $1/9$ | | | |
| Factorial $4 \times 4$ | 16 | | $(0.0, 0.0)$ $1/16$ | $(0.0, 4.0)$ $1/16$ | $(0.0, 8.0)$ $1/16$ | $(0.0, 12.0)$ $1/16$ | $(3.33, 0.0)$ $1/16$ | $(3.33, 4.0)$ $1/16$ |
| | | | $(3.33, 8.0)$ $1/16$ | $(3.33, 12.0)$ $1/16$ | $(6.67, 0.0)$ $1/16$ | $(6.67, 4.0)$ $1/16$ | $(6.67, 8.0)$ $1/16$ | $(6.67, 12.0)$ $1/16$ |
| | | | $(10.0, 0.0)$ $1/16$ | $(10.0, 4.0)$ $1/16$ | $(10.0, 8.0)$ $1/16$ | $(10.0, 12.0)$ $1/16$ | | |
| MED $(50, 80)$ | 11 | 0.003 | $(0.0, 0.0)$ $0.031$ | $(0.0, 12.0)$ $0.07$ | $(0.11, 4.58)$ $0.176$ | $(0.78, 3.07)$ $0.067$ | $(1.01, 12.0)$ $0.082$ | $(1.66, 0.0)$ $0.129$ |
| | | | $(4.33, 0.0)$ $0.176$ | $(6.4, 12.0)$ $0.033$ | $(10.0, 4.12)$ $0.145$ | $(10.0, 0.0)$ $0.046$ | $(10.0, 12.0)$ $0.045$ | |
| D-optimal | 7 | | $(0.0, 0.0)$ $0.143$ | $(0.0, 3.75)$ $0.143$ | $(0.0, 12.0)$ $0.142$ | $(1.11, 0.0)$ $0.143$ | $(3.79, 0.0)$ $0.143$ | $(10.0, 0.0)$ $0.143$ |
| | | | $(10.0, 12.0)$ $0.143$ | | | | | |

Table 7: Considered designs in robustness analyses setting 1 with information of used dose combination levels and corresponding weights. Additionally, the MED $(80, 90)$ design for $\gamma_t = 0.02$ from Table 5 was included in the robustness analysis. $K$ denotes the number of distinct dose combinations, and elb is the lower bound on their efficiency from (12) in Corollary 6.

| Design | $K$ | $1-\text{elb}$ | Doses and Corresponding Weights | | | | | |
|---|---|---|---|---|---|---|---|---|
| MED $(80, 90)$ $\gamma = 0.005$ | 7 | $4.3 \times 10^{-7}$ | $(0.0, 12.0)$ $0.021$ | $(0.0, 0.0)$ $0.006$ | $(3.03, 5.58)$ $0.131$ | $(3.04, 12.0)$ $0.307$ | $(10.0, 0.0)$ $0.02$ | $(10.0, 5.67)$ $0.331$ |
| | | | $(10.0, 12.0)$ $0.184$ | | | | | |
| Bayesian MED $(80, 90)$ | 7 | $1.2 \times 10^{-6}$ | $(0.0, 0.0)$ $0.006$ | $(0.0, 12.0)$ $0.022$ | $(2.91, 5.41)$ $0.111$ | $(3.03, 12.0)$ $0.305$ | $(10.0, 0.0)$ $0.021$ | $(10.0, 5.63)$ $0.33$ |
| | | | $(10.0, 12.0)$ $0.205$ | | | | | |
| MED $(10, 30)$ $\gamma_t = 0.02$ | 9 | $6.0 \times 10^{-8}$ | $(0.0, 0.0)$ $0.13$ | $(0.0, 12.0)$ $0.187$ | $(0.0, 4.88)$ $0.061$ | $(1.54, 3.18)$ $0.27$ | $(1.6, 12.0)$ $0.027$ | $(1.99, 0.0)$ $0.039$ |
| | | | $(10.0, 3.02)$ $0.032$ | $(10.0, 0.0)$ $0.238$ | $(10.0, 12.0)$ $0.016$ | | | |
| MED $(10, 30)$ $\gamma = 0.005$ | 7 | $4.2 \times 10^{-4}$ | $(0.0, 0.0)$ $0.139$ | $(0.0, 12.0)$ $0.064$ | $(0.0, 3.98)$ $0.23$ | $(1.53, 3.04)$ $0.259$ | $(2.05, 0.0)$ $0.22$ | $(10.0, 0.0)$ $0.076$ |
| | | | $(10.0, 12.0)$ $0.012$ | | | | | |
| Bayesian MED $(10, 30)$ | 9 | 0.0031 | $(0.0, 3.82)$ $0.18$ | $(0.0, 12.0)$ $0.117$ | $(0.0, 0.0)$ $0.121$ | $(1.39, 12.0)$ $0.022$ | $(1.6, 3.26)$ $0.207$ | $(1.97, 0.0)$ $0.168$ |
| | | | $(10.0, 0.0)$ $0.146$ | $(10.0, 2.8)$ $0.024$ | $(10.0, 12.0)$ $0.015$ | | | |

Table 8: Considered designs in robustness analyses setting 2 with information of used dose combination levels and corresponding weights. Additionally, the MED $(80, 90)$ design for $\gamma = 0.02$ from Table 5 was included in the robustness analysis. $K$ denotes the number of distinct dose combinations, and $\mathrm{elb}$ is the lower bound on their efficiency from (12) in Corollary 6.

| Design | $K$ | $1 - \mathrm{elb}$ | Doses and Corresponding Weights | | | | | |
|---|---|---|---|---|---|---|---|---|
| MED $(80, 90)$ $\gamma_t = -0.01$ | 9 | $4.0 \times 10^{-6}$ | $(0.0, 0.0)$ 0.018 | $(0.0, 12.0)$ 0.107 | $(0.0, 3.88)$ 0.014 | $(1.74, 12.0)$ 0.167 | $(1.94, 0.0)$ 0.017 | $(2.34, 4.67)$ 0.233 |
| | | | $(10.0, 0.0)$ 0.068 | $(10.0, 3.49)$ 0.218 | $(10.0, 12.0)$ 0.158 | | | |
| Bayesian MED $(80, 90)$ | 9 | 0.0014 | $(0.0, 0.0)$ 0.016 | $(0.0, 5.35)$ 0.05 | $(0.0, 12.0)$ 0.076 | $(2.16, 4.22)$ 0.109 | $(2.32, 12.0)$ 0.211 | $(2.68, 0.0)$ 0.036 |
| | | | $(10.0, 0.0)$ 0.095 | $(10.0, 4.61)$ 0.227 | $(10.0, 12.0)$ 0.18 | | | |

**Figures**

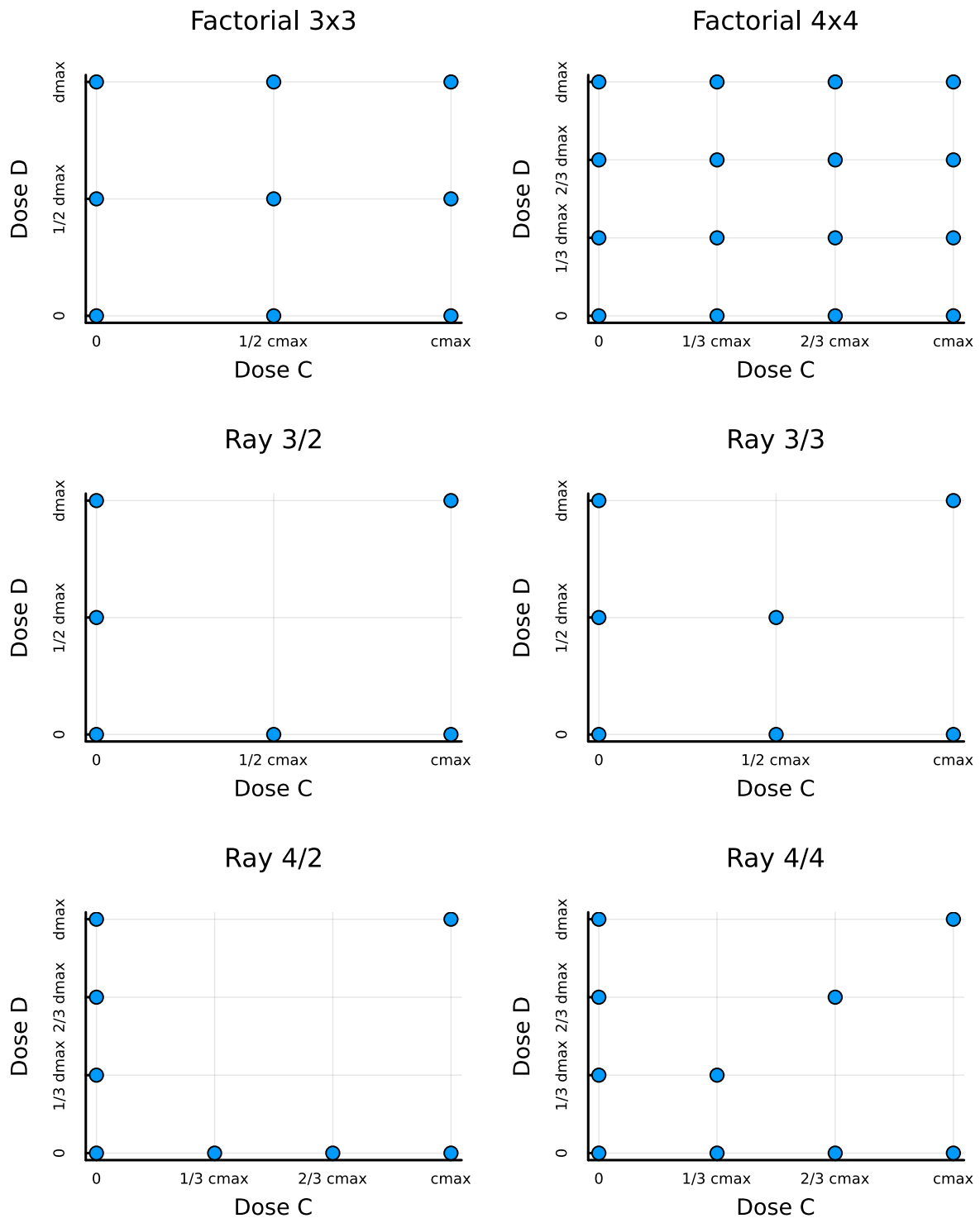
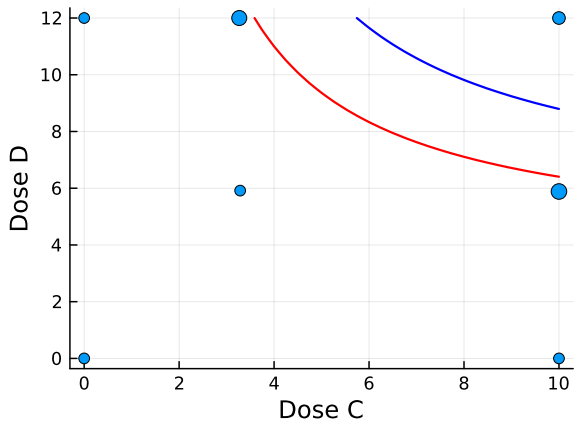
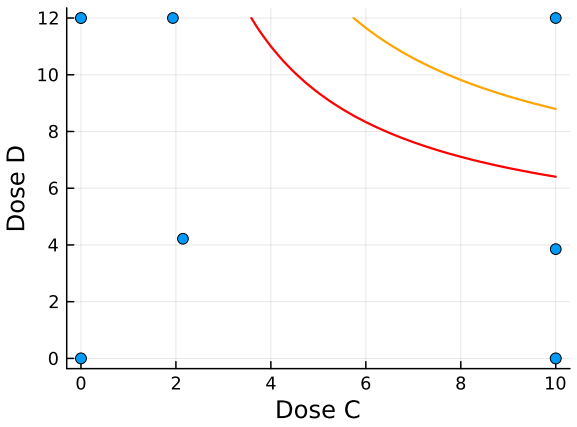



Figure 11: Showcase visualization of the considered factorial designs and ray designs. The considered number of support points can differ depending on the considered setting.

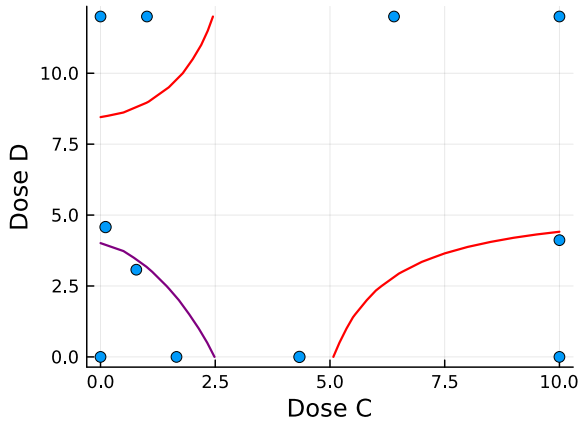



(a) Locally MED-optimal design with $\mathrm{MED}_{80}$ and $\mathrm{MED}_{90}$ sets.

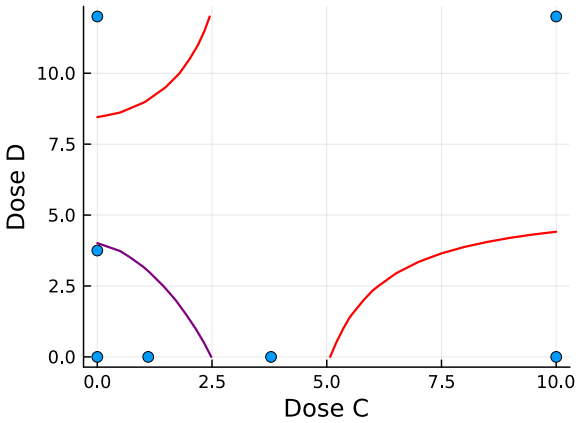



(b) D-optimal design with $\mathrm{MED}_{80}$ and $\mathrm{MED}_{90}$ sets.

Figure 12: Visualization of the considered designs and contour lines in Scenario 2, which includes their support points represented by dots, with their sizes corresponding to their weights.




(a) Locally MED-optimal design with $\mathrm{MED}_{50}$ and $\mathrm{MED}_{80}$ sets.




(b) D-optimal design with $\mathrm{MED}_{50}$ and $\mathrm{MED}_{80}$ sets.

Figure 13: Visualization of the considered designs and contour lines in Scenario 3, which includes their support points represented by dots, with their sizes corresponding to their weights.

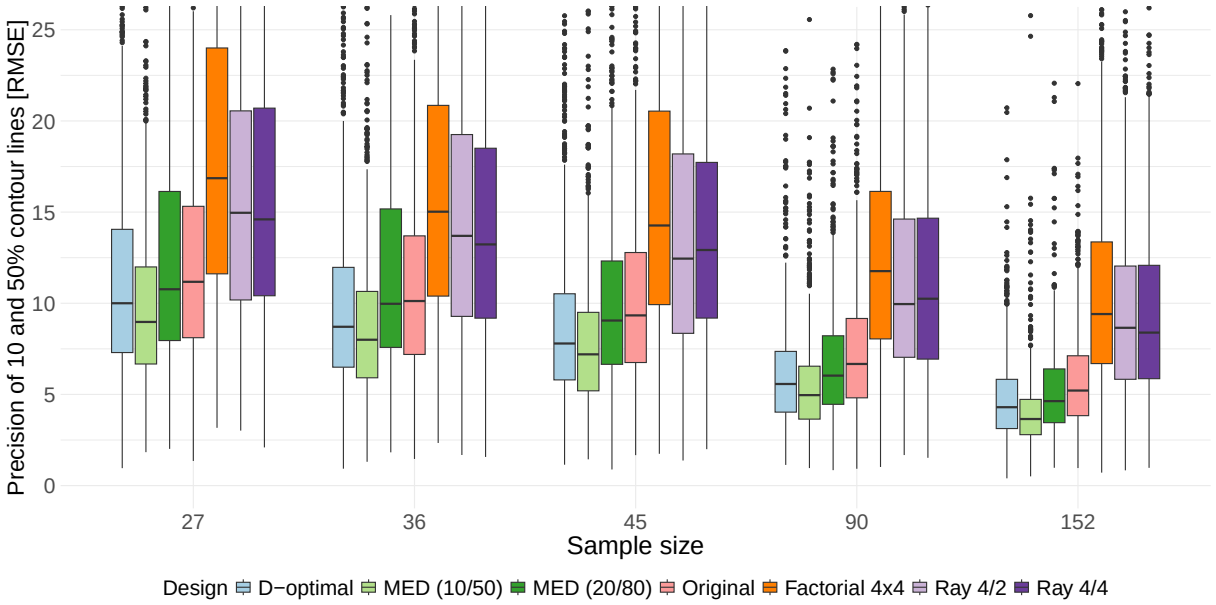



Figure 14: Extract of RMSE values grouped by design and total number of observations used in the simulation steps of Scenario 1. The locally $\mathrm{MED}(10, 50)$-optimal design shows the smallest RMSE values for all considered sample sizes compared to all other designs, especially the traditionally used factorial design and both ray designs.

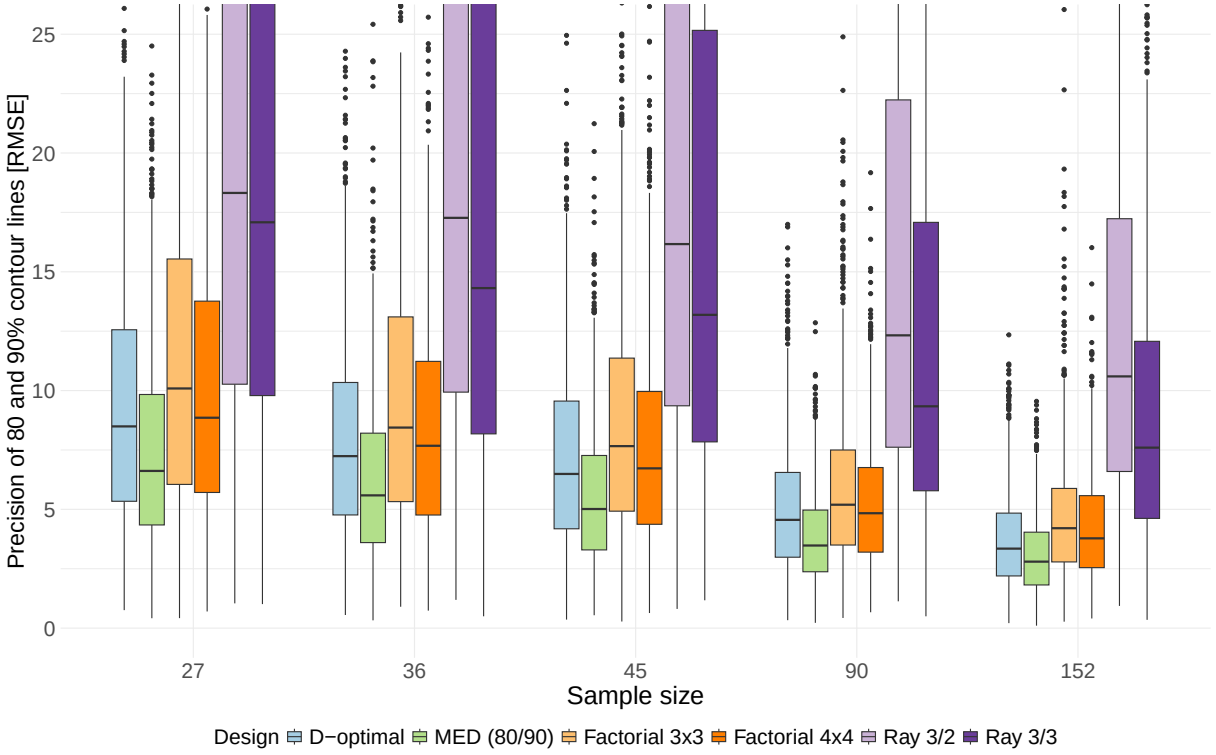



Figure 15: Extract of RMSE values grouped by design and total number of observations used in the simulation steps of Scenario 2. The locally $\mathrm{MED}$-optimal design outperforms all other designs for all numbers of measurements. In detail it shows the smallest RMSE-values and therefore the highest precision of the set of effective doses.

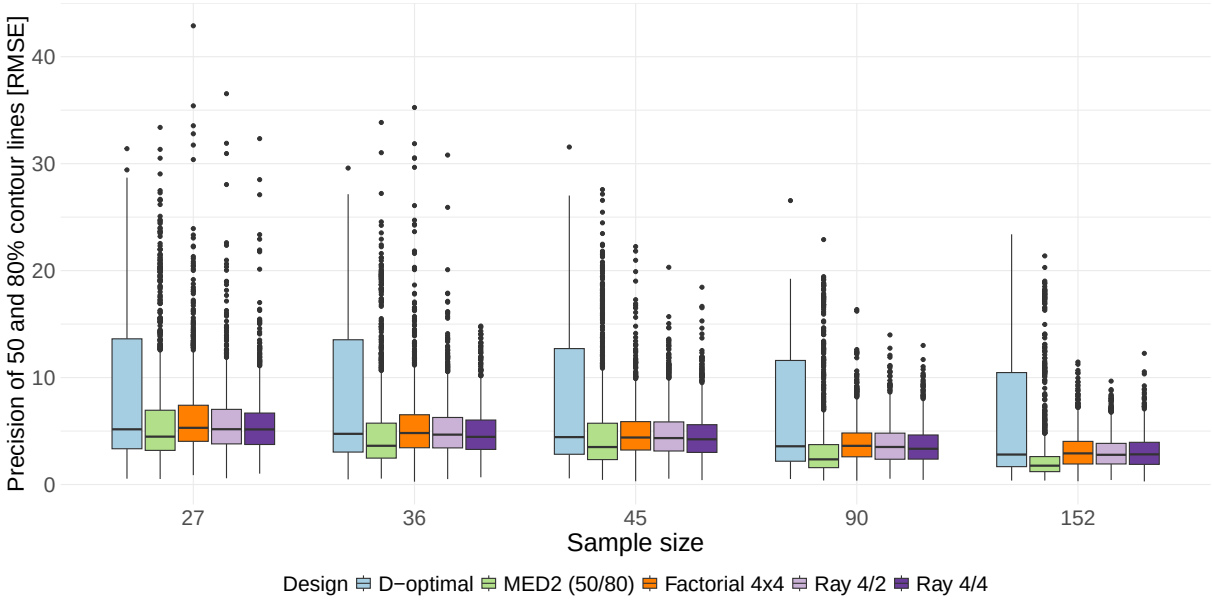



Figure 16: RMSE values grouped by design and total number of observations used in the simulation steps of Scenario 3. The locally MED-optimal design outperforms all other designs for all numbers of measurements. In detail it shows the smallest RMSE-values and therefore the highest precision of the set of effective doses.

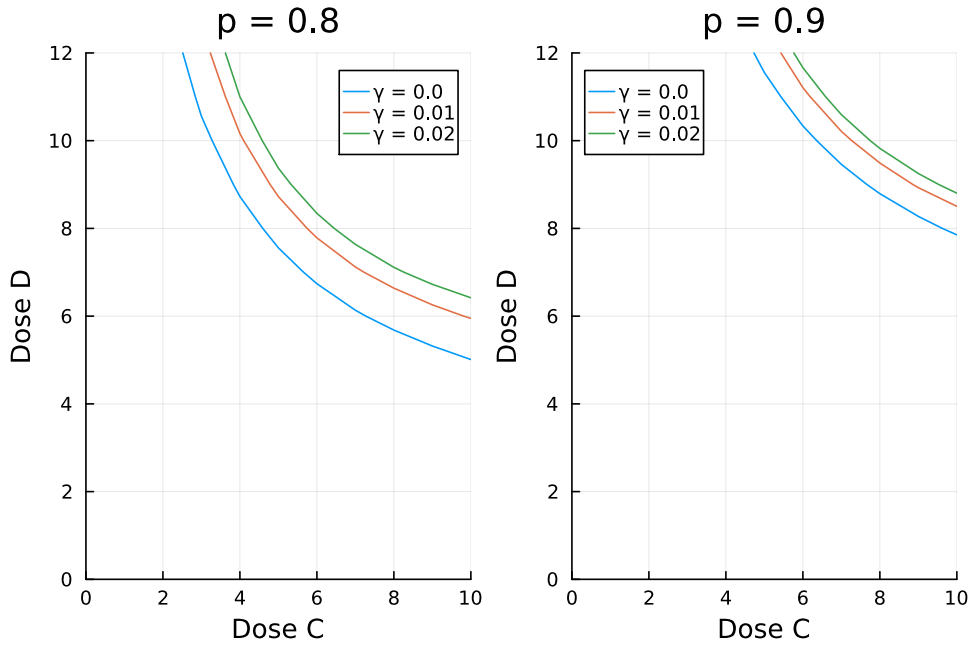
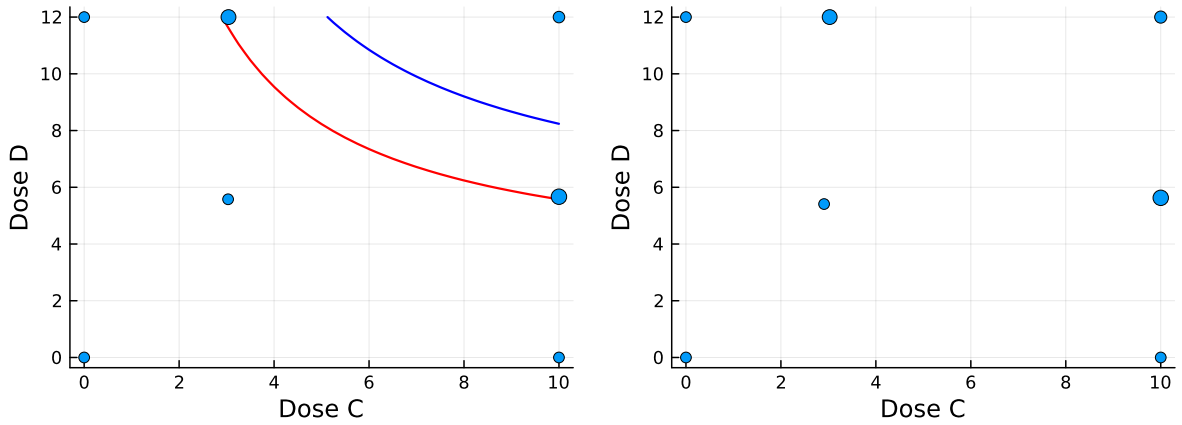



Figure 17: Visualization of $\mathrm{MED}_{80}$ and $\mathrm{MED}_{90}$ sets for robustness analysis setting 1 in case of different values of the interaction effect $\gamma$.




(a) Locally MED-optimal design with $\text{MED}_{80}$ and $\text{MED}_{90}$ sets for misspecified $\gamma = 0.005$.

(b) Bayesian MED-optimal design.

Figure 18: Visualization of the considered designs in the robustness analysis setting 1, which includes their support points represented by dots, with their sizes corresponding to their weights.

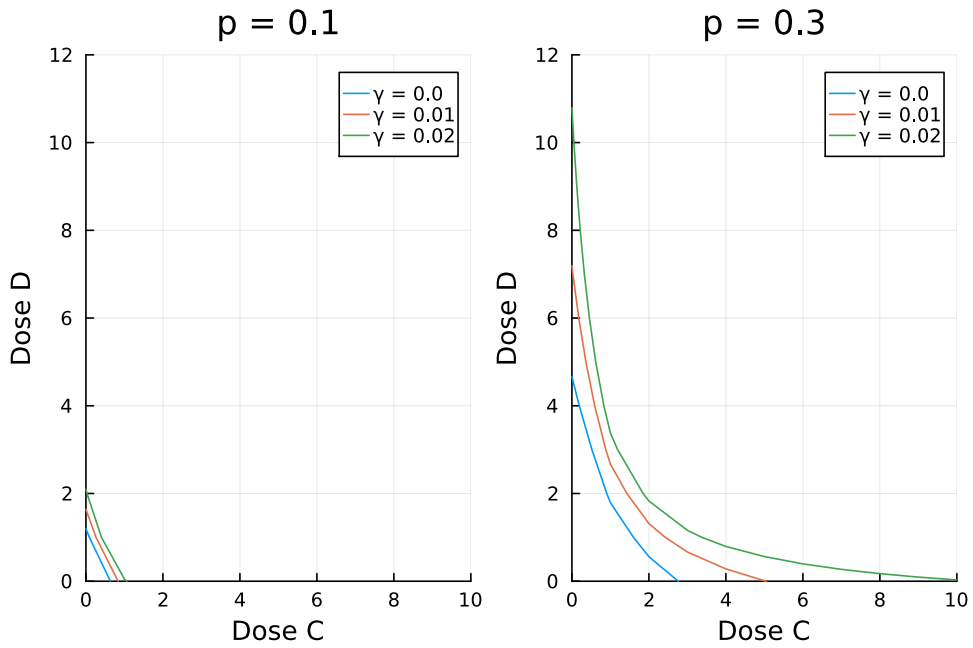
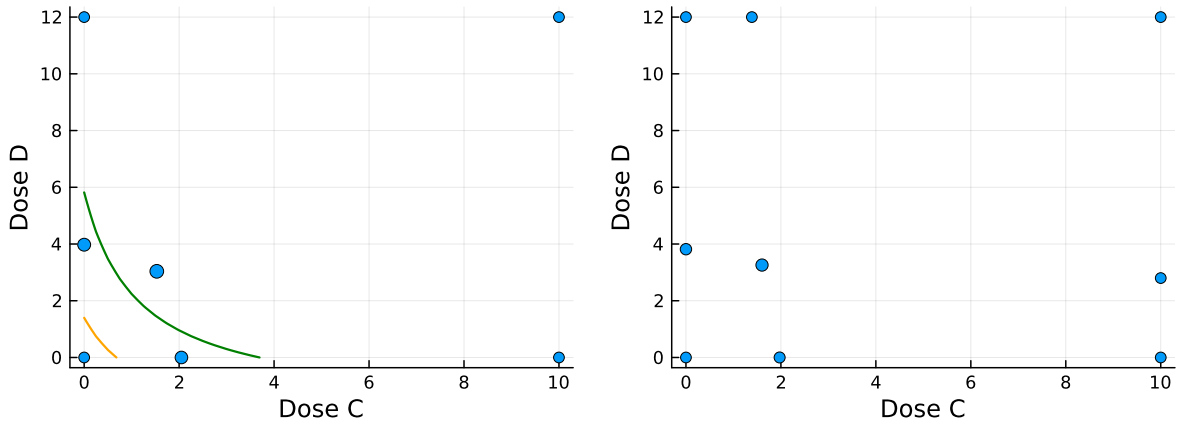



Figure 19: Visualization of $\text{MED}_{10}$ and $\text{MED}_{30}$ sets for robustness analysis setting 1 in case of different values of the interaction effect $\gamma$.




(a) Locally MED-optimal design with $\text{MED}_{10}$ and $\text{MED}_{30}$ sets for misspecified $\gamma = 0.005$.

(b) Bayesian MED design.

Figure 20: Visualization of the considered Bayesian designs in robustness analysis setting 1, which includes their support points represented by dots, with their sizes corresponding to their weights.

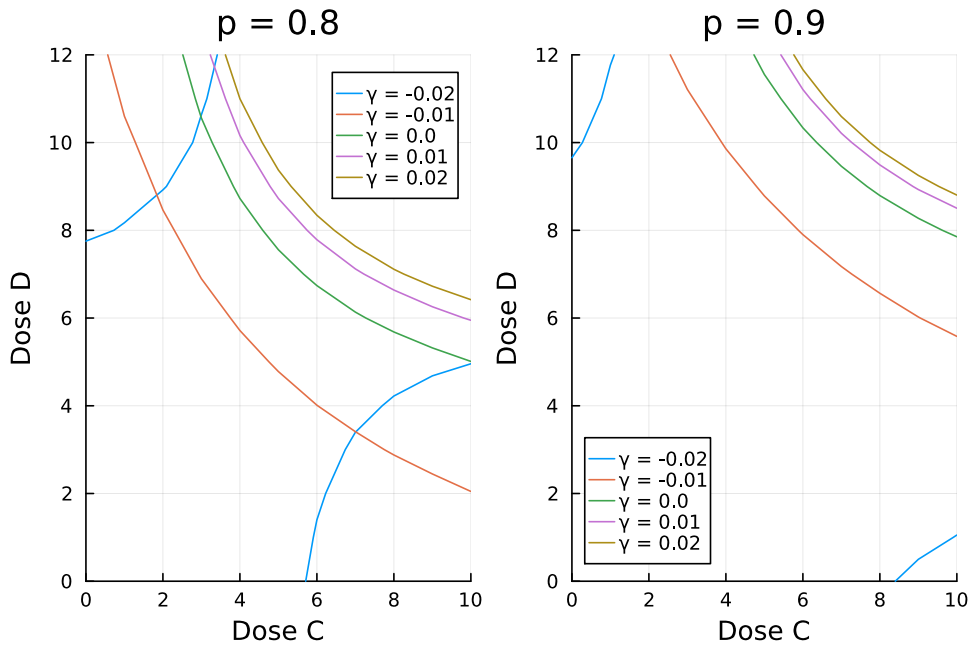
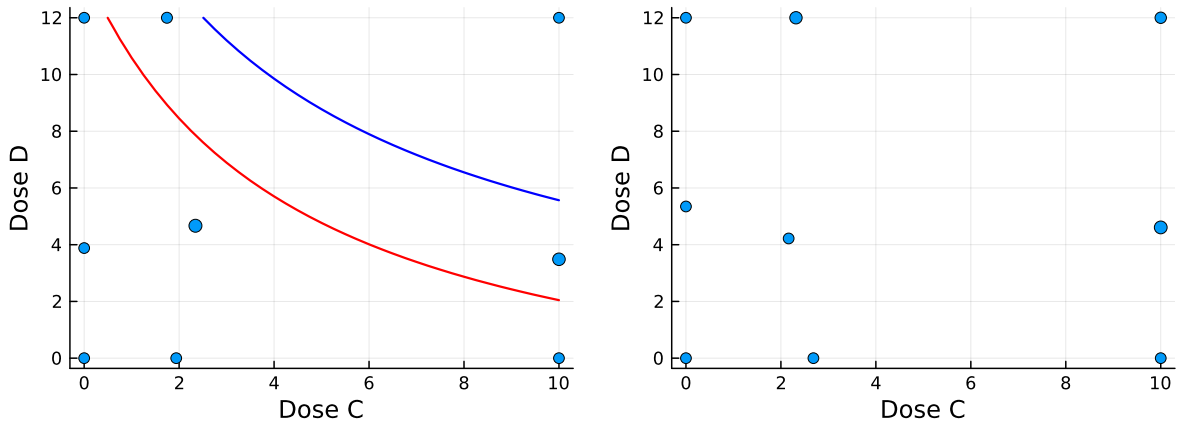



Figure 21: Visualization of $\text{MED}_{80}$ and $\text{MED}_{90}$ sets for the second setting in the robust analysis in case of different values of the interaction effect $\gamma$.




(a) MED design with $\text{MED}_{80}$ and $\text{MED}_{90}$ sets for $\gamma_t = -0.01$.

(b) Bayesian MED design.

Figure 22: Visualization of the considered Bayesian designs in robustness analysis setting 2, which includes their support points represented by dots, with their sizes corresponding to their weights.

## References


B. Almohaimeed and A. N. Donev. Experimental designs for drug combination studies. *Computational Statistics & Data Analysis*, 71:1077–1087, 2014. doi: 10.1016/j.csda.2013.01.007.

C. I. Bliss. The toxicity of poisons applied jointly. *Annals of Applied Biology*, 26(3):585–615, 1939. doi: 10.1111/j.1744-7348.1939.tb06990.x.

B. Bornkamp, J. Pinheiro, and F. Bretz. MCPMod: an R package for the design and analysis of dose-finding studies. *Journal of Statistical Software*, 29:1–23, 2009. doi: 10.18637/jss.v029.i07.

F. Bretz, J. C. Pinheiro, and M. Branson. Combining multiple comparisons and modeling techniques in dose-response studies. *Biometrics*, 61(3):738–748, 2005. doi: 10.1111/j.1541-0420.2005.00344.x.

F. Bretz, H. Dette, and J. C. Pinheiro. Practical considerations for optimal designs in clinical dose finding studies. *Statistics in Medicine*, 29(7-8):731–742, 2010. doi: 10.1002/sim.3802.

K. Chaloner. Bayesian design for estimating the turning point of a quadratic regression. *Communications in Statistics-Theory and Methods*, 18(4):1385–1400, 1989.

K. Chaloner and K. Larntz. Optimal Bayesian design applied to logistic regression experiments. *Journal of Statistical Planning and Inference*, 21(2):191–208, 1989. doi: 10.1016/0378-3758(89)90004-9.

T.-C. Chou. Preclinical versus clinical drug combination studies. *Leukemia & Lymphoma*, 49(11):2059–2080, 2008. doi: 10.1080/10428190802353591.

H. Dette and K. Schorning. Optimal designs for comparing curves. *Annals of Statistics*, 44(3):1103, 2016. doi: 10.1214/15-AOS1399.

V. V. Fedorov and S. L. Leonov. *Optimal design for nonlinear response models*. CRC Press, 2013. doi: 10.1201/b15054.

J. Foucquier and M. Guedj. Analysis of drug combinations: Current methodological landscape. *Pharmacology Research & Perspectives*, 3(3), 2015. doi: 10.1002/prp2.149.

T. Holland-Letz and A. Kopp-Schneider. Optimal experimental designs for estimating the drug combination index in toxicology. *Computational Statistics & Data Analysis*, 117:182–193, 2018. doi: 10.1016/j.csda.2017.08.006.

R. I. Jennrich. Asymptotic properties of non-linear least squares estimators. *The Annals of Mathematical Statistics*, 40 (2):633–643, 1969. doi: 10.1214/aoms/1177697731.

X. Jiang and A. Kopp-Schneider. Summarizing EC50 estimates from multiple dose-response experiments: a comparison of a meta-analysis strategy to a mixed-effects model approach. *Biometrical Journal*, 56(3):493–512, 2014. doi: 10.1002/bimj.201300123.

J. Kennedy and R. Eberhart. Particle swarm optimization. In *Proceedings of ICNN'95-international conference on neural networks*, volume 4, pages 1942–1948. IEEE, 1995. doi: 10.1109/ICNN.1995.488968.

J. Kiefer. General equivalence theory for optimum designs (approximate theory). *The Annals of Statistics*, 2(5):849–879, 1974. doi: 10.1214/aos/1176342810.

J. J. Lee, M. Kong, G. D. Ayers, and R. Lotan. Interaction index and different methods for determining drug interaction in combination therapy. *Journal of Biopharmaceutical Statistics*, 17(3):461–480, 2007. doi: 10.1080/10543400701199593.

S. Loewe. Die Mischarznei: Versuch einer Allgemeinen Pharmakologie der Arzneikombinationen. *Klinische Wochenschrift*, 6(23):1077–1085, 1927. doi: https://doi.org/10.1007/BF01890305.

W. E. Lorensen and H. E. Cline. Marching cubes: A high resolution 3D surface construction algorithm. In *Proceedings of the 14th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '87, pages 163–169. ACM, 1987. doi: 10.1145/37401.37422.

J. R. Magnus and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons Ltd, Chichester, 2nd edition, 1999. ISBN 047198633X.

E. Masoudi, H. Holling, B. P. Duarte, and W. K. Wong. A metaheuristic adaptive cubature based algorithm to find Bayesian optimal designs for nonlinear models. *Journal of Computational and Graphical Statistics*, 28(4):861–876, 2019. doi: 10.1080/10618600.2019.1601097.

F. Miller, O. Guilbaud, and H. Dette. Optimal designs for estimating the interesting part of a dose-effect curve. *Journal of Biopharmaceutical Statistics*, 17(6):1097–1115, 2007. doi: 10.1080/10543400701645140.

R. B. Mokhtari, T. S. Homayouni, N. Baluch, E. Morgatskaya, S. Kumar, B. Das, and H. Yeger. Combination therapy in combating cancer. *Oncotarget*, 8(23):38022–38043, 2017. doi: 10.18632/oncotarget.16723.

T. Papathanasiou, A. Strathe, R. V. Overgaard, T. M. Lund, and A. C. Hooker. Optimizing dose-finding studies for drug combinations based on exposure-response models. *The AAPS Journal*, 21:1–11, 2019. doi: 10.1208/s12248-019-0365-3.

L. Pronzato and E. Walter. Robust experiment design via stochastic approximation. *Mathematical Biosciences*, 75(1): 103–120, 1985. doi: 10.1016/0025-5564(85)90068-9.

F. Pukelsheim. *Optimal Design of Experiments*. SIAM, Philadelphia, 2006. doi: 10.1137/1.9780898719109.

F. Pukelsheim and S. Rieder. Efficient rounding of approximate designs. *Biometrika*, 79(4):763–770, 1992. doi: 10.2307/2337232.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2024. URL `https://www.R-project.org/`.

L. Rønneberg, A. Cremaschi, R. Hanes, J. M. Enserink, and M. Zucknick. bayesynergy: Flexible Bayesian modelling of synergistic interaction effects in in vitro drug combination experiments. *Briefings in Bioinformatics*, 22(6), 2021. doi: 10.1093/bib/bbab251.

L. Sandig. Kirstine.jl: A Julia package for Bayesian optimal design of experiments. *Journal of Open Source Software*, 9(97):6424, 2024. doi: 10.21105/joss.06424.

L. Schürmeyer and L. Sandig. Code for "Optimal designs for identifying effective doses in drug combination studies", 2025. doi: 10.5281/zenodo.15209227.

S. Silvey. *Optimal Design. An Introduction to the Theory for Parameter Estimation*, volume 1. Springer Science & Business Media, 2013. doi: 10.1007/978-94-009-5912-5.

R. Straetemans, T. O'Brien, L. Wouters, J. Van Dun, M. Janicot, L. Bijnens, T. Burzykowski, and M. Aerts. Design and analysis of drug combination experiments. *Biometrical Journal*, 47(3):299–308, 2005. doi: 10.1002/bimj.200410124.

L. Zhao, J. L. Au, and M. G. Wientjes. Comparison of methods for evaluating drug-drug interaction. *Frontiers in Bioscience (Elite Edition)*, 2:241, 2010. doi: 10.2741/e86.

Y. Zhou, A. Sloan, S. Menon, and L. Wang. Combination MCP-mod for two-drug combination dose-ranging studies. *Journal of Biopharmaceutical Statistics*, 35(2):257–270, 2024. doi: 10.1080/10543406.2024.2311254.

# 3. Article: Designs for the simultaneous inference of concentration–response curves

BMC Bioinformatics

RESEARCH

Open Access

# Designs for the simultaneous inference of concentration–response curves

Leonie Schürmeyer[1*], Kirsten Schorning[1] and Jörg Rahnenführer[1]

*Correspondence:
schuermeyer@statistik.tu-dortmund.de

[1] Department of Statistics, TU Dortmund University, Dortmund, Germany

## Abstract

**Background:** An important problem in toxicology in the context of gene expression data is the simultaneous inference of a large number of concentration–response relationships. The quality of the inference substantially depends on the choice of design of the experiments, in particular, on the set of different concentrations, at which observations are taken for the different genes under consideration. As this set has to be the same for all genes, the efficient planning of such experiments is very challenging. We address this problem by determining efficient designs for the simultaneous inference of a large number of concentration–response models. For that purpose, we both construct a *D*-optimality criterion for simultaneous inference and a *K*-means procedure which clusters the support points of the locally *D*-optimal designs of the individual models.

**Results:** We show that a planning of experiments that addresses the simultaneous inference of a large number of concentration–response relationships yields a substantially more accurate statistical analysis. In particular, we compare the performance of the constructed designs to the ones of other commonly used designs in terms of *D*-efficiencies and in terms of the quality of the resulting model fits using a real data example dealing with valproic acid. For the quality comparison we perform an extensive simulation study.

**Conclusions:** The design maximizing the *D*-optimality criterion for simultaneous inference improves the inference of the different concentration–response relationships substantially. The design based on the *K*-means procedure also performs well, whereas a log-equidistant design, which was also included in the analysis, performs poorly in terms of the quality of the simultaneous inference. Based on our findings, the *D*-optimal design for simultaneous inference should be used for upcoming analyses dealing with high-dimensional gene expression data.

**Keywords:** Optimal design, Gene expression, Nonlinear regression, High-dimensional data

## Introduction

An important problem in toxicology in the context of gene expression microarray data is the simultaneous inference of a large number of concentration–response relationships. While gene expression of each gene is observed individually within the

concentration–response experiments, the corresponding different experimental conditions, i.e. the concentrations, have to be the same for all genes. Therefore, the crucial question is how the different conditions should be allocated in order to reduce the costs of the expensive gene expression experiments while achieving a high informativeness of the resulting data.

Concerning the statistical analysis based on these experiments, the concentration–response relationships are typically modelled by nonlinear parametric curves to obtain certain parameters of interest for the individual genes, such as alert concentrations or the effect of a prespecified concentration on certain genes (see [1–3], among many others). Moreover, in each of these cases, the quality of the described inference substantially depends on the choice of the set of different conditions, at which observations are taken simultaneously for all genes. Thus, this paper is devoted to the construction of efficient sets of experimental conditions, also called optimal designs, that can be used for the simultaneous inference of a large number of nonlinear concentration–response curves.

Although determining optimal experimental designs for parametric curves has found considerable interest in the literature for the individual analysis of concentration–response relationships [4] and especially for the sigmoid Emax model [5–7], there is only a few literature on optimal design theory for multiple parametric curves. Dette and Schorning [8] constructed designs for the comparison of two different concentration–response curves, whereas Feller et al. [9] determine Determinant-optimal ($D$-optimal) designs for several concentration–response curves which share some parameters, but in both cases the resulting optimal designs are individually determined for each curve under consideration. Dror and Steinberg [10] developed robust designs based on a $K$-means algorithm for multiple experiments, but in the context of multivariate generalized linear models, which cannot directly be adapted to the situation considered in this paper. In the setting of high-dimensional microarray data, Dong et al. [11] introduced a procedure to obtain the maximally informative next experiment (MINE) for a high-dimensional linear model, in which the number of model parameters exceeds the sample size significantly. Recently, MINE was further investigated and extended to nonlinear models by Bouffier et al. [12] as well as McGee and Buzzard [13], respectively. Nevertheless, the MINE procedure is not applicable to the setting of this paper, as neither sequential approaches nor high-dimensional model parameters are included.

Instead, the design criteria developed in this paper aim for non-sequential designs that are efficient for all parametric concentration–response curves in terms of $D$-optimality. Therefore, the resulting designs for simultaneous inference contain information both about the allocation of the concentrations and about the relative amount of observations that should be taken at each concentration, respectively. In particular, we construct a $D$-optimality criterion for simultaneous inference by adapting the approach of Bayesian optimality criteria introduced by Chaloner [14] in combination with $D$-efficiencies (see [15], among many others). Although the resulting criterion is similar to the ones used in the case of one concentration–response curve, where the uncertainty of the true parameter is incorporated, its target is different, as it addresses the precise simultaneous inference of a large number of different curves. In a second approach, we use the large set of locally $D$-optimal designs obtained by considering the different curves individually and construct a cluster design based on a $K$-means cluster algorithm. In contrast to [10] the

resulting algorithm is based on the approximate designs and directly applicable to the situation of concentration–response modelling.

Further, we demonstrate that the resulting optimal designs yield a substantially more accurate statistical analysis, using an application of our developed methods on a gene expression data set provided by Krug et al. [16]. In this study the expression level of 54,675 genes exposed to valproic acid (short: VPA) is measured at eight different concentrations (including placebo), and we show that using the developed optimal designs results in significantly more precise model fits of the concentration–response curves than using the eight original concentrations. Moreover, we compare the optimal designs with commonly used designs in practice, in particular with the equidistant and log-equidistant design. Thereby, we illustrate that the log-equidistant design is clearly insufficient for such analyses, while the equidistant design performs well.

The criteria developed in this paper depend on prior knowledge on the genes, in particular on the distribution of the corresponding nonlinear parameters within the concentration–response modelling. Possible application scenarios are experiments where prior knowledge is available either from previous experiments or from results on similar experiments given in the literature. In the latter setting, our approach is especially useful when the aim of the planned experiment is the reproduction of a former experiment in another laboratory.

Furthermore, a huge advantage of our approach is its flexibility. If a different substance with different characteristics is of interest, our method is directly adjustable to the new setting. Besides there is no restriction on the prior knowledge used for the developed criteria. For example, if only specific previously known gene groups are of interest this can be incorporated in the prior distribution in order to construct a design for simultaneous inference for these specific gene groups.

The remaining paper is structured as follows. First we introduce the situation under consideration and present two new methods for constructing optimal designs for simultaneous inference in the section "Methods". Then we evaluate the performance of the developed methods on a real data example (VPA data set, see [16]). While the section "Designs for simultaneous inference of VPA-data" provides a detailed explanation of the construction of optimal designs based on this data set, in the section "Comparison of the designs" the performance of these new designs is compared to the original design and to commonly used designs in practice, in particular the equidistant and log-equidistant design. Here, we both analyse the theoretical efficiencies of the different designs under consideration and the results of an extensive simulation study. Finally, some conclusions and an outlook are given.

## Methods

### Classical optimal design theory

We assume that the same parametric model can be used to fit a curve describing the concentration–response relationship of each gene. More precisely, we assume that the data of each gene can be described by a nonlinear regression model

$$Y_{ij} = \eta(x_i, \theta) + \varepsilon_{ij} \quad j = 1, \dots, r_i; \quad i = 1, \dots, n, \tag{1}$$

where $\varepsilon_{ij}$ are independent centered normally distributed random variables, that is, $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$. This means that observations are taken at $n$ concentrations $x_1, \ldots, x_n$ which vary in the design space $\mathcal{X} = [0, x_{\max}] \subset \mathbb{R}$ and $r_i$ observations $Y_{i1}, \ldots, Y_{ir_i}$ are taken at each concentration $x_j$ $(j = 1, ..., n)$. Moreover, let $N = \sum_{i=1}^{n} r_i$ denote the total sample size. In general, a regression model $\eta(\cdot, \theta)$ with a $p$-dimensional parameter vector $\theta \in \Theta \subset \mathbb{R}^p$ is used to describe the dependence between the concentration (of a toxic compound) and the response.

Following Kiefer [17], we define an approximate design $\xi$ as probability measure with mass $w_i$ at the different support points $x_i \in \mathcal{X}$, which we denote by

$$\xi = \begin{pmatrix} x_1 & \ldots & x_n \\ w_1 & \ldots & w_n \end{pmatrix}.$$

If an approximate design is given and $N$ observations can be taken, a rounding procedure [18] is applied to obtain integers $r_i$ $(i = 1, \ldots, n)$ from the not necessarily integer values quantities $w_i N$, respectively. Then, under common assumptions of regularity and the assumption $\lim_{N\to\infty} \frac{r_i}{N} = w_i$ $(i = 1, \ldots, n)$ the maximum likelihood estimator $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_p)^T$ satisfies

$$\sqrt{N}(\hat{\theta} - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2 M^{-1}(\xi, \theta))$$

as $N \to \infty$, where the symbol $\xrightarrow{\mathcal{D}}$ denotes weak convergence. The matrix $M(\xi, \theta)$ is called information matrix of the design $\xi$ and is defined by

$$M(\xi, \theta) = \int_{\mathcal{X}} \frac{\partial}{\partial\theta}\eta(x, \theta)\left(\frac{\partial}{\partial\theta}\eta(x, \theta)\right)^T d\xi(x),$$

where $\frac{\partial}{\partial\theta}\eta(x, \theta)$ is the gradient of the regression function $\eta(x, \theta)$ with respect to the parameter $\theta \in \mathbb{R}^p$. Note that the gradient $\frac{\partial}{\partial\theta}\eta(x, \theta)$ depends on the unknown but true parameter vector, if the considered model $\eta(x, \theta)$ is nonlinear. The information matrix $M(\xi, \theta)$ is a measure of the information gained if the design $\xi$ is used. Consequently, designs that result in a large information matrix $M(\xi, \theta)$ in some sense, are appropriate. In practice, several criteria are measuring the quality of a design regarding the resulting information matrix and one of the most popular ones is the $D$-optimality criterion (see [17]). To be precise, a design $\xi_\theta^*$ is called locally $D$-optimal, as it is proposed by Chernoff [19], for estimating the parameter $\theta$ when it maximizes the concave functional

$$\psi_D(\xi, \theta) = \det(M(\xi, \theta))^{1/p} \tag{2}$$

among all designs $\xi$ on the design space $\mathcal{X}$, indicating the dependence of the $D$-optimal design on the parameter $\theta$. One key advantage of working with approximate designs and concave criteria is that convex optimization theory can be applied. As a consequence, a general equivalence theorem is available to verify whether a design is locally optimal among all designs. In particular, the locally $D$-optimality of a design $\xi_\theta^*$ can be validated by checking whether the inequality

$$d(x, \xi_\theta^*, \theta) = \frac{\partial}{\partial\theta}\eta^T(x, \theta)M^{-1}(\xi_\theta^*, \theta)\frac{\partial}{\partial\theta}\eta(x, \theta) - p \leq 0 \tag{3}$$

is satisfied for all $x \in \mathcal{X}$ (see [20]).

In order to investigate the quality of a (non-$D$-optimal) design $\xi$, we consider its $D$-efficiency which is defined by

$$\text{Eff}_D(\xi, \theta) = \frac{\psi_D(\xi, \theta)}{\psi_D(\xi^*_\theta, \theta)} \tag{4}$$

(see [21], among many others), and whose value is within the interval [0, 1] by definition. The better the design $\xi$ in terms of the $D$-optimality criterion $\psi_D$ is, the greater is its $D$-efficiency. Note that the $D$-efficiency also depends on the unknown parameter $\theta$, if the model in (1) is nonlinear.

#### Optimal designs for simultaneous inference

In the situation of the paper, let $G \in \mathbb{N}$ be the number of different genes considered in the experiment. Then $G$ corresponding different concentration–response curves (with different true parameter vectors) of the form (1) have to be fitted simultaneously using the same design $\xi$. Due to the dependence of the locally $D$-optimal designs on the unknown parameters, there is no locally $D$-optimal design which is appropriate for the simultaneous estimation of the $G$ different curves. In particular, let $\theta^{(1)} \neq \theta^{(2)}$ be two parameter vectors of two different curves. Then the $D$-efficiency of the locally $D$-optimal design $\xi^*_{\theta^{(1)}}$ might be low if used for estimating the curve where $\theta^{(2)}$ is the actual true parameter vector and vice versa.

Consequently, a design is required that provides high $D$-efficiencies for all $G$ parameter vectors and we introduce an optimality criterion that addresses this need. More precisely, let $\pi$ be a discrete distribution on the set $\Theta_G$ which contains the different parameter vectors of the $G$ different considered curves. Then a design $\xi^*_{\Theta_G}$ is called $D$-optimal for the simultaneous inference (short version: simultaneous $D$-optimal) if

$$\Psi(\xi, \pi) = \sum_{\theta \in \Theta_G} \pi(\theta) \text{Eff}_D(\xi, \theta), \tag{5}$$

is maximized by $\xi^*_{\Theta_G}$ among all designs $\xi$ on the design space $\mathcal{X}$. Moreover, it can be checked if a given design $\xi$ is simultaneous $D$-optimal by checking whether the inequality

$$s(x, \xi, \pi) = \sum_{\theta \in \Theta_G} \pi(\theta) \text{Eff}_D(\xi, \theta) d(x, \xi, \theta) \leq 0 \tag{6}$$

is satisfied for all $x \in \mathcal{X}$, where the function $d(x, \xi, \theta)$ is defined in (3). A Proof of this statement is given in the Additional file 1. Note that criteria that are of similar form as (5) were first introduced by Pronzato and Walter [22], Chaloner [14] as well as Chaloner and Larntz [23] and are also known as Bayesian or robust optimality criteria. These criteria are classically used if there is less knowledge about the unknown parameter value of one parametric curve which should be estimated. In this case a prior distribution, $\pi$, on the parameter space $\Theta$ is used to average the locally $D$-optimality criterion given in (2) with respect to different parameter values (see [24, 25], among others). Although the criteria are similar to the one defined in (5), we emphasize that the target of the latter mentioned is the finding of a design that results in good efficiencies for the $G$ different parametric curves. In particular, the simultaneous $D$-optimality criterion prevents being

affected by different sizes of the $G$ different information matrices by standardizing via the $D$-efficiencies.

Also, the simultaneous $D$-optimality criterion becomes more complex, the more complex the set $\Theta_G$ and the corresponding distribution $\pi$ are. This might result in numerical problems when the corresponding simultaneous $D$-optimal design has to be calculated for a great number $G$ of different parametric curves. One approach is the reduction of the support of the distribution $\pi$ to a smaller set $\tilde{\Theta} \subset \Theta_G$ that represents the complete set $\Theta_G$ appropriately.

Another approach for the construction of an appropriate design in the situation of simultaneous inference is motivated by a $K$-means cluster algorithm originally proposed by Hartigan and Wong [26]. More precisely, denote the support of a design $\xi$ by supp $(\xi)$ and let supp $(\xi_g^*)$ be the support of the locally $D$-optimal design for estimating the parameter $\theta^g$, $\theta^g \in \Theta_G$. Moreover, denote the intersection of all supports by $C_0 = \cap_{g=1}^G$ supp $(\xi_g^*)$ and the union of all supports by $C = \cup_{g=1}^G$ supp $(\xi_g^*)$. Fixing the number of different experimental conditions to $L \in \mathbb{N}^{\geq p}$, the $K$-means design with $L$ different experimental conditions is determined in four consecutive steps:

1. Determine $\tilde{L} \leq L$ different elements $c_1, \ldots, c_{\tilde{L}}$ in $C_0$ and set $K = L - \tilde{L}$.
2. Divide the set $C \setminus C_0$ into $K$ disjoint sets $C_1, \ldots, C_K$ that satisfy $\cup_{k=1}^K C_k = C \setminus C_0$, using the $K$-means algorithm with Euclidean distance. Moreover, calculate the center of the set $C_k$ by

$$\bar{c}_k = \frac{1}{\#C_k} \sum_{c_k \in C_k} c_k,$$

   for $k = 1, \ldots, K$, respectively, where $\#A$ denotes the number of different elements in a discrete set $A$.
3. Repeat the second step $J$ times. In the $j$-th step, sort the resulting cluster centers and denote them by $\bar{c}_{j\,(1)} < \cdots < \bar{c}_{j\,(K)}$. Calculate the mean of the $k$-th ordered center by

$$\tilde{c}_k = \frac{1}{J} \sum_{i=1}^{J} \bar{c}_{i\,(k)}$$

   for all $k = 1, \ldots, K$.
4. The $K$-means design with $L$ different experimental conditions is given by the probability measure with equal masses $\frac{1}{L}$ at the different experimental conditions $c_1, \ldots, c_{\tilde{L}}$ (see 1. step) and $\tilde{c}_1, \ldots, \tilde{c}_K$ (see 3. step). It is denoted by

$$\xi_L = \begin{pmatrix} c_1 & \cdots & c_{\tilde{L}} & \tilde{c}_1 & \cdots & \tilde{c}_K \\ \frac{1}{L} & \cdots & \frac{1}{L} & \frac{1}{L} & \cdots & \frac{1}{L} \end{pmatrix}.$$

Equal weights are used for the $K$-means design with $L$ experimental conditions $\xi_L$ for simplicity, other weights that incorporate the distribution of the different parameter values of the $G$ different parametric curves can also be used. The third step of the algorithm is included to obtain robustness with respect to the resulting clusters. Moreover, the algorithm can be further improved by using methods provided by Apon et al. [27] among many others.

Note that Dror and Steinberg [10] proposed a similar heuristic approach for constructing robust exact optimal designs based on a $K$-means algorithm for multivariate generalized linear models. While [10] use a rough grid over the parameter space and exact designs, all considered model parameters are included in our approach. Moreover, our approach is based on the corresponding locally $D$-optimal approximate designs. While the number of clusters $K$ is variable in the algorithm of [10], we fix the number of clusters in advance to a fixed $K$.

Although the $K$-means design is based on locally $D$-optimal designs, and therefore on convex design criteria, it is not possible to derive sufficient and necessary conditions of the form (6) to check the optimality of the $K$-means design.

#### The sigmoid Emax model

In the context of gene expression data, the concentration–response relationship often shows a sigmoidal course. Therefore, the regression model in (1) is frequently used with the sigmoid Emax function as regression function $\eta(\cdot, \theta)$ (see [28]). For a concentration $x \in \mathcal{X} = [0, x_{\max}]$ and a parameter vector $\theta = (\mathrm{E}_0, \mathrm{E}_{\max}, \mathrm{EC}_{50}, h) \in \mathbb{R}^4$, the sigmoid Emax function is defined by

$$\eta(x, \theta) = \mathrm{E}_0 + \frac{x^h \cdot \mathrm{E}_{\max}}{x^h + \mathrm{EC}_{50}^h}, \tag{7}$$

where the parameter $\mathrm{E}_0$ describes the effect at the placebo concentration, $x = 0$, and the parameter $\mathrm{E}_{\max}$ specifies the maximal effect associated with the considered compound. Moreover, the parameter $\mathrm{EC}_{50}$ denotes the mean effective concentration, which describes the concentration at which 50% of the maximal effect associated with the compound is attained and the hill parameter $h$ quantifies the slope of the regression function [28].

The sigmoid Emax function in (7) is nonlinear in the parameters $\mathrm{EC}_{50}$ and $h$ such that its information matrix $M(\xi, \theta)$ and the locally $D$-optimal design $\xi^*$ depend on these parameters. Due to the complexity of the sigmoid Emax function, the locally $D$-optimal designs maximizing the criterion in (2) cannot be determined analytically. However, results are available about the structure of the locally $D$-optimal design: In particular, Wang and Yang [7] proved that the locally $D$-optimal design consists of four different support points, while Li and Majumdar [6] derived that two of these support points equal the boundary points of the design space $\mathcal{X} = [0, x_{\max}]$. Using Lemma 5.1.3. stated in [29] the locally $D$-optimal design $\xi^*$ for the sigmoid Emax model is of the form

$$\xi^* = \begin{pmatrix} 0 & x_2 & x_3 & x_{\max} \\ 0.25 & 0.25 & 0.25 & 0.25 \end{pmatrix},$$

where $x_2, x_3 \in \mathcal{X}$ have to be calculated numerically in dependence on $\theta$.

### Designs for simultaneous inference of VPA-data

In the following sections, the construction of designs for simultaneous inference is illustrated by an application on a gene expression data set, called VPA-data [16]. In the section "Data", the VPA-data set is described, whereas the section "Data preprocessing and

analysis" provides a description of the initial data preprocessing and analysis steps. In the section "Construction of K-means and D-optimal design for simultaneous inference", we construct both the $K$-Means design and the simultaneous $D$-optimal design aiming for a precise inference of the concentration–response relationships of the VPA-data. All analyses were performed using the statistical software R, version 4.2.2 [30].

### Data

In a study proposed by Krug et al. [16] human embryonic stem cells were exposed to valproic acid (short: VPA). Originally the neurotoxicity was evaluated by conducting experiments with Affymetrix Human Genome U133 Plus 2.0 gene chips. The gene expression values for 54,675 probe sets were evaluated in a crude form. For the sake of clarity, the probe sets are simply considered as genes in the following. The design space $\mathcal{X}$ of potential concentrations is given by $\mathcal{X} = [0, 1000]$ and the responses were measured at the concentrations 25, 150, 350, 450, 550, 800, and 1000 $\mu$M conducted three times with different experiments. Additionally, the placebo concentration 0 was executed six times. Thus, 27 measurements at 8 different concentrations are available for each gene. The corresponding approximate design denoted as original design $\xi_{\text{orig}}$ is given by

$$\xi_{\text{orig}} = \begin{pmatrix} 0 & 25 & 150 & 350 & 450 & 550 & 800 & 1000 \\ \frac{2}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \end{pmatrix}.$$

### Data preprocessing and analysis

First, the data of the 54,675 genes are examined with respect to their biological activity. More precisely, we use the multiple contrast test of the multiple contrast procedure method (MCP-Mod) introduced by Bretz et al. [31] to detect genes whose concentration–response relationship follows a sigmoid Emax model defined by (7) to ensure a convenient model fit. Although we are reducing the data set we are not interested in a dimensionality reduction of the data set or identifying specific gene sets. Such approaches have been proposed by Azadifar et al. [32] and Rostami et al. [33], but are not considered in this paper. The MCP-Mod procedure requires specifications for the nonlinear parameter values of the used model, which are the $EC_{50}$ and $h$ parameter in case of the sigmoid Emax model. Following Duda et al. [34], we fix these parameters to $EC_{50} = 450$ and $h = 5.118$. Note that no adjustment for multiple testing was conducted. In contrast to multiple testing procedures, which aim at controlling the type I error, our goal is to develop and analyse methods on a huge number of genes. In doing so we want to identify all genes with a sensible sigmoidal model course. Thus we set the significance level to $\alpha = 0.01$. According to the MCP-Mod procedure 33,884 genes are not significant, which implies these cannot be modelled properly by the sigmoid Emax model and biological activity cannot be assumed for these genes. For 20,791 genes, the sigmoid Emax model is significant, which indicates that biological activity and a convenient sigmoid Emax model fit can be assumed.

We now concentrate on the analysis of the remaining 20,791 genes and the corresponding concentration–response relationships. In particular, we fit a regression model of the form (1) with the sigmoid Emax function (7) to the data of each gene using maximum-likelihood estimation. The estimation is provided by the function `fitMod` contained in

the R-package `DoseFinding` [35] with corresponding predefined parameter restrictions for the nonlinear parameters $\mathrm{EC}_{50} \in [0, 1500]$ and $h \in [0.05, 10]$. In the case of a fitted $\mathrm{EC}_{50} > 1000$, the concentration at which 50% of the maximal effect is attained lies outside the design space $\mathcal{X} = [0, 1000]$. From the point of view of experimental design, the constellation of an $\mathrm{EC}_{50}$-value outside the design space is not reasonable and will result in an insufficient model fit independent from the design. In particular, the model fits will nevertheless aim for an estimate of $\mathrm{EC}_{50}$ inside the design space. Thus, we restrict ourselves to the analysis of the concentration–response curves whose estimated parameter $\hat{\theta}$ is contained in the parameter space $\Theta = \mathbb{R} \times \mathbb{R} \times [0, 1000] \times [0.05, 10] \subset \mathbb{R}^4$. Furthermore, we removed data of other 85 genes, as their parameter combinations lead to numerical instabilities in the further analysis. Summarizing, the data set is reduced to $G = 15{,}233$ genes and we store the corresponding parameter estimates of the data set in a reduced parameter space $\Theta_G$, defined by

$$\Theta_G = \{\theta^g = (\hat{\mathrm{E}}_0, \hat{\mathrm{E}}_{\max}, \hat{\mathrm{EC}}_{50}, \hat{h})^T \mid \exists g \in \{1, \ldots, 15233\} : \theta^g \in \Theta\}. \tag{8}$$

We now focus on the distribution of the nonlinear parts of the parameter estimations, that is $(\mathrm{EC}_{50}, h)^T$ and analyse the distribution of the values contained in the two-dimensional set

$$\tilde{\Theta}_G = \{(\hat{\mathrm{EC}}_{50}, \hat{h})^T \mid \exists g \in \{1, \ldots, 15233\} : \theta^g \in \Theta\}. \tag{9}$$

on $\tilde{\Theta} = [0, 1000] \times [0.05, 10]$. Figure 1 illustrates the distribution of the steepness $h$ and the $\mathrm{EC}_{50}$ of every fitted sigmoid Emax model on $\tilde{\Theta}$ displaying a $(5 \times 5)$-grid classification on $\tilde{\Theta}$ of the parameter estimates contained in $\tilde{\Theta}_G$.

Due to the contrast levels an accumulation for genes with high steepness parameters independent of the $\mathrm{EC}_{50}$-value is clearly visible through the darker areas. In particular, for 31.7% of all genes the estimated steepness parameter is given by $\hat{h} = 10$, which is the upper bound for $h$ within the parameter space $\Theta$. For 56.2% of the considered genes it holds $\hat{h} \in (2, 10)$. Concentrating on the $\mathrm{EC}_{50}$-values (independent from $h$), 84.5 % of the estimates take values in the interval (200, 800). Both small values i.e. $\hat{EC}_{50} \leq 200$, and high values, i.e. $\hat{EC}_{50} > 800$, are rarely present (7.7% and 7.8%).

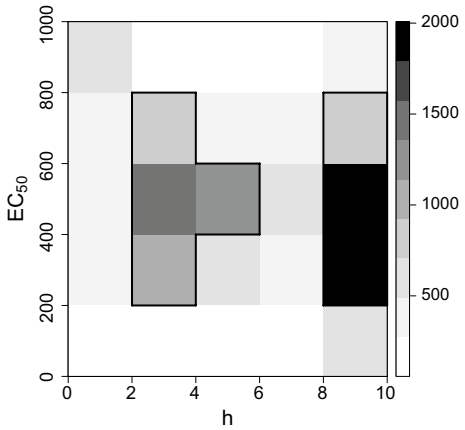



**Fig. 1** Heatmap of steepness $h$ and $\mathrm{EC}_{50}$ of $\tilde{\Theta}_G$, encircled areas correspond to frequencies higher 5%

**Table 1** Representative parameter vectors $\theta^j$ of parameter set $\Theta_7$ and corresponding weight distribution $\pi_7$ for the representated areas with frequencies higher than 5%

| **Representated Area** | $\Theta_7$ | $EC_{50}$ | ***h*** | $\pi_7(\theta^g)$ |
|---|---|---|---|---|
| $(200, 400] \times (2, 4]$ | $\theta^1$ | 298.81 | 2.53 | 0.1192 |
| $(400, 600] \times (2, 4]$ | $\theta^2$ | 575.00 | 3.77 | 0.1633 |
| $(600, 800] \times (2, 4]$ | $\theta^3$ | 758.84 | 2.72 | 0.0897 |
| $(400, 600] \times (4, 6]$ | $\theta^4$ | 469.36 | 4.00 | 0.1225 |
| $(200, 400] \times (8, 10]$ | $\theta^5$ | 310.60 | 10.00 | 0.2009 |
| $(400, 600] \times (8, 10]$ | $\theta^6$ | 501.97 | 10.00 | 0.2172 |
| $(600, 800] \times (8, 10]$ | $\theta^7$ | 747.88 | 10.00 | 0.0872 |

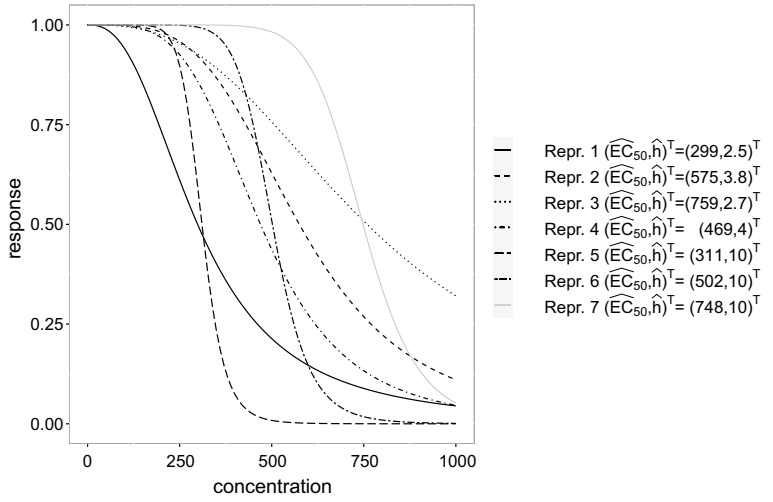



**Fig. 2** Sigmoid Emax model fits to the representative parameter vectors given in Table 1 where the linear parameters are fixed to $E_0 = 1$ and $E_{max} = -1$, respectively

Note that each of the seven areas in Fig. 1, which are encircled in black, indicates a frequency of more than 5% and the union of these seven areas constitutes 61% of the parameter estimates contained in $\Theta_G$. Therefore, we concentrate on the analysis of these areas, which are listed with corresponding representative estimates in Table 1. Note that the representative parameter estimates were drawn randomly from each of the seven areas, respectively, and that we denote the set of these vectors by $\Theta_7 = \{\theta^1, \ldots, \theta^7\}$.

In Fig. 2 the resulting estimated sigmoid Emax curves of the seven representative parameter estimates are shown, where the $E_0$ and $E_{\max}$ values are set to 1 and $-1$, respectively, for the sake of comparability. We observe that the representative parameter estimates result in both curves that are saturated within the design space $\mathcal{X} = [0, 1000]$ (i.e. representatives 4 and 6) and in curves that are still significantly decreasing at the upper bound of the design space $\mathcal{X}$ (see representative 1, 2, 3). Thus, the determination of a design that is suitable for the joint estimation of these different curves is demanding and we concentrate on that task in the following section.

**Construction of *K*-means and *D*-optimal design for simultaneous inference**

For the construction of the *K*-means design and the simultaneous *D*-optimal design, the locally *D*-optimal designs based on the $G = 15{,}233$ different parameter estimations in $\Theta_G$ are necessary. Moreover, the locally *D*-optimal designs for these parameters are needed for the calculation of the *D*-efficiencies defined by (4). Consequently, we first determined the locally *D*-optimal designs using the results of its principle structure stated in the section "The sigmoid Emax model". In particular, in the present situation the locally *D*-optimal design $\xi^g_{\text{Dopt}}$ for a particular parameter $\theta^g = (\hat{E}_0, \hat{E}_{\max}, \hat{E}_{50}, \hat{h})^T \in \Theta_G$ is of the form

$$\xi^g_{\text{Dopt}} = \begin{pmatrix} 0 & x_2 & x_3 & 1000 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{pmatrix}.$$

The additional support points $x_2, x_3$ were calculated numerically in dependence of $\theta^g$ using the particle swam optimization-algorithm (PSO), which is a heuristic optimization algorithm (see [36] for details). To check the *D*-optimality of the designs obtained by PSO, the inequality of the corresponding Equivalence Theorem given in (3) was checked, respectively. Note again that it holds $\xi^g_{\text{Dopt}} \neq \xi^{g'}_{\text{Dopt}}$ for $\theta^g \neq \theta^{g'}$, in general and a design that is *D*-optimal for $\theta^g$ might not be appropriate for the parameter $\theta^{g'}$ (see the section "Optimal designs for simultaneous inference" and "The sigmoid Emax model" for details).

For the construction of an equally weighted *K*-means design, we follow the procedure described in the section "Optimal designs for simultaneous inference". Imitating the property of the original design given in Table 2, which has 8 support points with the placebo concentration weighted twice, the total number of different support points for the *K*-means design is fixed to $L = 9$. Due to the structure of the 15,233 different locally *D*-optimal designs the intersection of all supports is given by $C_0 = \{0, 1000\}$, such that $\tilde{L} = 2$. Thus, the clustering is done for $K = 9 - 2 = 7$ clusters on the set $C \setminus C_0$ which contains the support points of the different locally *D*-optimal designs without the concentrations 0 and 1000. The resulting equally weighted *K*-means design with 9 support points (short: *K*-means design) can be found in Table 2 where the support points are rounded to integers for the sake of readability.

**Table 2** Considered designs with information of used concentrations and corresponding weights

| Design | Notation | Concentrations and corresponding weights | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Original | $\xi_{\text{orig}}$ | 0 | 25 | 150 | 350 | 450 | 550 | 800 | | 1000 |
| | | $\frac{2}{9}$ | $\frac{1}{9}$ | $\frac{1}{9}$ | $\frac{1}{9}$ | $\frac{1}{9}$ | $\frac{1}{9}$ | $\frac{1}{9}$ | | $\frac{1}{9}$ |
| Equidistant | $\xi_{\text{equi}}$ | 0 | 125 | 250 | 375 | 500 | 625 | 750 | 875 | 1000 |
| | | $\frac{1}{9}$ | $\frac{1}{9}$ | $\frac{1}{9}$ | $\frac{1}{9}$ | $\frac{1}{9}$ | $\frac{1}{9}$ | $\frac{1}{9}$ | $\frac{1}{9}$ | $\frac{1}{9}$ |
| Log-equidistant | $\xi_{\log}$ | 0 | 1 | 3 | 7 | 19 | 52 | 139 | 373 | 1000 |
| | | $\frac{1}{9}$ | $\frac{1}{9}$ | $\frac{1}{9}$ | $\frac{1}{9}$ | $\frac{1}{9}$ | $\frac{1}{9}$ | $\frac{1}{9}$ | $\frac{1}{9}$ | $\frac{1}{9}$ |
| *K*-means | $\xi_{\text{kmeans}}$ | 0 | 89 | 209 | 326 | 428 | 536 | 652 | 798 | 1000 |
| | | $\frac{1}{9}$ | $\frac{1}{9}$ | $\frac{1}{9}$ | $\frac{1}{9}$ | $\frac{1}{9}$ | $\frac{1}{9}$ | $\frac{1}{9}$ | $\frac{1}{9}$ | $\frac{1}{9}$ |
| Simultaneous *D*-optimal | $\xi_{\Theta_7}$ | 0 | 145 | 280 | 345 | 457 | 575 | 656 | 781 | 1000 |
| | | 0.17 | 0.05 | 0.12 | 0.12 | 0.11 | 0.14 | 0.03 | 0.06 | 0.20 |

For the determination of the simultaneous $D$-optimal design in (5), a discrete distribution $\pi$ on $\Theta_G$ is necessary. The choice of the uniform distribution on the parameter set $\Theta_G$ given by (8) is not appropriate in the case under consideration, as the number of different parameter estimates contained in this set is huge ($G = 15{,}233$) and the corresponding $D$-optimality criterion for simultaneous inference becomes numerically instable. Therefore, we recall the seven areas and the corresponding representative parameter estimates presented in Table 1 instead. We set the support of the considered distribution $\pi$ to the set $\Theta_7 = \{\theta^1, \ldots, \theta^7\}$, which are the representative parameter estimates of the seven significant parameter areas (see Table 1). For $\theta \in \Theta_7$, the probability $\pi(\theta) = \pi_7(\theta)$ is set to the readjusted relative frequency of the corresponding area, which is listed in the last column of Table 1. For $\theta \in \Theta_G \setminus \Theta_7$, it then follows $\pi_7(\theta) = 0$.

Using PSO based on the distribution $\pi_7$, we obtain the simultaneous $D$-optimal design given in Table 2 where the support points are again rounded to integers for the sake of readability. Note that the optimality of the design $\xi_{\Theta_7}$ can be checked by plotting the function $s(x, \xi_{\Theta_7}, \pi_7)$ given in (6) (see Additional file 1: Figure S1).

In the section "Comparison of the designs", the $K$-means design, the simultaneous $D$-optimal design and the original design are compared concerning different measures of performance. Furthermore, we include an equidistant and a log-equidistant design with nine support points on the design space $\mathcal{X} = [0, 1000]$ in the analysis, since such designs are commonly used in the context of gene expression data (see [37, 38], among many others). While Pinheiro and Bornkamp [39] argued that the log-equidistant design is superior to the equidistant design if used for the analysis of one concentration–response curve, we investigate whether this also holds true in the context of the simultaneous analysis of gene expression data.

All designs under consideration are shown in Table 2, an illustration of them is contained in the Additional file 1: Figure S2.

## Comparison of the designs

In the following sections, the performances of the different designs depicted in Table 2 are investigated when they are used for the estimation of the 15,233 concentration–response curves. In the section "Comparison with respect to the D-efficiencies", the designs are compared with respect to their $D$-efficiencies, whereas in the section "Comparison using a simulation study" the designs are used to simulate new concentration–response data for each of the 15,233 genes. Based on this data, new concentration–response curves are estimated and compared to the curves obtained by the original VPA-data.

### Comparison with respect to the *D*-efficiencies

As stated in the section "Optimal designs for simultaneous inference", the performance of a given design $\xi$ can be measured using the $D$-efficiency $\text{Eff}_D(\xi, \theta)$ defined in (4), where $\theta \in \Theta$ is the assumed true parameter vector. The greater the $D$-efficiency is, the better the corresponding design performs.

We analyse the $D$-efficiencies of the designs depicted in Table 2 for all parameter vectors contained in $\Theta_G$, i.e. the parameter set containing the $G = 15{,}233$ significant parameter vectors for the corresponding genes (see Eq. (8) and the section "Data preprocessing

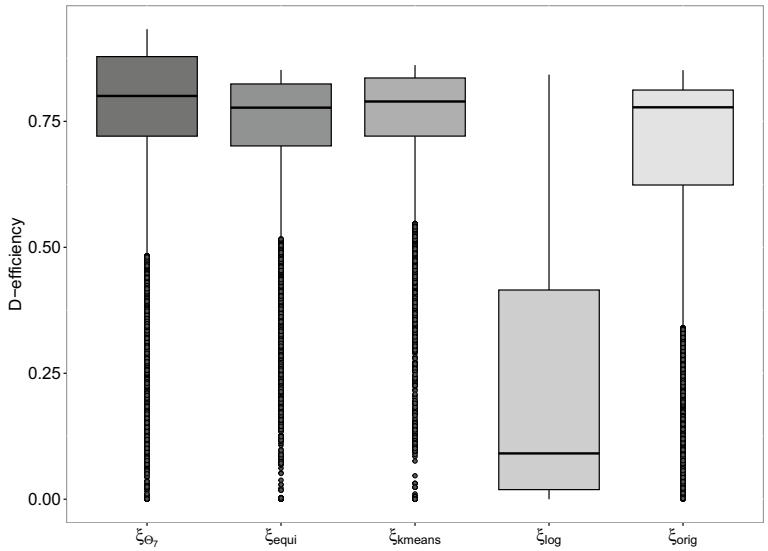



**Fig. 3** *D*-efficiencies of the different designs under consideration assuming the different parameter vectors $\theta \in \Theta_G$

**Table 3** Descriptive parameters of the *D*-efficiencies regarding each design, minimum, 0.25%- and 0.75%-quantiles, median, mean and maximum

| Design | Min | 0.25% | Median | Mean | 0.75% | Max |
|---|---|---|---|---|---|---|
| Simultaneous *D*-optimal | 0.000 | 0.721 | 0.801 | 0.745 | 0.879 | 0.933 |
| Equidistant | 0.000 | 0.701 | 0.777 | 0.716 | 0.824 | 0.852 |
| *K*-means | 0.000 | 0.721 | 0.789 | 0.732 | 0.836 | 0.861 |
| Log-equidistant | 0.000 | 0.019 | 0.091 | 0.228 | 0.415 | 0.843 |
| Original | 0.000 | 0.624 | 0.778 | 0.682 | 0.812 | 0.852 |

and analysis" for details). In Fig. 3, the resulting *D*-efficiencies are presented as box plots for each considered design, respectively. Whereas Table 3 depicts the corresponding benchmark *D*-efficiencies. i.e. their minima, maxima, quantiles, medians and means. Based on the *D*-efficiencies the log-equidistant design performs worst, whereas the simultaneous *D*-optimal design performs best.

More precise, the log-equidistant design $\xi_{\log}$ has a *D*-efficiency smaller than 0.5 for more than 75% of the parameter vectors contained in the parameter set $\Theta_G$ and a median *D*-efficiency of 0.09, which indicates a bad performance with respect to the *D*-optimality criterion for most of the considered parameter vectors.

In contrast 75% of the *D*-efficiencies of the simultaneous *D*-optimal design are greater than 0.72 and its median efficiency is equal to 0.80. In particular, the maximal *D*-efficiencies of the equidistant, original and *K*-means design are smaller than the 75%-quantile of the simultaneous *D*-optimal design, respectively. Thus, for 25% of all genes the simultaneous *D*-optimal design provides higher *D*-efficiencies than the maximal *D*-efficiency of the other designs, respectively. Nevertheless, the *K*-means design also results in high *D*-efficiencies, 75% of the *D*-efficiencies are greater than 0.72 and its median is given by 0.79. The *D*-efficiencies of the equidistant design are similar to the *D*-efficiencies of the

$K$-means design, which can be explained by the similar structure of these designs. Note that both designs are equally weighted and their support points are similarly distributed over the design space $\mathcal{X} = [0, 1000]$ (cf. Table 2 and Figure S2, see Additional file 1). Considering the original design, 75% of its $D$-efficiencies are greater than 0.62.

Note that there are several outliers for all designs under consideration. We investigated that parameter vectors that lead to very small $D$-efficiencies of the original design result in significantly small $D$-efficiencies of the other designs and almost vice versa. In particular, parameter combinations with $EC_{50} \leq 200$ and large steepness $h > 8$ lead to small $D$-efficiencies independent from the design.

Although the simultaneous $D$-optimal design performs best, when the $D$-efficiencies are compared without restrictions, a worse performance could be possible for parameter vectors $\theta$, that are not part of the seven areas included in the distribution $\pi_7$. Therefore, we also investigate the $D$-efficiencies of the different designs grouped by the nonlinear parameters of the sigmoid Emax model, i.e. $EC_{50}$ and $h$. Analogously to the overall analysis the simultaneous $D$-optimal design shows the highest $D$-efficiencies for most of the groups: Only for a few parameter constellations, that were not considered in the construction of the design, lower $D$-efficiencies occurred. Regarding the remaining considered designs, the structure of their $D$-efficiencies was similar to the one in the overall analysis, respectively. A detailed description and additional figures of the grouped analysis can be found in the Additional file 1.

Summarizing the simultaneous $D$-optimal design performs best with respect to the $D$-efficiencies, whereas the log-equidistant design performs worst. The $K$-means and the equidistant design perform well, resulting in similar $D$-efficiencies for the considered cases. The original design results in $D$-efficiencies which are in principle allocated between the ones of the log-equidistant and the equidistant design.

### Comparison using a simulation study

In this section, we report the results of a simulation to investigate the performance of the different designs in Table 2 in scenarios that imitate concentration–response experiments for the 15,233 significant genes contained in the VPA-data set. In the section "Simulation study setup", we introduce the design of the simulation study, including its assumptions and scenarios. Further, we describe the normalized root mean square error (NRMSE) which is used to evaluate the performance of the different designs within the study. In the section "Simulation results", we summarize the results of the simulation study with respect to the NRMSE.

#### *Simulation study setup*

Imitating the original data set of Krug et al. [16], we investigate the performance of the five different designs if used for the simultaneous estimation of the concentration–response relationships corresponding to the 15,233 significant genes of the data set. We consider simulation parameters denoted in Table 4.

More precisely, we consider six different sample sizes $N = 18, 27, 36, 45, 63, 90$ for each design given in Table 2. For the equidistant, log-equidistant and $K$-means design varying the sample size $N$ leads to equal repetitions at every concentration, as these designs are equally weighted, whereas for the original design, there are twice as much

**Table 4** Simulation parameters

| Parameter | Variation/Value |
|---|---|
| Sample size | $N \in \{18, 27, 36, 45, 63, 90\}$ |
| Design | $\xi_{\text{orig}}, \xi_{\text{equi}}, \xi_{\text{logequi}}, \xi_{\text{kmeans}}, \xi_{\Theta_7}$ |
| Gene | $g \in \{1, \ldots, 15233\}$ |
| Model parameter | $\theta^g \in \Theta_G = \{\theta^1, \ldots, \theta^{15,233}\}$ |
| Simulation step | $j \in \{1, \ldots, 500\}$ |
| $\sigma$ | $0.2 \cdot \lvert E^g_{\max} \rvert$ |

repetitions at placebo ($x = 0$) as at the remaining non-placebo concentrations. In case of unequal repetitions as attained for the simultaneous $D$-optimal design the procedure of efficient rounding according to Pukelsheim and Rieder [18] is used to obtain integer numbers of repetitions.

Further, we assume that the concentration–response relationship of each significant gene $g$ is described by the nonlinear regression model (1), where the regression function is given by the sigmoid Emax model in (7) with the corresponding true parameter given by the estimate $\theta^g \in \Theta_G$ given by (8). The errors in model (1) are assumed to be normally distributed with standard deviation $\sigma = 0.2 \cdot |E^g_{\max}|$, where $E^g_{\max}$ is the maximal effect of gene $g$, respectively. This results in $\mathcal{S} = 5 \cdot 6 = 30$ different scenarios in total, and for each scenario, we obtain data from 15,233 concentration–response relationships.

We used $N_{\text{sim}} = 500$ simulation runs for each scenario and in each simulation step, the sigmoid Emax model is fitted to the data of each gene separately.

We use the Root Mean Squared Error (RMSE) to evaluate the performance of the different designs. For a given scenario $S$ (out of the $\mathcal{S} = 30$ scenarios), let $\eta(\cdot, \hat{\theta}^g_{jS})$ denote the estimated sigmoid Emax model with corresponding estimated model parameter $\hat{\theta}^g_{jS}$ for the data generated for gene $g$ in simulation $j$ ($g = 1, \ldots, 15233$, $j = 1, \ldots, 500$). Moreover, let $\eta(\cdot, \theta^g)$ denote the data generating sigmoid Emax model of the $g$-th significant gene. Following Cheema [40], the RMSE (of the $S$-th scenario for the $g$-th gene) is then given by

$$\text{RMSE}(g, S) = \frac{1}{N_{\text{sim}}} \sum_{j=1}^{N_{\text{sim}}} \sqrt{\frac{1}{1001} \sum_{i=0}^{1000} \left( \eta(x_i, \hat{\theta}^g_{jS}) - \eta(x_i, \theta^g) \right)^2},$$

where $x_0, \ldots, x_{1000}$ are given by the sequence $0, 1, \ldots, 1000 \in \mathcal{X}$. There is high variability in the ranges of the expression values across genes, which is not accounted for by the RMSE. For example, if we consider two genes with the same RMSE value of 2, but with different response ranges, e.g. 4 and 10, the RMSE value for the gene with larger range shows a higher model precision compared to the gene with a smaller range, although this is not reflected directly by the RMSE. In order to obtain comparability between the curves associated to the different genes, it is useful to standardize the RMSE. This is achieved by dividing the RMSE by $\text{E}^g_{\max}$, which is the maximal range of the curve corresponding to gene $g$. Thus, it holds:

$$\mathrm{NRMSE}(g, S) = \frac{\mathrm{RMSE}(g,S)}{E^g_{\max}}. \tag{10}$$

Note that the smaller the NRMSE is, the closer the fitted model is to the true concentration–response relationship.

#### *Simulation results*

In the section "Results of the different designs with fixed sample size", we present the results for the different designs contained in Table 2, where the sample size is fixed to the sample size $N = 27$, which coincides with the sample size of the original data set (see the section "Data" for details). In the section "Variation of sample size", we analyse the influence of the sample size on the performance of the different designs.

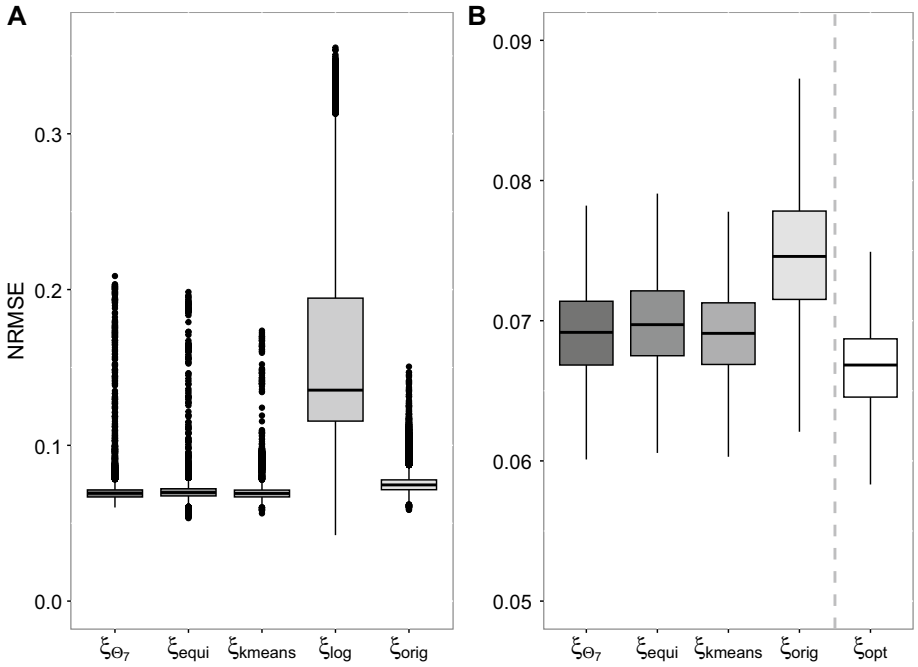



**Fig. 4** **A** NRMSEs for each gene with respect to the different designs. **B** NRMSEs for each gene with respect to the different designs, where the values of the log-equidistant designs and the outliers are removed

**Table 5** Descriptive parameters of the *D*-efficiencies regarding each design, minimum, 0.25%- and 0.75%-quantiles, median, mean and maximum

| Design | Min | 0.25% | Median | Mean | 0.75% | Max |
|---|---|---|---|---|---|---|
| Simultaneous *D*-optimal | 0.060 | 0.067 | 0.069 | 0.071 | 0.071 | 0.209 |
| Equidistant | 0.053 | 0.068 | 0.070 | 0.070 | 0.072 | 0.198 |
| *K*-means | 0.056 | 0.067 | 0.069 | 0.070 | 0.071 | 0.174 |
| Log-equidistant | 0.042 | 0.116 | 0.135 | 0.161 | 0.194 | 0.355 |
| Original | 0.059 | 0.072 | 0.075 | 0.076 | 0.078 | 0.151 |
| Locally *D*-optimal | 0.053 | 0.065 | 0.067 | 0.067 | 0.069 | 0.147 |

***Results of the different designs with fixed sample size***

In the left part of Fig. 4, we display the NRMSE defined by (10) for the 15,233 curves with box plots grouped by the different designs under consideration. Also Table 5 depicts corresponding benchmark values of the NRMSEs like minima, maxima, quantiles, medians and means.

It is clearly visible that the NRMSEs corresponding to the log-equidistant designs are greater than the NRMSEs of the other designs. In particular, almost 50% of the considered genes have an NRMSE, that is almost twice as large as for the other designs. Moreover, the NRMSEs obtained with the log-equidistant design, are extremely varying compared to the others. Thus, it follows that using the log-equidistant design is not reasonable for the data at hand and we restrict ourselves to the analysis of the NRMSEs obtained for the original, the equidistant, the *K*-means, and the simultaneous *D*-optimal design, respectively. For that purpose, we also remove the outliers, as no structure of these could be detected in dependence of the design choice. Further analyses resulted in observations similar to the ones for the *D*-efficiencies: In particular, outliers occur mostly for extreme parameter combinations such as $\mathrm{EC}_{50} \leq 2$ and $h > 8$.

The right part of Fig. 4 shows the NRMSEs without outliers grouped by all designs apart from the log-equidistant design. Additionally, the NRMSE of the locally *D*-optimal design of each gene *g* is provided, respectively. Note that the locally *D*-optimal designs, in general called $\xi_{\mathrm{opt}}$, lead to the smallest NRMSEs and constitute the best choice for evaluating each gene separately. Nevertheless, these designs are not applicable in the experiment under consideration, which is done simultaneously for all genes. Consequently, the box plot of the NRMSEs based on the locally *D*-optimal designs has to be interpreted as benchmark.

Apart from the locally optimal designs, the best results with respect to the NRMSE are achieved by the simultaneous *D*-optimal design and the *K*-means, followed by the NRMSEs of the equidistant design. For instance, the median is 0.069 for both $\xi_{\Theta_7}$ and $\xi_{\mathrm{kmeans}}$ and 0.070 for the equidistant design. These designs result in NRMSEs that are close to the ones of the locally *D*-optimal designs, which indicates a good performance with respect to the NRMSE. Finally, the original design results in larger NRMSEs than the other designs. In particular, the lower quartile of NRMSEs is given by 0.072, which is even greater than the upper quartiles of the equidistant, *K*-means, and simultaneous *D*-optimal design.

Similarly to the theoretical analysis of the *D*-efficiencies in the section "Comparison with respect to the D-efficiencies", we investigated the NRMSEs grouped by the parameters $\mathrm{EC}_{50}$ and *h*. Summarizing, the comparison of the RMSEs of the designs (stratified to different parameter constellations) leads to similar results as within the total analysis. A detailed description and corresponding figures can be found in the Additional file 1.

Summarizing, the designs constructed in the section "Construction of K-means and D-optimal design for simultaneous inference" outperform the original design with respect to the NRMSEs. Moreover, the equidistant design, which has support points similar to the ones of the *K*-means design, results in appropriate NRMSEs, whereas for the log-equidistant design NRMSEs are substantially higher in comparison to all other designs. Both the simultaneous *D*-optimal design and the *K*-means result in the best NRMSEs and therefore in the best simultaneous inference of the 15,233 concentration–response relationships.

***Variation of sample size***

We consider the NRMSEs of the different designs when the sample size $N$ is varied, that is $N = 18, 27, 36, 45, 63, 90$. In Fig. 5, we display the NRMSEs grouped by design and the total sample sizes.

For all considered designs the NRMSEs are decreasing with increasing sample size, which implies an increase in the precision of the corresponding model fits. In particular, the median NRMSEs of all designs are located within 0.084 and 0.092 for $N = 18$ measurements and decrease to values between 0.038 and 0.042 for $N = 90$. Thus, the NRMSEs are almost reduced to half of the values by increasing the sample size, if the sample size is multiplied by 4. This effect is well explained by the convergence rate of the maximum likelihood estimator, which is $\sqrt{N}$. It follows, that the absolute reduction of the NRMSEs is higher for smaller sample size, in particular, if $N = 27$ observations are used instead of $N = 18$. Thus, in the situation under consideration, it is reasonable to consider at least $N = 27$ observations.

Comparing the NRMSEs of the original design to the NRMSEs of the simultaneous $D$-optimal design (or the equidistant and $K$-means design, respectively), the following can be observed: The NRMSEs of the original design for $N = 36$ are similar to the NRMSEs of the simultaneous $D$-optimal design for $N = 27$. Similar observations are possible if the sample sizes are increased step wisely. That means that at least 9 more observations are necessary, if the original design is used, in order to achieve the precision of the simultaneous $D$-optimal design.

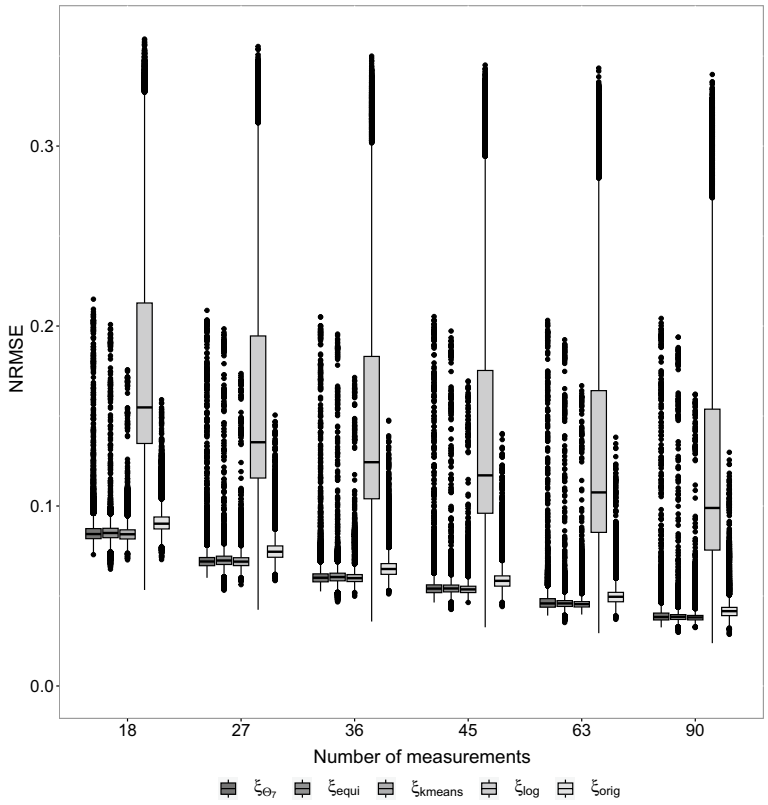



**Fig. 5** NRMSE values for each gene regarding different designs with varying number of measurements

## Conclusion and outlook

This paper introduced two ways to construct designs that address the problem of the simultaneous inference of a large number of concentration–response relationships: the *K*-means design which is based on locally *D*-optimal designs of individual concentration–response curves and the *D*-optimal design for simultaneous inference which incorporates the distribution of the nonlinear parameters of the different concentration–response curves. In order to investigate the performance of these designs, we used the VPA-data set by Krug et al. [16] and constructed the corresponding designs for the relevant concentration–response relationships contained in this data set. Then the designs were compared to the design that was originally used for generating the VPA-data set and to an appropriate log-equidistant and an equidistant design. The comparison was done in terms of *D*-efficiencies and in terms of the performance for the simultaneous estimation of 15,233 concentration–response curves in a simulation study (imitating the original VPA-data).

In terms of *D*-efficiency and terms of the simulation study, we observed similar results: the simultaneous *D*-optimal design results in the most precise model fits for the different curves under consideration, the model fits obtained if the *K*-means and the equidistant design used are also appropriate. The log-equidistant design performs worse and the corresponding precision of the model fits is the lowest. Consequently, it is not recommendable to use the log-equidistant design if a large number of different concentration–response relationships (with considerably different shapes) should be estimated.

In general, it follows that the simultaneous *D*-optimal design improves the inference substantially. While the *K*-means design also performs well in the situation under consideration, it might be less feasible due to its construction. The construction of the simultaneous *D*-optimal design is straightforward by including a rough distribution of the nonlinear parameter values. Note that the equidistant design also performs well with respect to the considered measures in the situation under consideration, but it might perform worse in others. The advantage of the simultaneous *D*-optimal design is its flexibility: it can easily be adapted to the situation at hand, based on a distribution which can be predefined by the user.

The paper is based on the assumption that all concentration–response relationships are modelled appropriately by the same nonlinear regression function (with varying parameters). Different parametric regression functions could be assumed and the distribution of the occurrence of these curves could be included in the *D*-optimality criterion for simultaneous inference. Moreover, we restrict ourselves to the *D*-optimality criterion, other criteria could also be used to construct a design for simultaneous inference, e.g. addressing the precise estimation of the $EC_{50}$ values or the prediction of responses at predefined concentrations. We leave the extension of these approaches to future research.

## Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12859-023-05526-3.

**Additional file 1.** Equivalence Theorem for $D$-optimal design for simultaneous inference, Additional Figures, Comparison of the designs grouped by parameter $EC_{50}$ and $h$.

**Acknowledgements**
The authors thank the anonymous reviewers for their valuable suggestions.

**Author contributions**
All authors provided a concept of this work. LS and KS analysed the results. LS and KS wrote the manuscript. LS, KS and JR reviewed the manuscript.

**Funding**
Open Access funding enabled and organized by Projekt DEAL. This work has been supported by the Research Training Group "Biostatistical Methods for High-Dimensional Data in Toxicology" (RTG 2624, Project P5) funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation - Project Number 427806116).

**Data availibility**
The data was originally provided by Krug et al. [16]. The data set as it is used in this publication can be found at https://github.com/schuermeyer/Designs-for-the-simultaneous-inference-of-concentration-response-curves. In addition the R-code and the main functions used in our analysis are provided.

## Declarations

**Ethics approval and consent to participate**
Not applicable.

**Consent for publication**
Not applicable.

**Competing interests**
The authors declare that they have no competing interests.

Received: 6 April 2023 Accepted: 9 October 2023
Published online: 19 October 2023

**References**

1. Möllenhoff K, Schorning K, Kappenberg F. Identifying alert concentrations using a model-based bootstrap approach. Biometrics. 2022. https://doi.org/10.1111/biom.13799.
2. Kappenberg F, Grinberg M, Jiang X, Kopp-Schneider A, Hengstler JG, Rahnenführer J. Comparison of observation-based and model-based identification of alert concentrations from concentration-expression data. Bioinformatics. 2021. https://doi.org/10.1093/bioinformatics/btab043.
3. House JS, Grimm FA, Jima DD, Zhou Y-H, Rusyn I, Wright FA. A pipeline for high-throughput concentration response modeling of gene expression for toxicogenomics. Front Genet. 2017;8:168.
4. Collins MD, Cui EH, Hyun SW, Wong WK. A model-based approach to designing developmental toxicology experiments using sea urchin embryos. Arch Toxicol. 2022; 96:919–932.
5. Dragalin V, Hsuan F, Padmanabhan SK. Adaptive designs for dose-finding studies based on sigmoid e max model. J Biopharm Stat. 2007;17(6):1051–70.
6. Li G, Majumdar D. D-optimal designs for logistic models with three and four parameters. J Stat Plan Inference. 2008;138(7):1950–9.
7. Wang T, Yang M. Adaptive optimal designs for dose-finding studies based on sigmoid emax models. J Stat Plan Inference. 2014;144:188–97.
8. Dette H, Schorning K. Optimal designs for comparing curves. Ann Stat. 2016;44(3):1103.
9. Feller C, Schorning K, Dette H, Bermann G, Bornkamp B. Optimal designs for dose response curves with common parameters; 2017;45(5):2102–2132.
10. Dror HA, Steinberg DM. Robust experimental design for multivariate generalized linear models. Technometrics. 2006;48(4):520–9.
11. Dong W, Tang X, Yu Y, Nilsen R, Kim R, Griffith J, Arnold J, Schüttler H-B. Systems biology of the clock in *Neurospora crassa*. PLoS ONE. 2008;3(8):3105.
12. Bouffier AM, Arnold J, Schüttler HB. A mine alternative to d-optimal designs for the linear model. PLoS ONE. 2014;9(10): 110234.
13. McGee RL, Buzzard GT. Maximally informative next experiments for nonlinear models. Math Biosci. 2018;302:1–8.
14. Chaloner K. Bayesian design for estimating the turning point of a quadratic regression. Commun Stat Theory Methods. 1989;18(4):1385–400.
15. Dette H, Haines LM, Imhof LA. Maximin and Bayesian optimal designs for regression models. Stat Sin. 2007;17:463–80.
16. Krug AK, Kolde R, Gaspar JA, Rempel E, Balmer NV, Meganathan K, Vojnits K, Baquié M, Waldmann T, Ensenat-Waser R, Jagtap S, Evans RM, Julien S, Peterson H, Zagoura D, Kadereit S, Gerhard D, Sotiriadou I, Heke M, Natarajan K, Henry M, Winkler J, Marchan R, Stoppini L, Bosgra S, Westerhout J, Verwei M, Vilo J, Kortenkamp A, Hescheler JR, Hothorn L, Bremer S, van Thriel C, Krause KH, Hengstler JG, Rahnenfuhrer J, Leist M, Sachinidis A. Human embryonic stem cell-derived test systems for developmental neurotoxicity: a transcriptomics approach. Arch Toxicol. 2013;87(1):123–43.
17. Kiefer J. General equivalence theory for optimum designs (approximate theory). Ann Stat. 1974;2(5):849–79.
18. Pukelsheim F, Rieder S. Efficient rounding of approximate designs. Biometrika. 1992;79(4):763–70.
19. Chernoff H. Locally optimal designs for estimating parameters. Ann Math Stat. 1953;24:586–602.

20. Pukelsheim F. Optimal design of experiments. Philadelphia: SIAM; 2006.
21. Fedorov VV, Leonov SL. Optimal design for nonlinear response models. Cambridge: CRC Press; 2013.
22. Pronzato L, Walter E. Robust experiment design via stochastic approximation. Math Biosci. 1985;75(1):103–20.
23. Chaloner K, Larntz K. Optimal Bayesian design applied to logistic regression experiments. J Stat Plan Inference. 1989;21(2):191–208.
24. Chaloner K. A note on optimal Bayesian design for nonlinear problems. J Stat Plan Inference. 1993;37(2):229–35.
25. Chaloner K, Verdinelli I. Bayesian experimental design: a review. Stat Sci. 1995;10(3):273–304.
26. Hartigan JA, Wong MA. Algorithm as 136: a k-means clustering algorithm. J R Stat Soc Ser C (Appl Stat). 1979;28(1):100–8.
27. Apon A, Robinson F, Brewer D, Dowdy L, Hoffman D, Lu B. Initial starting point analysis for k-means clustering: a case study. Publications. 2006;22.
28. Macdougall J. Analysis of dose–response studies—e max model. In: Dose finding in drug development, pp. 127–145. New York: Springer; 2006.
29. Silvey SD. Optimal design. London: Chapman and Hall; 1980.
30. R Core Team: R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria 2022. R Foundation for Statistical Computing. https://www.R-project.org/.
31. Bretz F, Pinheiro JC, Branson M. Combining multiple comparisons and modeling techniques in dose–response studies. Biometrics. 2005;61(3):738–48. https://doi.org/10.1111/j.1541-0420.2005.00344.x.
32. Azadifar S, Rostami M, Berahmand K, Moradi P, Oussalah M. Graph-based relevancy–redundancy gene selection method for cancer diagnosis. Comput Biol Med. 2022;147: 105766.
33. Rostami M, Forouzandeh S, Berahmand K, Soltani M, Shahsavari M, Oussalah M. Gene selection for microarray data classification via multi-objective graph theoretic-based method. Artif Intell Med. 2022;123: 102228.
34. Duda JC, Kappenberg F, Rahnenführer J. Model selection characteristics when using mcp-mod for dose–response gene expression data. Biom J. 2022;64(5):883–97.
35. Bornkamp B, Pinheiro J, Bretz F, Sandig L. DoseFinding: planning and analyzing dose finding experiments; 2021. R package version 1.0-2. https://CRAN.R-project.org/package=DoseFinding
36. Kennedy J, Eberhart R. Particle swarm optimization. In: Proceedings of ICNN'95-international conference on neural networks, 1995;4:1942–1948. IEEE.
37. Parny M, Bernad J, Prat M, Salon M, Aubouy A, Bonnafé E, Coste A, Pipy B, Treilhou M. Comparative study of the effects of ziram and disulfiram on human monocyte-derived macrophage functions and polarization: involvement of zinc. Cell Biol Toxicol. 2021;37:379–400.
38. Snijders KE, Fehér A, Táncos Z, Bock I, Téglási A, van den Berk L, Niemeijer M, Bouwman P, Le Dévédec SE, Moné MJ, et al. Fluorescent tagging of endogenous heme oxygenase-1 in human induced pluripotent stem cells for high content imaging of oxidative stress in various differentiated lineages. Arch Toxicol. 2021;95(10):3285–302.
39. Pinheiro J, Bornkamp B. Designing phase II dose-finding studies: sample size, doses, and dose allocation weights. In: Handbook of methods for designing. Monitoring, and analyzing dose-finding trials, pp. 229–246. New York: Chapman and Hall/CRC; 2017.
40. Cheema JR. Some general guidelines for choosing missing data handling methods in educational research. J Mod Appl Stat Methods. 2014;13(2):53–75.


## Publisher's Note

# Supplementary material

Leonie Schürmeyer, Kirsten Schorning and Jörg Rahnenführer

## Equivalence Theorem for the $D$-optimal design for simultaneous inference

For the proof of statement (6) methods of convex analysis have to be used (see Silvey (1980) for details). As the criterion $\Psi$ is concave due to the concavity of $\psi_D$, a design $\xi^*$ maximizes $\Psi$ if and only if its derivative in $\xi^*$ in direction $(1-\alpha)\,\xi^* + \alpha\xi_x$ is smaller or equal to zero for all $x \in \mathcal{X}$. In particular, the derivative of the criterion $\Psi(\xi, \pi)$ evaluated in a design $\xi$ in direction $(1-\alpha)\,\xi + \alpha\xi_x$ has to be calculated, where $\xi_x$ is the design that puts weight "1" to an arbitrary but fixed point $x \in \mathcal{X}$.
In the situation of the criterion $\Psi(\xi, \pi)$ we consider the derivative:

$$\frac{\partial}{\partial\alpha}\Psi(\xi,\pi) = \lim_{\alpha\to 0}\frac{\Psi\left((1-\alpha)\xi + \alpha\xi_x\right) - \Psi\left(\xi,\pi\right)}{\alpha}$$

Applying classical rules of differentiation, this term can be rewritten by:

$$\frac{\partial}{\partial\alpha}\Psi(\xi,\pi) = \sum_{\theta\in\Theta}\frac{\pi(\theta)}{\psi_D\left(\xi^*,\theta\right)}\frac{\partial}{\partial\alpha}\psi_D\left(\xi,\theta\right) \tag{E1}$$

where $\frac{\partial}{\partial\alpha}\psi_D\left(\xi,\theta\right)$ is the derivative of the $D$-optimality criterion $\psi_D\left(\xi,\theta\right) = \left(\det\left(M(\xi,\theta)\right)\right)^{\frac{1}{p}}$ in $\xi$ in direction $(1-\alpha)\xi + \alpha\xi_x$ it holds:

$$\begin{aligned}\frac{\partial}{\partial\alpha}\psi_D\left(\xi,\theta\right) =& \frac{1}{p}\left(\det(M(\xi,\theta))\right)^{\frac{1}{p}-1} \\ & \lim_{\alpha\to 0}\frac{\det\left(M((1-\alpha)\xi+\alpha\xi_x)\right) - \det\left(M(\xi,\theta)\right)}{\alpha} \\ =& \frac{1}{p}\left(\det(M(\xi,\theta))\right)^{\frac{1}{p}-1}\det\left(M(\xi,\theta)\right) \\ & \lim_{\alpha\to 0}\frac{\alpha\mathrm{tr}\left(M^{-1}(\xi,\theta)\left\{M(\xi_x,\theta) - M(\xi,\theta)\right\}\right)}{\alpha},\end{aligned} \tag{E2}$$

where the last equality follows by the fact that $\det(A+H) - \det(A) = \det(A)\mathrm{tr}(A^{-1}H)$, for matrices $A, H \in \mathbb{R}^{p\times p}$, for $H$ small enough.

Rewriting $M(\xi_x, \theta) = \left(\frac{\partial}{\partial\theta}\eta(x,\theta)\right)\left(\frac{\partial}{\partial\theta}\eta(x,\theta)\right)^T$, (E2) can be reformulated by:

$$\begin{aligned}\frac{\partial}{\partial\alpha}\psi_D(\xi,\theta) =& \frac{1}{p}\left(\det\left(M(\xi,\theta)\right)^{\frac{1}{p}}\right) \\ & \left\{\left(\frac{\partial}{\partial\theta}\eta(x,\theta)\right)^T M^{-1}(\xi,\theta)\left(\frac{\partial}{\partial\theta}\eta(x,\theta)\right) - p\right\} \\ =& \frac{1}{p}(\det(M(\xi,\theta)))^{\frac{1}{p}} d(x,\xi,\theta),\end{aligned}$$

where $d(x,\xi,\theta)$ is given by the function depicted in (3). Inserting (E2) in formula (E1), we obtain $\frac{\partial}{\partial\alpha}\Psi(\xi,\pi) = s(x,\xi,\pi)$, where $s(x,\xi,\pi)$ is given in (6).

## Additional Figures

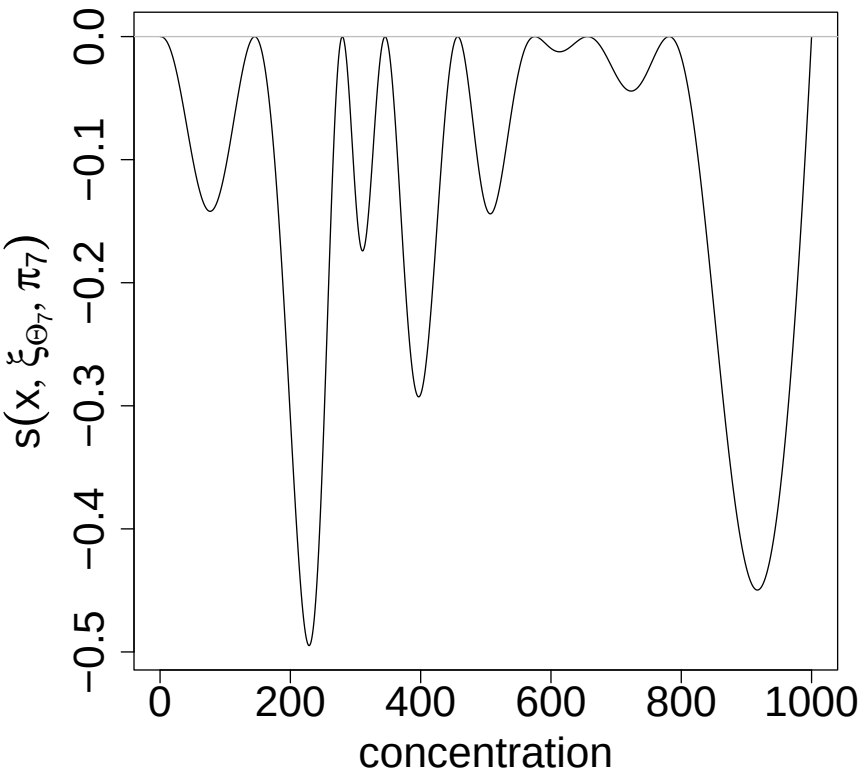



Figure S1: The plot shows the function $s(x, \xi_{\Theta_7}, \pi_7)$ in (6). Since the function is non-positive on the design space $\mathcal{X} = [0, 1000]$, the design $\xi_{\Theta_7}$ is $D$-optimal for simultaneous inference.

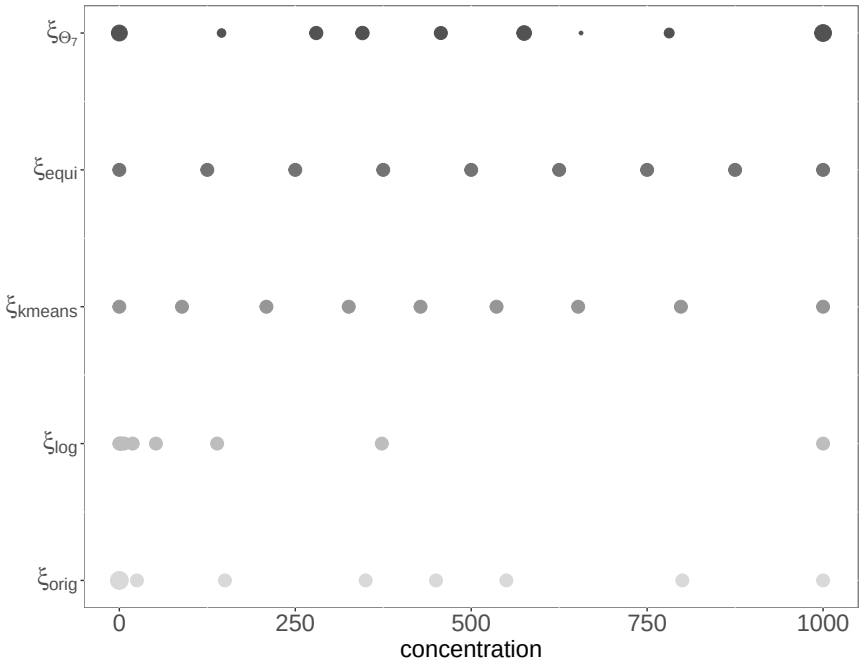



Figure S2: Support points of all considered designs with information of corresponding weights by size

# Comparison of the different designs grouped by parameter $EC_{50}$ and $h$

## Comparison of the $D$-efficiencies grouped by parameter $EC_{50}$ and $h$

In the left panel of Figure S3 the relative $D$-efficiencies grouped by design and $EC_{50}$ are shown, where the outliers are removed for the sake of clarity. The corresponding plot including outliers can be found in Figure S4A. Note that we could not observe any structure in the outliers with respect to the design.

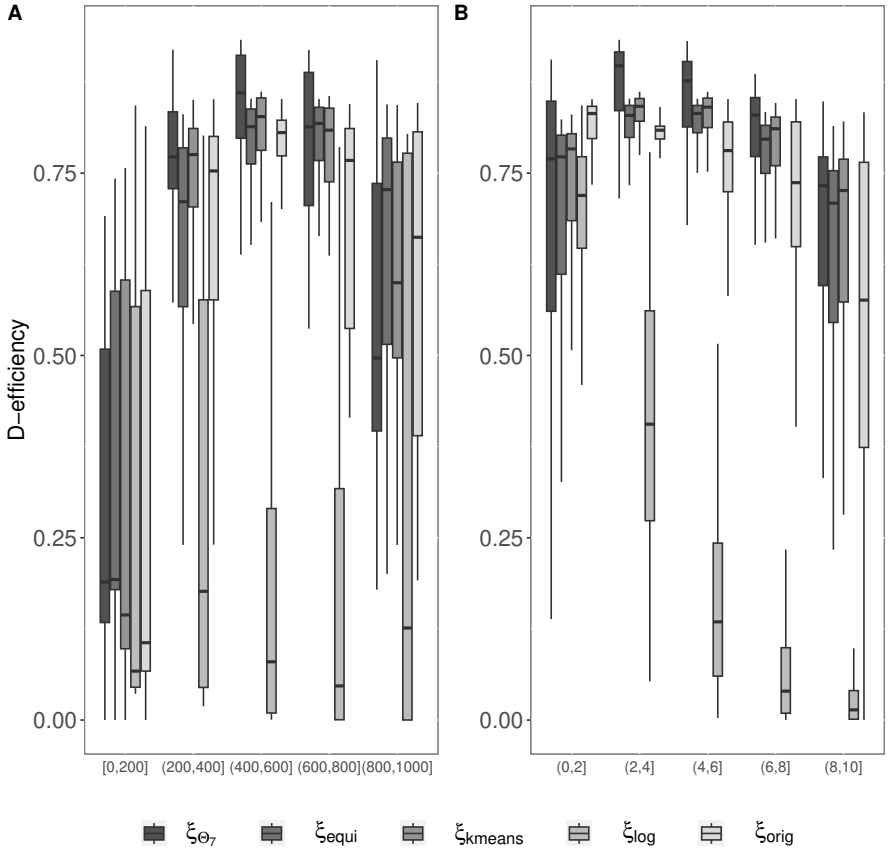



Figure S3: **A**: $D$-efficiency values for each gene regarding different designs divided by parameter $h$, **B**: $D$-efficiency values for each gene regarding different designs divided by parameter $EC_{50}$. The outliers are removed in both Figures.

The simultaneous $D$-optimal design leads to the greatest $D$-efficiencies for $\mathrm{EC}_{50} \in (200, 800]$. Note that for 84.5% of the 15 233 considered curves it holds $\mathrm{EC}_{50} \in (200, 800]$ and that the support of the used distribution $\pi_7$ only contains $\mathrm{EC}_{50}$ within $(200, 800]$, such that this effect is not surprising. For $\mathrm{EC}_{50} \leq 200$ (7.7% of all genes), the simultaneous $D$-optimal design still performs well compared to the other designs with respect to the medium $D$-efficiencies. Considering the complete distribution of $D$-efficiencies the equidistant design results in the best $D$-efficiencies, although the $D$-efficiencies are small for all considered designs if $\mathrm{EC}_{50} \leq 200$. If $\mathrm{EC}_{50} > 800$ (7.8% of all genes), the simultaneous $D$-optimal design and the log-equidistant design result in the smallest $D$-efficiencies. In this case, the equidistant design and the original designs perform best, as these designs contain two support points that are close to the assumed $\mathrm{EC}_{50}$, which are $875, 1000$ and $800, 1000$, respectively.

In the right panel of Figure S3 the relative $D$-efficiencies grouped by design and the steepness parameter $h$ are shown, where the outliers are also removed for the sake of clarity. The corresponding plot including outliers can be found in Figure S4B. Note that we could not observe any structure in the outliers with respect to the design. For all groups with $h > 2$ (87.9% of all genes), the simultaneous $D$-optimal design leads to the greatest $D$-efficiencies compared to the other designs, both with respect to the medians and the lower quartiles. The original design's $D$-efficiencies are substantially smaller for $h > 2$, whereas the equidistant and the $K$-means design have similar $D$-efficiencies that lie between the ones of the original and the simultaneous $D$-optimal design. For $h \leq 2$ (12.1% of genes), the original design results in the highest $D$-efficiencies. In particular, the lower quartile of the $D$-efficiencies of the original design is greater than the upper quartiles of the $D$-efficiencies of the equidistant, log-equidistant and the $K$-means design. The $D$-efficiencies of the simultaneous $D$-optimal design are most varying, with minimal value 0.07 and highest $D$-efficiency 0.91 for $h \leq 2$. This effect is explained by considering the discrete distribution $\pi_7$ that is used for the construction of $\xi_\theta^*$: $\pi_7(\theta) = 0$ for all $\theta = (\mathrm{EC}_{50}, h)$ with $h \leq 2$, which means, it does not aim for high $D$-efficiencies for $h \leq 2$.

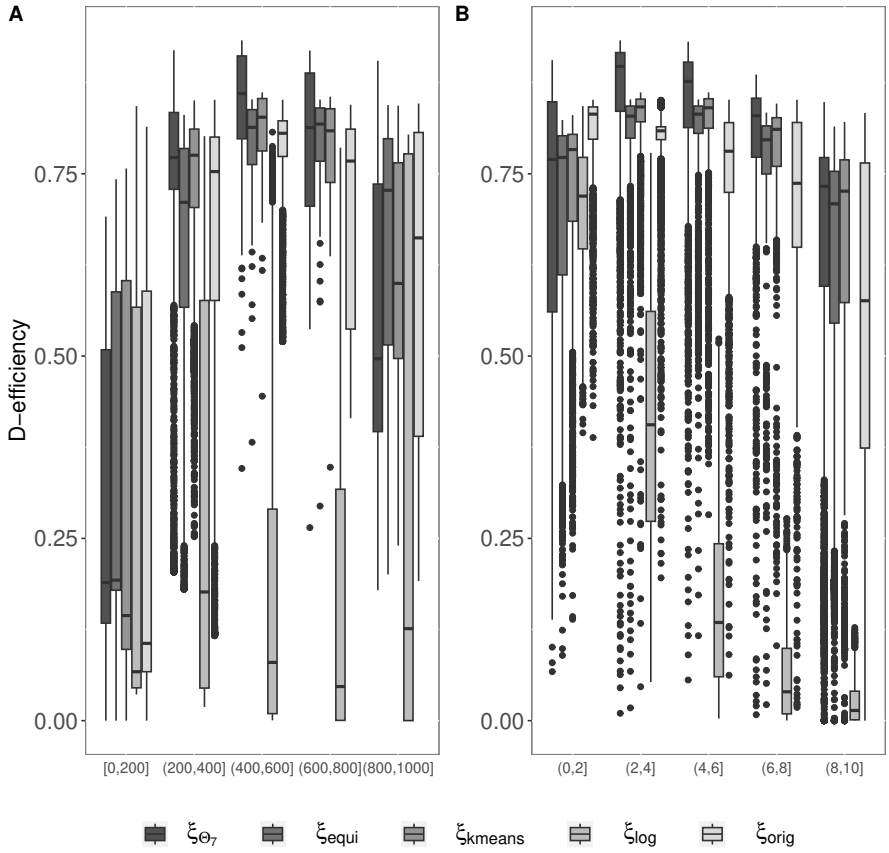



Figure S4: **A**: $D$-efficiency values for each gene regarding different designs divided by parameter $h$, **B**: $D$-efficiency values for each gene regarding different designs divided by parameter $\mathrm{EC}_{50}$

## Simulation results grouped by parameter $\mathbf{EC}_{50}$ and $h$

In the left part of Figure S5 the NRMSEs are shown grouped by $\mathrm{EC}_{50}$ and design, where the outliers are removed for the sake of clarity. The corresponding plot including outliers can be found in Figure S6A. Note that we could not observe any structure in the outliers with respect to the design.

If $\mathrm{EC}_{50} \in [200, 600)$, the box plots of the NRMSEs are similar to the ones of the general analysis for all designs under consideration. In particular, the simultaneous $D$-optimal design and the $K$-means still perform best, whereas the NRMSEs of the equidistant design are slightly larger. If $\mathrm{EC}_{50} \in [600, 1000)$, the NRMSE based on the equidistant design are the smallest compared to the other designs. Note that we observed a similar behaviour of the $D$-efficiencies of

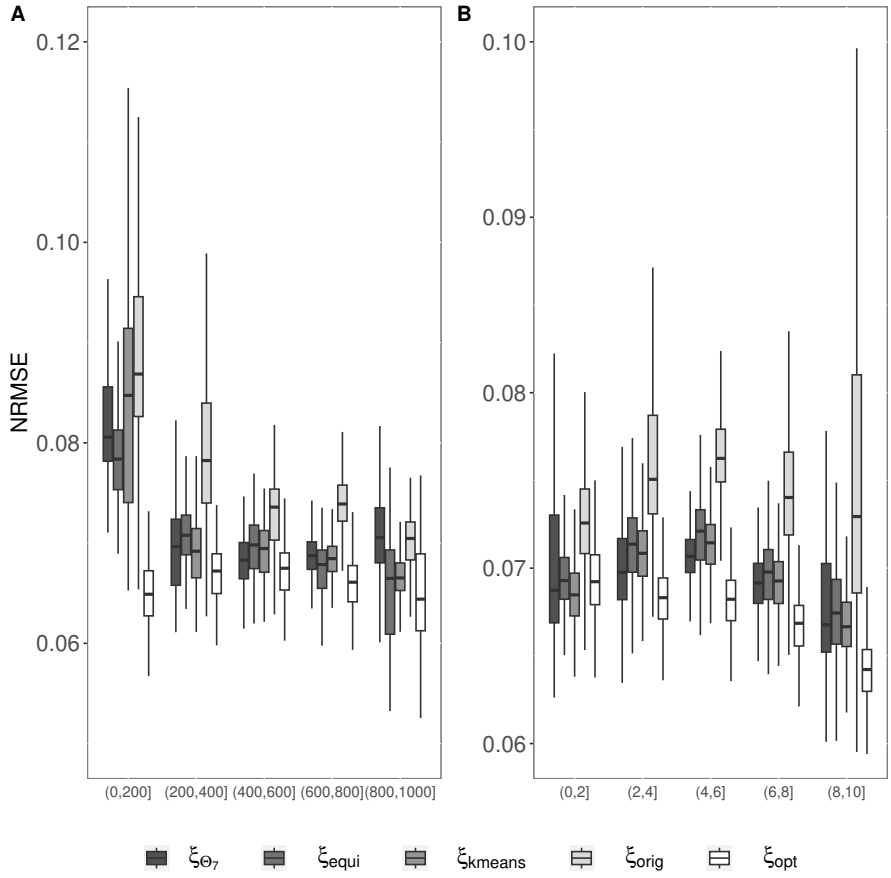



Figure S5: **A**: NRMSE values for each gene regarding different designs with 27 measurements and grouped by parameter $EC_{50}$, **B**: NRMSE values for each gene regarding different designs with 27 measurements and grouped by parameter $h$. The outliers are removed in both Figures.

the equidistant design for that case. For small $EC_{50}$-values, that is $EC_{50} \leq 200$, the considered designs result in higher NRMSE in general, whereas the ranking of the NRMSEs with respect to the designs stays almost the same. In particular, the equidistant, the $K$-means and the simultaneous $D$-optimal design still have similar NRMSEs, whereas the equidistant slightly outperforms the others. The original design performs worse with respect to the NRMSE independent from the restriction on the $EC_{50}$-value.

In the right part of Figure S5, the NRMSEs are shown grouped by $h$ and the design, where the outliers are removed for the sake of clarity. The corresponding

plot including outliers can be found in Figure S6B. Note that we could not observe any structure in the outliers with respect to the design. As for the $EC_{50}$, the NRMSEs of the different designs are only slightly varying if considered restricted to $h$. In particular, the simultaneous $D$-optimal design results in higher varying NRMSEs with a median equal to 0.069 and 0.75%-quantile at 0.073, for $h \leq 2$. Note that we observed a similar effect for the corresponding $D$-efficiencies, which could be explained by the fact that values $h \leq 2$ were not considered for the construction of this design. If $h \in (8, 10]$, the original design results in higher varying NRMSEs. Note that this effect was also observed for the corresponding $D$-efficiencies of that design (see Figure S3.B).

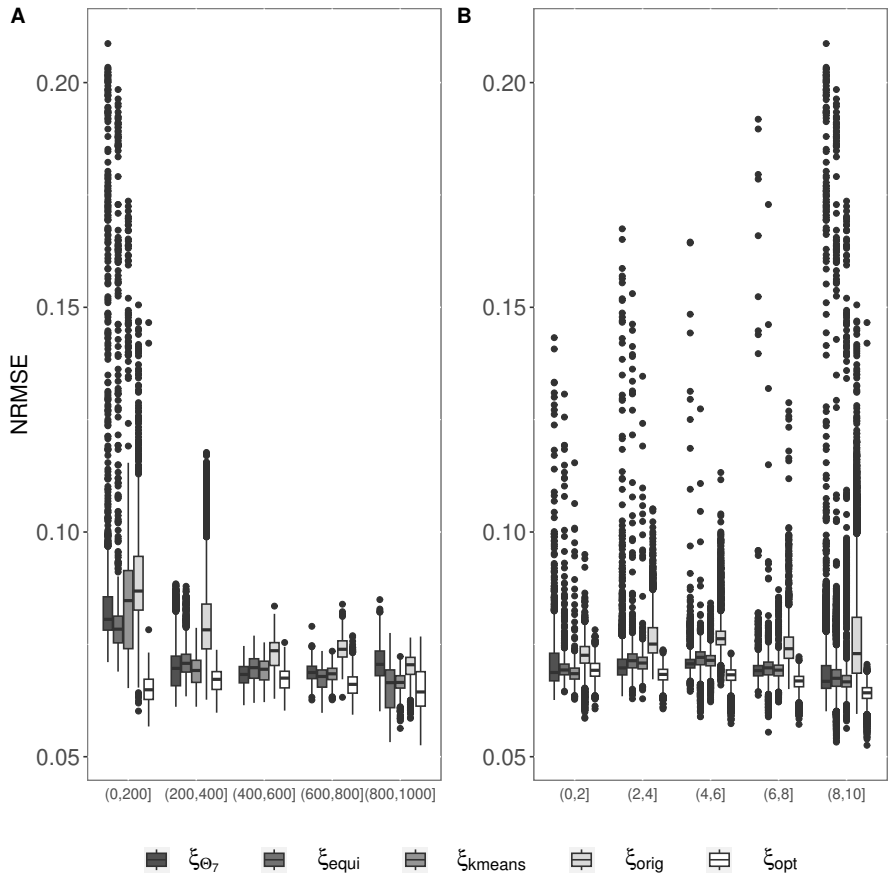



Figure S6: **A**: NRMSE values for each gene regarding different designs with 27 measurements and grouped by parameter $EC_{50}$, **B**: NRMSE values for each gene regarding different designs with 27 measurements and grouped by parameter $h$

# Eidesstattliche Versicherung (Affidavit)

Schürmeyer, Leonie
Name, Vorname
(Surname, first name)

184109
Matrikel-Nr.
(Enrolment number)

Belehrung:

Wer vorsätzlich gegen eine die Täuschung über Prüfungsleistungen betreffende Regelung einer Hochschulprüfungsordnung verstößt, handelt ordnungswidrig. Die Ordnungswidrigkeit kann mit einer Geldbuße von bis zu 50.000,00 € geahndet werden. Zuständige Verwaltungsbehörde für die Verfolgung und Ahndung von Ordnungswidrigkeiten ist der Kanzler/die Kanzlerin der Technischen Universität Dortmund. Im Falle eines mehrfachen oder sonstigen schwerwiegenden Täuschungsversuches kann der Prüfling zudem exmatrikuliert werden, § 63 Abs. 5 Hochschulgesetz NRW.

Die Abgabe einer falschen Versicherung an Eides statt ist strafbar.

Wer vorsätzlich eine falsche Versicherung an Eides statt abgibt, kann mit einer Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft werden, § 156 StGB. Die fahrlässige Abgabe einer falschen Versicherung an Eides statt kann mit einer Freiheitsstrafe bis zu einem Jahr oder Geldstrafe bestraft werden, § 161 StGB.

Die oben stehende Belehrung habe ich zur Kenntnis genommen:

Official notification:

Any person who intentionally breaches any regulation of university examination regulations relating to deception in examination performance is acting improperly. This offence can be punished with a fine of up to EUR 50,000.00. The competent administrative authority for the pursuit and prosecution of offences of this type is the chancellor of the TU Dortmund University. In the case of multiple or other serious attempts at deception, the candidate can also be unenrolled, Section 63, paragraph 5 of the Universities Act of North Rhine-Westphalia.

The submission of a false affidavit is punishable.

Any person who intentionally submits a false affidavit can be punished with a prison sentence of up to three years or a fine, Section 156 of the Criminal Code. The negligent submission of a false affidavit can be punished with a prison sentence of up to one year or a fine, Section 161 of the Criminal Code.

I have taken note of the above official notification.

Ort, Datum
(Place, date)

Unterschrift
(Signature)

Titel der Dissertation:
(Title of the thesis):

Optimal design theory of dose-response experiments in toxicology

Ich versichere hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Titel selbstständig und ohne unzulässige fremde Hilfe angefertigt habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie wörtliche und sinngemäße Zitate kenntlich gemacht.
Die Arbeit hat in gegenwärtiger oder in einer anderen Fassung weder der TU Dortmund noch einer anderen Hochschule im Zusammenhang mit einer staatlichen oder akademischen Prüfung vorgelegen.

I hereby swear that I have completed the present dissertation independently and without inadmissible external support. I have not used any sources or tools other than those indicated and have identified literal and analogous quotations.

The thesis in its current version or another version has not been presented to the TU Dortmund University or another university in connection with a state or academic examination.*

*Please be aware that solely the German version of the affidavit ("Eidesstattliche Versicherung") for the PhD thesis is the official and legally binding version.

Ort, Datum
(Place, date)

Unterschrift
(Signature)