UHASSELT
KNOWLEDGE IN ACTION

Maastricht University

Faculty of Sciences
School for Information Technology
Master of Statistics and Data Science

Master's thesis

Optimization of triple combination therapy experimental design for in-vitro studies

Marie Sebald

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics and Data Science, specialization Biostatistics

SUPERVISOR :

Prof. dr. Olivier THAS

2024
2025

Transnational University Limburg is a unique collaboration of two universities in two countries: the University of Hasselt and Maastricht University.

UHASSELT
KNOWLEDGE IN ACTION

www.uhasselt.be
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

Maastricht University

Faculty of Sciences
School for Information Technology
Master of Statistics and Data Science

Master's thesis

Optimization of triple combination therapy experimental design for in-vitro studies

Marie Sebald

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics and Data Science, specialization Biostatistics

SUPERVISOR :

Prof. dr. Olivier THAS

Contents

List of Tables

List of Figures

Abstract

Background: Multi-drug combinations constitute a promising treatment approach for complex diseases. However, the systematic exploration of interactions among their constituents in in-vitro experiments poses challenges when more than two drugs are combined. In the presented work, three-drug combinations are considered and suitable experimental designs are reviewed and implemented in the context of a simulation study focusing on synergistic interactions.

Methods: Seven experimental designs - among them uniform maximum power designs and variations of ray and checkercube designs - are examined with respect to their performance in seven different monotherapy scenarios. Main performance metrics are sensitivity of the synergy detection and power to detect at least five synergistic points. These are evaluated based on 200 simulation runs. Data is generated under the Bliss independence model and synergies are added to combination points in varying areas of the dose space. Data is analyzed under the Bliss independence model using pointwise deviation assessments between observed effects and expected effects under additivity. The experimental budget is assumed to consist of 384 runs, allowing for up to 96 unique design points and four replicates.

Results: The power to detect at least five synergistic points appeared promising for most designs, with values mostly lying above 90%. The sensitivity varied, with uniform maximum power designs and checkercube designs showing, on average, higher sensitivities than ray designs. A checkercube with four effective concentrations on the axes showed exceptional performance with regard to sensitivity, but was less robust with regard to power, as it sampled in a limited range of the dose space and sometimes missed the simulated synergy area. Scenarios with high slopes generally posed greater difficulties for synergy detection across all designs.

Conclusion: In practical experimentation, it is important to systematically explore as large as possible and as relevant as possible areas in the dose space with a limited number of design points. Few principled experimental designs specifically dedicated to multi-drug combinations exist in the literature, with uniform designs constituting a notable exception. However, conventional approaches, such as checkercube designs, altogether show promise, even when the experimental budget and thus the achievable resolution on the checkercube axes are limited.

1 Introduction

Drug discovery faces numerous difficulties in both pre-clinical and clinical development phases. A primary reason for the eventual failure of many candidate drugs is intolerable toxicity at clinically effective doses (Sun et al., 2022). The emergence of drug resistance poses another obstacle (Makhoba et al., 2020). In the light of these problems, combination therapies consisting of multiple drugs constitute a promising approach towards treatments that are not only more effective, but also safer and less susceptible to the emergence of drug resistance. Intuitively, enhanced efficacy may stem from synergies between drugs in the sense that the constituent drugs in a combination may boost each other's efficacy. This may allow to administer lower doses of the individual drugs, thereby limiting their dose-related toxicities (Calzetta et al., 2024). Reduced drug resistance may follow from different mechanisms of action of the constituent drugs in a combination, making it more challenging for diseases processes to find escape routes (Malik et al., 2020). Multi-drug combinations already represent the standard of care across a spectrum of diseases and are particularly prevalent in oncology (Makhoba et al., 2020; Zimmermann et al., 2007).

Yet, the systematic exploration of multi-drug combinations poses unique challenges, especially during early drug discovery (Zimmermann et al., 2007): First, there is a very large number of combinations of agents that could be explored. Second, interactions between the agents in any given multi-drug combination need to be studied thoroughly but efficiently to detect those combinations, whose constituents do not counteract each other but ideally enhance each other's efficacy, creating synergies. The first problem can be addressed by prioritizing rationally designed drug combinations, whose identification can possibly be aided by artificial intelligence tools (Boshuizen & Peeper, 2020; J. Chen et al., 2025). The second problem can be addressed by studying the prioritized drug combinations based on informative experimental designs, accompanied by suitable statistical analyses, in in-vitro settings. In doing so, it is important to test several dose combinations since the interaction patterns may vary with dose (Calzetta et al., 2024; Jonker et al., 2005).

This master's thesis is concerned with experimental designs for in-vitro studies of combinations of three drugs, subsequently referred to as triple combinations. Given prior knowledge of the constituent drugs' dose-response curves and a limited experimental budget, the central challenge is to select the dose combinations to be tested in an in-vitro experiment in such a way that the power of detecting synergies between the constituent drugs is maximized. The thesis is organized

as follows: Chapter 2 introduces key concepts relevant to drug combination studies, including the definition of the term synergy. Based on these concepts, the research question is further refined. Chapter 3 provides a literature review of experimental designs applicable to triple combination studies. Chapter 4 describes a simulation set-up, in which some of these designs are implemented and evaluated with respect to their performance in scenarios informed by real-world applications in oncology. Chapter 5 summarizes the simulation results, while chapter 6 discusses the strengths and limitations of the simulation-based research and suggests further reasearch venues. Chapter 7 sheds light on ethical considerations before chapter 8 draws conclusions. Of note, in this master's thesis, all concepts and methods are discussed from the perspective of drug combination studies. However, the presented material is equally relevant to other fields, in particular toxicology, where the investigation of multiple chemicals' combined toxicity has gained increased attention throughout the last decade (J. Y. Liu & Sayes, 2024; Martin et al., 2021).

2 From Definitions to Research Questions

Early discovery of single drugs focuses on signals of efficacy; however, early discovery of multi-drug combinations does not only need to account for the overall efficacy of combinations, but also for possible interactions between the constituents (D. Chen et al., 2015; J. Tang et al., 2015). To examine drug interactions, several dose combinations of the agents under consideration are typically first tested in diseased cells. Those combinations whose constituents show synergistic interactions in these in-vitro experiments - implying that the drugs enhance each other's efficacy, ideally across a broad dose range - are then often prioritized for further development (Pemovska et al., 2018). To be able to detect synergistic interactions, it is essential to formalize the concept of synergies and define the reference standard against which their efficacy enhancement is measured.

2.1 Drug Interactions as Deviations from Additivy

Even in the absence of drug interactions, administering a combination of drugs is typically expected to produce a greater effect than any individual drug alone, when applied as a monotherapy at the same dose level present in the combination (Chou, 2006; Greco et al., 1995). Therefore, comparing the combination effect to that of any monotherapy may be a relatively lenient criterion for claiming the presence of interactions. As a result, the most common models for defining drug interactions instead compare the observed combination effect to the expected combination effect under a no-interaction assumption in order to determine whether interactions are present. When the observed combination effect coincides with the expected combination effect under a no-interaction assumption, the effect is said to be additive at the respective dose combination. When the observed combination effect lies above (below) the expected effect, there is a synergistic (antagonistic) interaction between the drugs at the respective dose combination. Drug interactions are thus generally characterized by deviations from a reference model describing the expected additive effect (Calzetta et al., 2024; Foucquier & Guedj, 2015; Greco et al., 1995). This raises the question which reference models — also known as null models, as they represent the null hypothesis of additive action — should be used to assess drug interactions. A simple effect summation model, claiming that the expected additive effect of a combination should equal the sum of the effects of the constituent drugs at their respective dose levels, does typically not constitute a suitable reference model, i.a. because it lacks pharmacological plausibility (Calzetta et al.,

2024; Foucquier & Guedj, 2015). The following sections introduce two more reasonable reference models, Loewe additivity and Bliss independence, which have a long-standing history and constitute the most frequently used null models to date (J. Tang et al., 2015). In line with the focus of this master's thesis, the models are discussed in the context of triple combination therapies.

2.1.1 Loewe Null Model

The Loewe model, promoted by Loewe & Muischnek (1926), is based on the dose equivalence principle, stating that for any dose of one particular agent one can find an equivalent dose of another agent, reaching the same effect. This also implies that - for a given effect (e.g. the observed effect achieved by a certain three-drug combination) - one can always find a dose of any of the constituents reaching the same effect when administered as a monotherapy. Hence, following the dose equivalence principle, one can formally re-express a combination effect in three different ways in the absence of interactions (Lederer et al., 2018, 2019)

$$f_{ABC} \equiv f(x_A, x_B, x_C) = f_A(x_A + f_A^{-1}(f_B(x_B)) + f_A^{-1}(f_C(x_C))) \tag{1}$$

$$= f_B(x_B + f_B^{-1}(f_A(x_A)) + f_B^{-1}(f_C(x_C))) \tag{2}$$

$$= f_C(x_C + f_C^{-1}(f_A(x_A)) + f_C^{-1}(f_B(x_B))) \tag{3}$$

where f_{ABC} denotes the observed combination effect at doses x_A of drug A, x_B of drug B and x_C of drug C, $f_A(\cdot)$, $f_B(\cdot)$, $f_C(\cdot)$ denote monotonic monotherapy dose-response curves and, correspondingly, $f_A^{-1}(\cdot)$, $f_B^{-1}(\cdot)$, $f_C^{-1}(\cdot)$ denote their inverses, determining the dose needed to reach a certain effect. Naturally, this reasoning can only hold for drugs with dose-response curves that share the same minimum and maximum effect (Vlot et al., 2019). Now, additionally assuming that the drugs have a constant relative potency across the entire dose resp. effect range, the two less potent drugs in a triple combination can simply be seen as dilutions of the most potent drug in the combination. If one knows the doses of the monotherapies that reach a certain effect level, any fractional combination of these doses summing up to 1 should then achieve the same effect level (Lederer et al., 2019). Thus, looking at the combination effect f_{ABC} observed at a particular dose combination (x_A, x_B, x_C), under Loewe additivity one would expect that

$$\frac{x_A}{f_A^{-1}(f_{ABC})} + \frac{x_B}{f_B^{-1}(f_{ABC})} + \frac{x_C}{f_C^{-1}(f_{ABC})} = 1 \tag{4}$$

where the denominators always denote the monotherapy doses needed to achieve the combination effect f_{ABC} (Calzetta et al., 2024; Foucquier & Guedj, 2015). Deviations from the value 1 indicate deviations from Loewe additivity. Following the proposal by Berenbaum (1977), one can define the interaction index τ as

$$\tau = \frac{x_A}{f_A^{-1}(f_{ABC})} + \frac{x_B}{f_B^{-1}(f_{ABC})} + \frac{x_C}{f_C^{-1}(f_{ABC})} \begin{cases} < 1 & \rightarrow \text{ synergy} \\ = 1 & \rightarrow \text{ additivity} \\ > 1 & \rightarrow \text{ antagonism} \end{cases} \tag{5}$$

Synergy thus means that the total dose required to reach a certain combination effect f_{ABC} is lower than expected based on the drugs' relative potencies, reflecting the dose-effect based nature of the Loewe framework (Foucquier & Guedj, 2015; Pearson et al., 2023; Tallarida, 2012; Wooten et al., 2021). To assess whether and where interactions exist, one can calculate the interaction index at each combination point. Alternatively, one can proceed graphically and use equation (4) to define a isocontour plane in 3D with drug doses on the axes and the plane defining all dose combinations that should achieve the combination effect f_{ABC} under Loewe additivity. If the actual dose combination (x_A, x_B, x_C) under consideration lies below (above) this plane, this indicates synergy (antagonism). Under the stringent conditions underlying the Loewe additivity model, the isocontour plane is a straight, triangular plane connecting the intercepts defined by the doses $(f_A^{-1}(f_{ABC}), 0, 0)$, $(0, f_B^{-1}(f_{ABC}), 0)$ and $(0, 0, f_C^{-1}(f_{ABC}))$ (Greco et al., 1995; J. Y. Liu & Sayes, 2024; Rider & Simmons, 2018; Tallarida, 2011, 2012).

However, neither an assessment based on the interaction index nor an assessment based on the isocontour plane permits a natural assessment of the strength of interactions (Calzetta et al., 2024; J. Tang et al., 2015). Therefore, it may be preferable to calculate deviations between the observed combination effect and the expected combination effect under Loewe additivity on the effect scale (J. Tang et al., 2015). This requires treating the combination effect f_{ABC} in equation (4) as unknown and solving for it, given the values of (x_A, x_B, x_C) at a particular combination points and the dose-response curves $f_A(\cdot)$, $f_B(\cdot)$, $f_C(\cdot)$. The solution is then the expected combination effect under Loewe additivity, which may be denoted by f_{ABC}^{Loewe} for enhanced clarity. Note that equation (4) is often nonlinear due to nonlinear monotherapy dose-response curves, requiring numerical solution approaches (J. Tang et al., 2015). Alternatively, one can derive f_{ABC}^{Loewe} from equations (1),

(2) or (3), allowing for an analytic solution; the results coincide for drugs with constant relative potencies (Lederer et al., 2018, 2019).

The Loewe additivity model inherently satisfies the sham combination principle, meaning that the combination of several dilutions of the same agent is always declared additive in this framework. From a mechanistic viewpoint, the Loewe model is particularly intuitive for drugs that act with the same mechanism of action, but differ in their potency. Yet, the mechanistic rationale should not be overemphasized as the mechanism of action is often unknown and possibly more complex than any simple reference model can capture; moreover, the goal of identifying interactions precisely consists in rejecting the null hypothesis that the mechanistic implications of the reference model for additivity are true (Lederer et al., 2018; Makariadou et al., 2024; Thas et al., 2022).

The main limitation of the Loewe model is that its validity is restricted to cases in which all agents cover the same effect range and have a constant relative potency (Greco et al., 1992). This keeps the model formulation consistent and mathematically tractable as all dose combinations expected to reach a certain effect can be identified by compensating decreases in one drug's dose by increases in another drug's dose at a constant ratio (Foucquier & Guedj, 2015; Lederer et al., 2018; Tallarida, 2012). Extensions of the Loewe model are complex and mostly limited to two-drug combinations (see e.g. Calzetta et al., 2024 for an overview). Finally, note that Loewe additivity is undetermined when an observed combination effect f_{ABC} exceeds the maximum possible effect of the monotherapies, illustrating that the dose equivalence principle underlying equations (1)-(5) does not only rely on a shared effect range between the monotherapies but also between the monotherapies and the combinations (Pearson et al., 2023; Yadav et al., 2015).

2.1.2 Bliss Null Model

The Bliss model, introduced by Bliss (1939), is a purely effect-based model: It is based on comparing effects at certain monotherapy vs. combination doses rather than comparing monotherapy vs. combination doses needed to achieve certain effects. A further distinction from the Loewe model is the probabilistic nature of the Bliss model, making it suitable for monotherapies whose effect ranges span the interval $[0, 1]$ and can thus be interpreted as probabilities; also combination effects are assumed to be restricted to the interval $[0, 1]$. This is important since the Bliss model assumes that the observed combination effect f_{ABC} is the result of a probabilistic process

in which the three drugs act statistically independently of each other, describing a scenario that is most plausible for drugs with different mechanisms of action. For killing-type responses such as the percentage of cell death, one can rely on the probability of at least one drug causing cells to die and then apply the standard formula for the union of three statistically independent events to obtain the expected additive response at a given dose combination (x_A, x_B, x_C) (Kortenkamp & Altenburger, 2010; L. Zhang & Yang, 2016):

$$\begin{aligned} f_{ABC}^{\text{Bliss}} =& f_A(x_A) + f_B(x_B) + f_C(x_C) - f_A(x_A)\, f_B(x_B) \; - f_A(x_A)\, f_C(x_C) \; - \\ & f_B(x_B)\, f_C(x_C) \; + f_A(x_A)\, f_B(x_B)\, f_C(x_C) \end{aligned} \tag{6}$$

with $f_i(x_i)$ denoting the monotherapy effect of drug i at dose x_i with $i \in \{A, B, C\}$. If the observed combination effect f_{ABC} differs from the one predicted under Bliss independence, this provides evidence of interactions (J. Tang et al., 2015): A higher (lower) combination effect than f_{ABC}^{Bliss} indicates synergy (antagonism) at the respective combination dose. Note that in the case of survival-type responses such as the percentage of cell viability, one can either look at the intersection (instead of the union) of three statistically independent events - here, cells surviving the exposure to the doses of drugs A, B and C in a combination - or transform the responses to killing before applying equation (6) (Greco et al., 1995; Wooten et al., 2021; Yadav et al., 2015).

If the monotherapy responses $f_i(\cdot)$ are not naturally bounded to the 0-1 range or if the monotherapies are partial responders with different maximum effects, suitable rescalings may be required before employing the Bliss independence model, see e.g. Pearson et al. (2023), Thas et al. (2022) or Calzetta et al. (2024): The former two discuss rescalings aimed at maintaining the probabilistic underpinning of the Bliss model intact by rescaling based on each drug's maximum achievable response relative to the maximum achievable response across all three drugs, while the latter suggest to continue using a system's theoretical maximum response as an anchor, even in cases where the monotherapies are partial responders and thus cannot achieve the theoretical maximum response. Whenever the original model based on 0–1 effect ranges or a rescaled version preserving the probabilistic underpinning is used, the Bliss independence model is by definition not applicable in situations where the observed combination effect exceeds the maximum effect attainable by the most effective monotherapy, similar to the Loewe additivity model discussed in the previous section (Wooten et al., 2019, 2021). Yet, unlike the Loewe additivity model, the Bliss independence model does not treat drugs as dilutions of each other and is thus also not inherently sham compliant: Except for particular shapes of the dose-response curves, one may thus

erroneously conclude that there is a deviation from additivity when combining different doses of the same drug in a so-called sham experiment, which supporters of Loewe additivity claim to be logically inconsistent (Calzetta et al., 2024; Foucquier & Guedj, 2015; Greco et al., 1995). Otherrs dismiss the relevance of sham compliance as a main criterion for the sensibility of null models since it is unlikely that one will ever combine drugs with the exact same dose-response profiles in actual combination experiments (Wooten et al., 2021).

2.1.3 The Choice of Null Models

Both the Loewe and Bliss null models have been introduced almost a century ago (Bliss, 1939; Loewe & Muischnek, 1926). In the light of their specific limitations and also their differing perspectives on deviations from additivity, there have been intense discussions whether they constitute suitable reference models for the assessment of drug interactions and, if so, which of the two models should be preferred (Greco et al., 1992; J. Tang et al., 2015). To date, there has been no consensus on either of these questions, leading to "a fragmented landscape" (Meyer et al., 2020) or even "anarchy" (Calzetta et al., 2024) in the field of drug interaction assessment.

Nonetheless, in principle, the majority of researchers seems to recognize that both Loewe and Bliss null models constitute useful empirical reference models, without aspiring to a universally correct mechanistic description of additive action (Greco et al., 1992; J. Tang et al., 2015). In case both null models are applicable, there is a call for applying them in combination and possibly only claiming the presence of interactions when the Loewe and Bliss models indicate consistent deviations from additivity (J. Tang et al., 2015; L. Zhang & Yang, 2016). This accounts for the fact that the conclusions from both models may differ when used on the same data, particularly when monotherapy dose response curves are very steep or very shallow (Greco et al., 1992, 1995; J. Tang et al., 2015; Vlot et al., 2019). A minority of researchers regards both Loewe and Bliss models as inadequate reference models for additive action. Some of them propose their own null models; for a review on recent proposals see e.g. Calzetta et al. (2024). Others, such as Geary (2013), recommend to revert back to the rather lenient highest single agent model, which was originally proposed by Gaddum (1940). Using $\max(f_A(x_A), f_B(x_B), f_C(x_C))$ as criterion, one would claim synergy (antagonism) when the observed combination effect f_{ABC} at a given combination (x_A, x_B, x_C) is above (below) the effect of the most potent monotherapy (Berenbaum, 1989; J.

Tang et al., 2015). Relying on this model also avoids that antagonism is concluded in situations where the combination effect is higher than the highest single agent effect, but lower than the expected f_{ABC}^{Loewe} or f_{ABC}^{Bliss}, in which case antagonism may not be of practical relevance (J. Tang et al., 2015; Yadav et al., 2015).

Beyond discrepancies, all null models discussed here share certain features: Their conclusions are inherently specific to certain dose combinations, recognizing that some dose combinations may be additive, while others may be synergistic or antagonistic (Greco et al., 1992, 1995; Harbron, 2010). Moreover, they all rely on monotherapy data when deriving the expected additive effect. Importantly, for an unbiased comparison between the expected and observed combination effect, monotherapy data should always be measured under the same conditions as combination data, ideally in the same experiment (Chou, 2008). Regarding experimental design, it must be kept in mind that the null models do not allow to identify from which particular constituents in a multi-drug combination the interactions originate, unless one also samples and tests all subcombinations (Chou, 2008; Foucquier & Guedj, 2015; Tekin et al., 2017). Additionally, the statistical significance of interactions cannot be assessed, unless one includes replicates of the combination points in the experiment. This is of vital importance since in-vitro experiments exhibit a non-negligible amount of noise, but still frequently ignored. In both the Loewe and Bliss frameworks, adequate significance testing may be confronted with practical challenges (Calzetta et al., 2024; Foucquier & Guedj, 2015; Rønneberg et al., 2021; L. Zhang & Yang, 2016), making the use of bootstrapping an attractive option, especially when assessing deviations from additive expectations in nonparametric contexts (Makariadou et al., 2024; Mao & Guo, 2023; Thas et al., 2022).

2.2 Research Question: Detecting Synergies in Triple Combinations

This master's thesis is concerned with reviewing and evaluating experimental designs for in-vitro experiments aimed at uncovering synergistic interactions among the constituents of triple combination therapies. As a point of departure, it is assumed that previous experimentation has allowed to estimate the dose-response curves $f_i(\cdot)$ for all three monotherapies. The central question is then how to best leverage the existing information on monotherapies when designing a triple combination experiment. This question is considered within the context of early drug discovery, where an efficient screening for synergistic interactions is particularly important.

In the first stage, a literature review on experimental designs applicable to triple combination therapies is conducted. In the second stage, the performance of selected experimental designs is assessed through simulation studies. These simulations concentrate on the implementation and evaluation of the chosen designs using hypothetical examples, which are considered representative for actual three-drug combinations that could be encountered in the field of oncology. The most important criteria with regard to which the designs are evaluated are their sensitivity and their power to detect a certain threshold number of synergistic points.

The response of interest is the percentage killing of tumor cells when exposed to different drug doses in suitable in-vitro assays (Vlot et al., 2019; Zoli et al., 2001). The experimental budget is assumed to be limited to 384 runs, including replicates. This constraint is driven by the standard format of well plates used in in-vitro experiments: Standard plates contain either 96 or 384 wells, each of which is covered by tumor cells and can be exposed to different drug doses in our setting. Using one 384-well plate or four 96-well plates e.g. allows to study 96 unique dose combinations, tested in four replicates to assess variability in the response. This seems to be a sensible size of an experiment, balancing necessary resources and achievable accuracy of estimates (Makariadou et al., 2024; Thas et al., 2022). In this setting, the task is to rationally select up to 96 unique dose combinations (x_A^j, x_B^j, x_C^j) with $j \in \{1, ..., 96\}$ for in-vitro testing such that the chance of detecting synergies - if present - is maximized.

Of note, we explicitly do not attempt to identify the most effective triple combination therapy out of a larger pool of candidates, neither regarding the selection of constituent drugs nor regarding the selection of doses to be applied. In contrast, we analyze up to 96 combinations of three specific drugs A, B, C and attempt to find out whether they exhibit synergistic interactions and, if so, at which dose combinations. In doing so, priority is given to detecting net interactions between the three drugs with one-step designs; explaining how the net interactions originate among the constituents is left to follow-up experiments. For brevity, we generally refer to drugs and doses rather than compounds and concentrations, despite considering an in-vitro setting.

3 Literature Review: Designs for Triple Combination Experiments

This chapter provides a literature review, describing experimental designs for investigating interactions in three-drug combinations in general terms. To begin with, it highlights why optimal experimental designs are rarely feasible and more pragmatic design approaches are needed.

3.1 Challenges in Optimizing Experimental Designs for Drug Combinations

In principle, any experimental design should be carefully constructed to best suit the intended data analysis. Yet, for drug combination studies, there are not only several possible null models defining additive action; there are also several possible data analysis approaches based on these null models. The latter range from locally assessing the deviation between the observed effect and expected effect under additivity at each design point to fitting a response surface model over the entire design space and thereby facilitating interpolations between design points (Gennings, 1996). Response surface models may consider different types of responses (H. B. Fang et al., 2017; Greco et al., 1995; Lederer et al., 2019; L. Zhang & Yang, 2016): Some approaches focus on modeling the dose-effect surface, while incorporating interaction parameters; other approaches directly model deviations from the additive effect surface or an interaction index surface. Embedded in these models, interactions may still be defined in analogy to Loewe or - less frequently - Bliss criteria. Irrespective of the details, all response surface approaches share common challenges: When a response surface is modeled as a function of K drugs, this generally results in a K+1 dimensional surface with K dimensions for the drug doses and one dimension for the response, prohibiting visualizations for more than two drugs and often requiring a rather extensive grid of design points to allow reliable fitting of complex parametric or nonparametric models.

A noteworthy exception is the case of a fixed ratio combination where the total drug dose is escalated while the mixture proportion remains constant. In this special case, one may regard the fixed ratio combination of K drugs as a new drug and fit a conventional two-dimensional dose response curve to it (Straetemans et al., 2005). Still, even in this case, experimental design is challenged by nonlinearities: As monotherapies commonly have nonlinear dose response curves, drug combinations must be expected to follow nonlinear patterns as well. If one were to use optimal design theory, one would therefore need to specify both a suitable parametric form and suitable param-

eter values in order to derive optimal designs based on common optimality criteria such as c- or d-optimality (Hather et al., 2013; Holland-Letz & Kopp-Schneider, 2021; Rider & Simmons, 2018). Both specifications should be close to the truth, as optimal designs are generally sensitive to the assumptions under which they were derived: If the actual system behaves very differently than assumed, supposedly optimal designs may perform even worse than commonly used pragmatic designs (Hather et al., 2013). Altogether, one may may conclude that there is rarely enough prior information to implement optimal experimental designs, especially when multi-drug combinations are investigated during early drug discovery[1]. The following discussion therefore largely focuses on pragmatic approaches and reviews variants of commonly used designs such as checkercube or ray designs. An emphasis is also placed on space-filling designs, which acknowledge that there is limited information about where to expect synergies and, consequently, attempt to uniformly cover relevant segments of the design space.

3.2 Checkercube Design

Given that optimal designs based on conventional optimality criteria are not applicable here, one has to resort to pragmatic experimental designs to explore the design space, which is typically taken to be a rectangle for two-drug combinations or a cube for three-drug combinations when no restrictions on combinations are enforced. The edges of the rectangle resp. cube are determined by the drugs' activity ranges. One comprehensive experimental design for examining drug interactions throughout the entire design space is the checkerboard (K=2 drugs) and, by extension, checkercube (K=3 drugs) design (Prichard & Shipman, 1990).

This full-factorial design approach involves selecting D dose levels for each drug and testing all D^K possible combinations (Gennings, 1996). Typically, five to ten dose levels per drug are chosen to ensure a broad exploration of different combinations in the design space and enable both monotherapy dose-response modeling and potential response surface analysis (Rider & Simmons, 2018; Stein et al., 2015). To effectively cover the full activity range of each drug, a practical approach is to use logarithmically spaced serial dilutions, usually including a zero or inactive dose to allow

[1]For fixed ratio combinations of two drugs there are, however, some optimal design approaches, partially building on simplifying assumptions. See e.g. the work of Hather et al. (2013), Sperrin et al. (2015), Holland-Letz et al. (2020) and Holland-Letz & Kopp-Schneider (2021).

assessing lower-order interactions and establish baseline effects (Gennings, 1996; Gonzales et al., 2015). This design is straightforward to implement and can explore a wide range of dose combinations with different dose ratios. Yet, for more than two drugs and reasonably high resolution (e.g., seven to eight dose levels per drug), in practice it often exceeds the experimental budget, especially when considering the need to replicate observations in order to assess the variability in the response. In essence, an exponential increase in the number of design points with the number of drugs leads to a curse of dimensionality (Cokol et al., 2017). Additionally, from a synergy detection perspective, the checkerboard resp. checkercube design allocates many runs to high-dose combinations where monotherapies may already produce near-maximal responses, limiting the potential to detect synergistic effects at these points (Martinez-Irujo et al., 1996). To address this, one could conduct high resolution testing for monotherapies while restricting combination testing to lower dose levels, particularly when possibly undetected antagonisms at high dose levels do not constitute a concern (Harbron, 2010). Further limitations are that the checkerboard resp. checkercube design offers little room for optimizing the design (Holland-Letz & Kopp-Schneider, 2021), apart from selecting suitable dilution factors for monotherapies (Donev & Tobias, 2011). In this regard, there exists no definitive guidance in the context of drug interaction assessments, since an optimal selection would require information about the dose regions where interactions are expected. Modifying the traditional approach based on serial dilutions, C. Chen et al. (2018) suggest to use effective concentrations (EC) - e.g. doses reaching $0, 20, 50, 80\%$ of the maximal effect - as factor levels on the checkerboard/-cube's monotherapy axes, especially when considering a low-resolution design such as a 4x4(x4) design. In simulations within a pharmacodynamic context, this approach yielded a comparable performance in terms of correct classifications to a 10x10x(10) high-resolution design.

3.3 Ray Design

To reduce the experimental burden when studying interactions among more than two drugs, the fixed ratio ray design is a widely adopted strategy (Berenbaum, 1978; Chou & Talalay, 1984; Tallarida et al., 1997). This approach involves testing combinations of drugs with a fixed dose ratio, such as a 1:1:1 ratio in a simple three-drug example, at several different total doses. Consequently, one samples dose combinations along a line or "ray" in the design space, starting from the origin and

extending in the direction dictated by the chosen dose ratio. It is customary to explore several dose ratios and thus construct several rays. Along each ray, the relative proportions of the drugs - i.e. the mixture proportions - remain fixed, while the total dose varies, e.g. according to logarithmically spaced dilutions with D=5-10 different total doses tested (Mayer & Janoff, 2007; Nørgaard & Cedergreen, 2010; White et al., 2005). Monotherapies can be treated like rays for which the relative proportions of the other drugs have been fixed at zero. Altogether, for R rays, the ray design results in $R \times D$ design points, a considerably lower number than in the checkercube design. Moreover, owing to the fixed mixture proportions, each fixed ratio mixture can be interpreted like a new drug (Foucquier & Guedj, 2015; Gennings, 1996; Straetemans et al., 2005). Since several doses of the respective mixtures are sampled, dose-response curves can be fitted.

The existence of dose-response curves for the fixed ratio mixtures and the monotherapies lends itself to a Loewe-based interaction assessment: One may use the interaction index or the isocontour plane corresponding to an effect level of interest to determine whether a lower combination dose can achieve this effect level than predicted under additivity (Berenbaum, 1978; D. Chen et al., 2015; Gennings, 1996; Tallarida, 2012). This analysis may be embedded in a parametric inferential framework, e.g. fitting dose-response curves for all rays simultaneously and treating the interaction index as a ray-specific parameter, thereby increasing efficiency (Fan & Zhang, 2014; Straetemans et al., 2005; Straetemans & Bijnens, 2010). Yet, pointwise assessments of deviations from additivity remain possible and may be necessary if a ray does not show a consistent interaction pattern, implying that interactions do not only depend on the dose ratio but also on the dose level (Fan & Zhang, 2014; H. B. Fang et al., 2010; Harbron, 2010). Response surface analyses are only possible if a sufficient number of rays with good coverage of the design space was sampled (Foucquier & Guedj, 2015; Ning et al., 2014; White et al., 2005). Otherwise the conclusions are specific to the tested dose ratios and should not be inter- or extrapolated (Gennings, 1996). This illustrates the advantages and disadvantages of ray designs compared to checkercube designs beyond the run number: The latter explore the design space more thoroughly by analyzing a wider range of dose ratios but they do not contribute as much information on each dose ratio as the former (Rider & Simmons, 2018), e.g. not allowing to fit dose-response curves (except for the diagonal of the checkercube, which corresponds to a ray if identical dilution factors are used for all monotherapies, see Chou, 2014). The extensive information on mixtures provided by ray designs is often perceived as useful when moving from in-vitro to in-vivo settings (Mayer & Janoff, 2007).

For the implementation of ray designs for a given number of rays R and points D per ray, two aspects are important: First, the specification of the dose ratios and, second, the specification of the total doses to be tested. The ratios cannot be optimized without information about where interactions are expected (Almohaimeed & Donev, 2014); the total doses can be chosen D-optimally in simple two-drug settings using a parametric framework based on monotherapy information and either the assumption of additivity or the assumption of specific interaction patterns (Almohaimeed & Donev, 2014; Holland-Letz et al., 2020; Holland-Letz & Kopp-Schneider, 2021; Sperrin et al., 2015). Yet, robustness to the assumptions remains a concern and there is currently no extension to three-drug settings. The following subsections therefore discuss pragmatic approaches.

3.3.1 Equipotent Ray, Equipartition Ray and Ray Contour Designs

A common strategy for defining dose ratios in ray designs is to express them in terms of effective concentrations, with the EC50 being a particularly common reference point (Berenbaum, 1978; Chou & Talalay, 1984; Cokol-Cakmak et al., 2018; Straetemans et al., 2005). The EC50, defined as the dose producing a response halfway between a drug's baseline and maximum effect, is a critical measure of drug potency. It is explicitly parameterized in sigmoidal dose-response models, such as the four-parameter logistic regression model, making it both a biologically relevant and a convenient criterion for specifying dose ratios of interest. Often, one attempts to construct equipotent rays by selecting a 1:1:1 ratio based on the drugs' EC50 values, translating into a $\text{EC50}_A : \text{EC50}_B : \text{EC50}_C$ ratio in a concentration metric such as $\mu g/l$ (Chou, 2014; Holland-Letz & Kopp-Schneider, 2021; Ritz et al., 2021). By specifying the ratios in terms of EC50-equivalent units, the contribution of the different drugs to the combination effect shall be balanced. The equipotency argument for this fixed ratio combination, however, only holds when drugs cover the same effect range such that the EC50 corresponds to the same absolute effect level. Even then, true equipotency may only hold at the EC50 values: Moving away from this reference point may result in potency differences due to e.g. different slopes of the monotherapies, even when the dose ratio is maintained at the $\text{EC50}_A : \text{EC50}_B : \text{EC50}_C$ ratio along a ray.

Nonetheless, the idea of constructing an (approximately) "equipotent" ray based on the EC50 values is frequently applied in practice. Berenbaum (1978) and Cokol-Cakmak et al. (2018) motivate this approach geometrically from a perspective of Loewe additivity, claiming that going through

the center of the isobole (two-drug case) resp. isocontour plane (three-drug case) corresponding to the half-maximal effect yields maximum information about any possible curvature of this line resp. plane, which would in turn indicate deviations from additivity. Yet, this argument appears problematic for two reasons: First, the isobole resp. isocontour plane may even be curved under additivity if the assumption of constant relative potencies made in the Loewe framework does not hold (Lederer et al., 2018, 2019; Tallarida, 2011, 2012). Second, as e.g. noted by Almohaimeed & Donev (2014) or Brennan et al. (2022), the optimal dose ratio for detecting drug interactions always depends on the true interaction pattern. Specifically, interactions may or may not manifest at a 1:1:1 ratio based on the EC50 values; if they do not manifest at this ratio, this does not mean that there are no interactions at other dose ratios. Therefore, rays following several different dose ratios should be sampled. Ignoring the technicalities, one may take a simplified approach for specifying a range of dose ratios of interest, as suggested by Berenbaum (1978) and formalized by Dou et al. (2011) under the term "equipartition ray design": One may construct rays by extending lines from the origin through equidistant anchor points on an (approximate) "isobole" resp. "isocontour plane", with the latter obtained by connecting the drugs' EC50 values, even if this may not always yield an accurate isobole or isocontour plane. In the three-drug setting, one may then take inspiration from a simplex lattice design when choosing mixture proportions evenly spread throughout the resulting plane (Berenbaum, 1978; Berenbaum et al., 1983; Tallarida, 2012; White et al., 2005). This approach may allow a systematic and pragmatic exploration of several non-equipotent mixing ratios, in addition to an "equipotent" mixing ratio in the plane's center.

Equipotent as well as equipartition ray designs discussed so far mostly concern the selection of dose ratios for rays; the placement of total doses on the rays is then often achieved by e.g. twofold serial dilutions, starting at a certain maximum dose and decreasing dose from there, either based on an actual drug dose or EC50 equivalent units (D. Chen et al., 2015; Chou, 2008, 2014; Nørgaard & Cedergreen, 2010). Almohaimeed & Donev (2014), in contrast, do not make a proposal for selecting dose ratios but for placing total doses on R rays of interest. They suggest to focus on particular effect levels and then compute the total doses corresponding to these effect levels under the assumption of additivity for all rays of interest. Hence, each ray ultimately has total doses placed at points corresponding to the same effect levels, e.g. $20, 30, 40, 50, 60, 70, 80\%$ effect either in absolute terms or in relative terms if there is a common maximum effect. In terms of analysis, one can make use of the fact that the points belonging to the same contour (i.e. to

the same effect level under additivity), but placed on different rays are expected to achieve the same effect and also the same interaction index value; one can thus test for deviations of either the observed effects or the interaction index values from their reference value (the additive effect resp. an interaction index of 1) at each effect level in an ANOVA setting. This by default yields conclusions about interactions at several effect levels rather than focusing on one particular effect level, as frequently done in the context of equipotent or equipartition rays. It may be particularly useful when only parts of the effect range are of interest and/or e.g. either low or high doses shall be avoided, thereby making it more difficult to use analyses based on modeling rays' entire dose response curves (Almohaimeed & Donev, 2014). The so called ray-contour design was originally introduced in a parametric two-drug setting based on a five-parameter sigmoidal dose-response curve for responses with distributions belonging to the exponential family under a Loewe additivity framework. Yet, it appears easily extensible to the Bliss independence framework, which simplifies the identification of total doses needed to achieve certain effect levels in a fixed dose ratio setting with three drugs, irrespective of the shape of the underlying dose response curve. Of note, while the number of effect levels of interest may be reduced to save experimental resources, it is important to include a sufficiently high number of rays in the ray-contour approach; sample size calculations can be conducted based on the ANOVA model (Almohaimeed & Donev, 2014). As an alternative to the ray-contour approach for placing total doses on the rays, Almohaimeed & Donev (2014) also discuss optimal designs and make proposals for two-drug settings, incorporating pseudo-Bayesian D-optimal designs where the value of the optimality criterion is integrated over a prior distribution for the model parameters. Unfortunately, extensions of this approach to ray designs with more than two drugs do not appear to be readily available.

3.3.2 Uniform Ray Design

Uniform ray designs attempt to account for two main considerations in the context of ray designs (S. S. Liu et al., 2016; Y. H. Zhang et al., 2008, 2010): First, several mixture ratios should be examined. Second, the selection of specific mixture ratios is largely arbitrary in the absence of information on the expected location of interactions. In result, a space-filling approach appears desirable to reduce the risk of missing interactions. The construction of uniform ray designs is based on uniform design theory as developed by K.-T. Fang et al. (2000) and reviewed in K.-T. Fang et al. (2018).

Using a uniform design table, often referred to as U-type table, one can uniformly select factor-level combinations from a certain factor-level table of interest[2]. This results in combinations which are approximately uniformly distributed in the K-dimensional factor-level space. In the context of uniform ray designs, one can select R such combinations from the factor-level space, possibly containing effective concentrations of the K=3 drugs; these combinations can then be expanded into R rays with several total doses on each ray. Effectively, each sampled combination point determines a set of mixture proportions, which characterizes the direction of a particular ray in the design space (S. S. Liu et al., 2016; Y. H. Zhang et al., 2010).

When implementing uniform ray designs, one typically resorts to existing $U_n(q^m)$ tables developed for a particular number n of experiments involving m factors with q levels where $q \leq n$ as the generation of such U-type tables is laborious (K.-T. Fang et al., 2018; S. S. Liu et al., 2016). These tables have m columns, each of which is filled with permutations of numbers in $\{1, ..., q\}$. Their n rows determine uniformly spread combination points: When applied to a specific factor-level table of interest, a row (a, b, c) in the U-type table would indicate a combination with factor 1 at its a-th level, factor 2 at its b-th level and factor 3 at its c-th level with $a, b, c \in \{1, ..., q\}$ (K.-T. Fang et al., 2018; S. S. Liu et al., 2016). In the context of uniform ray designs for drug combination experiments, $n \equiv$ R corresponds to the number of rays while $m \equiv$ K refers to the number of drugs involved and q refers to dose levels of interest, e.g. effective concentrations. One would then match each row of the uniform table to the factor-level table and compute the ray-specific mixture proportions for the drugs A, B and C as $p_A = \frac{x_{A,a}}{x_{A,a}+x_{B,b}+x_{C,c}}$, $p_B = \frac{x_{B,b}}{x_{A,a}+x_{B,b}+x_{C,c}}$, $p_B = \frac{x_{C,c}}{x_{A,a}+x_{B,b}+x_{C,c}}$ (S. S. Liu et al., 2016; H. X. Tang et al., 2016). When no U-type table is available for the specific m and q of interest, one may look for a larger U-type table and select the best possible subset of columns. One may also merge adjacent factor levels together if needed (S. S. Liu et al., 2016).

While the number of factor levels is partially determined by available U-type tables, there exists no clear guidance on how to best design the underlying factor-level table when creating uniform rays. Yet, generally, it appears desirable to choose factor levels such that the mixtures constructed based on the U-type table cover truly different mixture ratios; in this regard, factor levels that are too close to each other may be counterproductive (S. S. Liu et al., 2016). At the same time, extreme mixture

[2]U-type tables are created by minimizing a discrepancy measure reflecting the deviation from a uniform distribution based on number-theoretic methods, monte-carlo methods or stochastic search algorithms. For moderate sample sizes, they achieve notably better uniformity than random sampling from a uniform distribution (K.-T. Fang et al., 2018).

ratios (where one component dominates the mixture effect while other components are present at almost ineffective concentrations), may also be undesirable in many contexts. This illustrates that uniformity as a design criterion in the construction of rays does not obliterate the need to carefully consider the experimental context. Moreover, working with an explicit factor-level table may have practical benefits in terms of the titration of the required drug doses, but it may limit the uniformity of the mixture ratios. Another approach may therefore be to only implicitly work with factor levels: Following the ideas of general uniform design theory, one may use a uniform design table with as many factor levels as rays to be constructed and - instead of equating each row in the table to a factor-level combination - translate each row to a point in a [0,1] space of the required dimensionality. This point can then be scaled according to the range of the design space of interest, which is particularly simple for rectangular design spaces, and leads to a uniform coverage of said space (K.-T. Fang et al., 2000, 2018). Ultimately, each point's coordinates again define a set of mixture proportions or, equivalently, a direction for a ray in the design space. The downside of this approach is that it only takes the maximum doses into account for a potency alignment, thereby possibly resulting in more extreme mixtures. As in any ray design, if one aims at determining which component in a mixture is responsible for the detected interaction, one must construct rays with subsets of the factors (e.g. investigate rays for all pairs of drugs in a triple combination while maintaining the ray ratios analogous to the triple combination rays, apart from setting one component to zero). An example of such an up-to-down testing approach in the context of uniform ray designs is provided by J. Zhang et al. (2014).

3.4 Uniform Maximum Power Design

This section introduces another type of uniform design, often referred to as uniform maximum power design, which is available for two- and three-drug combinations (Tan et al., 2003, 2009). The idea is similar to the uniform ray design: In the absence of knowledge where to expect synergy, design points should be uniformly placed in a sensible experimental domain to efficiently cover said domain. However, the mechanics underlying the design are noteably more involved. To begin with, it is important to clarify that design points are not simply placed uniformly throughout the cube-shaped dose space in a three-drug application. Instead, the uniform maximum-power design is embedded in the framework of Loewe additivity and builds on a model describing the

additive effect of the drugs A, B, C. If this model can be brought into an additive structure involving three linearly independent functions $g(z_1)$, $g(z_2)$ and $g(z_3)$, with the variables z_1, z_2 and z_3 being restricted to a domain whose shape allows uniform sampling, then uniformly sampling points (z_1, z_2, z_3) on this transformed domain is optimal in the sense that it maximizes the minimum power of a lack-of-fit test for detecting departures from additivity over a very broad class of possible departures. This property was proven by Wiens (1991) in the context of general approximately linear regression problems and exploited in the context of two- and three-drug combination studies in seminal works by Tan et al. (2003), H. B. Fang et al. (2008), H. B. Fang et al. (2009), Tan et al. (2009) and H. B. Fang et al. (2017). Owing to the complexity of the approach, the subsequent paragraphs provide technical details based on Tan et al. (2009).

First, within the framework of Loewe additivity, the dose equivalence principle can be applied and the observed effect of any combination dose (x_A, x_B, x_C) can be re-expressed

$$f_{ABC}(x_A, x_B, x_C) = f_A(X_A) = f_B(X_B) = f_C(X_C) \tag{7}$$

with X_A, X_B and X_C denoting isoeffective monotherapy doses, reaching the combination effect $f_{ABC}(x_A, x_B, x_C)$. Now, consider drug A as the reference drug: Under additivity, the drug A equivalent of a combination dose (x_A, x_B, x_C) is determined by the potencies of drugs B and C relative to drug A. At any particular combination, it is expected that

$$f_{ABC}(x_A, x_B, x_C) = f_A\left(x_A + \frac{X_A}{X_B}x_B + \frac{X_A}{X_C}x_C\right) \tag{8}$$

i.e., the monotherapy dose of drug A that yields an equivalent effect as the combination dose (x_A, x_B, x_C) amounts to $X_A = x_A + \frac{X_A}{X_B}x_B + \frac{X_A}{X_C}x_C$ under Loewe additivity, with X_A, X_B and X_C implicitly depending on the observed combination effect $f_{ABC}(x_A, x_B, x_C)$. This shall be generalized to any possible combination effect, resulting in a regression model

$$y \equiv f_{ABC}(x_A, x_B, x_C) = f_A(x_A + \rho(X_B)x_B + \rho(X_C)x_C) \tag{9}$$

for the additive effect with $\rho(X_i)$ being an implicit function of (x_A, x_B, x_C), expressing the potency of drug $i \in \{\text{B,C}\}$ relative to drug A. In a semi-parametric setting, one can then test for lack-of-fit of this additive regression model against a general alternative model

$$y = f_A(x_A + \rho(X_B)x_B + \rho(X_C)x_C) + f(x_A, x_B, x_C) + \epsilon \tag{10}$$

where $f(x_A, x_B, x_C)$ captures any systematic deviation from additivity of an unspecified, possibly non-parametric form $f(\cdot)$ and ϵ captures a random error term, assumed to be normally distributed with mean 0 and variance σ^2. If the additive model (9) were to exhibit a lack-of-fit based on a classical F-test comparing the model-based error to the pure error, this would provide evidence for $f \neq 0$ and thus indicate the presence of interactions between drugs A, B and C. Yet, from a design perspective, the power of this lack-of-fit test is optimized if and only if the following two conditions can be met (Tan et al., 2003, 2009; Wiens, 1991)

- the additive model can be expressed in a generalized additive structure as

$$\begin{aligned} y =& f_A(x_A + \rho(X_B)x_B + \rho(X_C)x_C) + f(x_A, x_B, x_C) \approx \\ & \alpha_1 g_1(z_1) + \alpha_2 g_2(z_2) + \alpha_3 g_3(z_3) \end{aligned} \tag{11}$$

  with (z_1, z_2, z_3) derived from a one-to one invertible transformation $(x_A, x_B, x_C) \in S_0 \rightarrow (z_1, z_2, z_3) \in S$ from the the original cube-shaped dose space S_0 to a transformed design space S, assuming the form $z_i = \varphi_i(x_A, x_B, x_C)$ for $i = 1, 2, 3$. Here, $g_1(\cdot), g_2(\cdot), g_3(\cdot)$ are linearly independent functions and α_i are known parameters. The general alternative model then corresponds to

$$y \approx \alpha_1 g_1(z_1) + \alpha_2 g_2(z_2) + \alpha_3 g_3(z_3) + g(z_1, z_2, z_3) + \epsilon \text{ for } (z_1, z_2, z_3) \in S \tag{12}$$

  with $g(z_1, z_2, z_3) = f(x_A, x_B, x_C)$ assumed orthogonal to $(g_1(z_1), g_2(z_2), g_3(z_3))$ but otherwise unspecified. The lack-of-fit test then corresponds to testing $H_0 : g = 0$.

- Uniform combination points (z_1^j, z_2^j, z_3^j) can be sampled on S. Making use of the one-to-one invertible transformation, the uniformly sampled points (z_1^j, z_2^j, z_3^j) can ultimately be transformed into dose combinations (x_A^j, x_B^j, x_C^j) to be tested in the experiment.

This approach does not require parametric assumptions about the nature of the interactions and allows sample size calculations based on a non-central F-distribution with non-centrality parameter $\delta = \frac{N}{\sigma^2}\eta^2$ where N is the total run number including replicates and $\eta^2 = \int_S g^2(z_1, z_2, z_3) dz_1 dz_2 dz_3$ is the assumed effect size. Following Tan et al. (2009), the

approach is illustrated for the case of monotherapies with log-linear dose-response curves

$$\begin{aligned} f_A(X_A) &= \alpha_A + \beta_A \ \log(X_A) \\ f_B(X_B) &= \alpha_B + \beta_B \ \log(X_B) \\ f_C(X_C) &= \alpha_C + \beta_C \ \log(X_C) \end{aligned} \tag{13}$$

where, without loss of generality, it is subsequently assumed that $\beta_C \leq \beta_B \leq \beta_A$ in the context of decreasing dose-response curves. For better distinction, upper-case X_i is again used to clearly indicate a monotherapy dose of drug $i \in \{A, B, C\}$. The potencies of drugs B and C relative to drug A can then be expressed for isoeffective doses X_A and X_B resp. X_A and X_C as

$$\rho(X_B) = \frac{X_A}{X_B} = \frac{f_A^{-1}(f_B(X_B))}{X_B} = \frac{\exp\left(\frac{\alpha_B + \beta_B \log(X_B) - \alpha_A}{\beta_A}\right)}{X_B} =$$

$$= \exp\left(\frac{\alpha_B - \alpha_A}{\beta_A}\right) X_B^{\frac{\beta_B}{\beta_A}-1} = \rho_0 X_B^{\frac{\beta_B}{\beta_A}-1} \quad \text{with} \quad \rho_0 = \exp\left(\frac{\alpha_B - \alpha_A}{\beta_A}\right) \tag{14a}$$

$$\rho(X_C) = \frac{X_A}{X_C} = \rho_1 X_C^{\frac{\beta_C}{\beta_A}-1} \quad \text{with} \quad \rho_1 = \exp\left(\frac{\alpha_C - \alpha_A}{\beta_A}\right) \tag{14b}$$

These relative potencies are constant when $\beta_A = \beta_B = \beta_C$, otherwise they vary with dose resp. effect levels. For constant relative potencies, the additive model can be brought into the desired structure with three separate logarithmic expressions by simple algebraic manipulations:

$$\begin{aligned} y &= \alpha_A + \beta_A \log(x_A + \rho_0 x_B + \rho_1 x_C) = \\ &= \alpha_A + \beta_A \log(x_A + x_B + x_C) + \beta_A \log\left(\frac{x_A + \rho_0 x_B + \rho_1 x_C}{x_A + x_B + x_C}\right) = \\ &= \alpha_A + \beta_A \log(x_A + x_B + x_C) + \\ &\qquad \beta_A \log\left(\frac{x_A}{x_A + x_B + x_C} \cdot \frac{x_A + x_B + \frac{\rho_1}{\rho_0} x_C}{x_A + x_B + \frac{\rho_1}{\rho_0} x_C} + \frac{\rho_0 x_B + \rho_1 x_C}{x_A + x_B + x_C}\right) = \\ &= \alpha_A + \beta_A \log(z_1) + \beta_A \log((1 - \rho_0) z_2 + \rho_0) + \beta_A \log\left(\left(1 - \frac{\rho_1}{\rho_0}\right)(1 - z_3) + \frac{\rho_1}{\rho_0}\right) \end{aligned} \tag{15}$$

with

$$
\begin{aligned}
z_1 &= x_A + x_B + x_C \\
z_2 &= \frac{x_A}{x_A + x_B + \frac{\rho_1}{\rho_0} x_C} \\
z_3 &= \frac{x_C}{x_A + x_B + x_C}
\end{aligned}
\tag{16}
$$

This specific structure allows to sample z_1 independently of z_2 and z_3 since the former relates to a sum while the latter relate to fractions bounded between $(0, 1)$. It is useful for generating uniform points (z_1^j, z_2^j, z_3^j) in a domain $S = \{(z_1, z_2, z_3) : Z_L < z_1 < Z_H, (z_2, z_3) \in \mathcal{V}_2\}$ with $\mathcal{V}_2 = \{(z_2, z_3) : z_i > 0 \, \forall \, i = \{2, 3\}, \sum_{i=2}^{i=3} < 1\}$. Here, Z_L and Z_H can be taken as lower and upper limts of the total dose according to drug A, allowing to restrict attention to a particular effect range of interest under additivity (e.g. by choosing Z_L and Z_H such that $20 - 80\%$ effect are reached based on drug A's dose-response curve). The structure of $\mathcal{V}_2$ then ensures that all components - i.e. drugs A, B and C - are present in the resulting mixtures. Geometrically, the domain S has the shape of a triangular prism in the space formed by z_1, z_2 and z_3. A simple algorithm for uniformly sampling in this domain is provided by Tan et al. (2009). In the case with constant relative potencies, the additive model has the above closed form expression, which allows for an analytic derivation of x_A, x_B and x_C by the inverse transformations of system (16). In the case with non-constant relative potencies, the additive model lacks a closed-form expression. Yet, the general expression (10) can be reformulated by isolating the dose-dependent components of the relative potency terms into a distinct term to the greatest extent possible:

$$
\begin{aligned}
y &= \alpha_A + \beta_A \log(x_A + \rho(X_B) x_B + \rho(X_C) x_C) = \\
&= \alpha_A + \beta_A \log\left(x_A + \rho_0 X_B^{\frac{\beta_B}{\beta_A} - 1} x_B + \rho_1 X_C^{\frac{\beta_C}{\beta_A} - 1} x_C \right) = \\
&= \alpha_A + \beta_A \log\left(x_A + \rho_0 X_B^{\frac{\beta_B}{\beta_A} - 1} x_B + \rho_1 \psi x_C \right) = \\
&= \alpha_A + \beta_A \log\left(x_A + \rho_0^{\frac{\beta_B}{\beta_A}} \rho_1^{1 - \frac{\beta_B}{\beta_A}} \psi^{\frac{\beta_C(\beta_B - \beta_A)}{\beta_B(\beta_C - \beta_A)}} x_B + \rho_1 \psi x_C \right)
\end{aligned}
\tag{17}
$$

with the function $\psi = \psi(x_A, x_B, x_C) = X_C^{\frac{\beta_C}{\beta_A} - 1}$ depending on (x_A, x_B, x_C) through X_C, with the latter again denoting the monotherapy dose of drug C that is as effective as a

particular combination dose (x_A, x_B, x_C). Notice that X_B is also an isoeffective monotherapy dose of drug B such that $\alpha_B + \beta_B \log(X_B) = \alpha_C + \beta_C \log(X_C)$, thus leading to $X_B = \exp^{\frac{\alpha_C - \alpha_B}{\beta_B}} \psi^{\frac{\beta_A \beta_C}{\beta_B(\beta_C - \beta_A)}}$ and eventually to equation (17). Regarding X_C, under Loewe additivity it must hold that $\alpha_A + \beta_A \log(x_A + \rho(X_B)x_B + \rho(X_C)x_C) = \alpha_C + \beta_C \log(X_C)$, i.e. X_C reaches the same effect as the drug A equivalent dose of the combination (x_A, x_B, x_C). This allows to re-express $X_C = \exp\left(\frac{\alpha_A - \alpha_C}{\beta_C}\right)(x_A + \rho(X_B)x_B + \rho(X_C)x_C)^{\frac{\beta_A}{\beta_C}}$, resulting in

$$\begin{aligned} \psi = X_C^{\frac{\beta_C}{\beta_A} - 1} &= \left(\rho_1^{-\frac{\beta_A}{\beta_C}} (x_A + \rho(X_B)x_B + \rho(X_C)x_C)^{\frac{\beta_A}{\beta_C}} \right)^{\frac{\beta_C}{\beta_A} - 1} = \\ &= \left(\frac{x_A}{\rho_1} + \left(\frac{\rho_0}{\rho_1}\right)^{\frac{\beta_A}{\beta_B}} \psi^{\frac{\beta_C(\beta_B - \beta_A)}{\beta_B(\beta_C - \beta_A)}} x_B + \psi x_C \right)^{1 - \frac{\beta_A}{\beta_C}} \end{aligned} \tag{18}$$

The expressions (17)-(18), reflecting the additive model, are complex and do not yet prove useful for the derivation of a uniform design. However, ψ can be approximated as described by Tan et al. (2009). The steps involve i.a. a linear approximation of $\psi^{\frac{\beta_C(\beta_B - \beta_A)}{\beta_B(\beta_C - \beta_A)}} x_B$, followed by a binomial expansion of the right-hand side of expression (18) with terms up to the second order, eventually yielding a quadratic equality that can be solved for ψ. The solution is

$$\begin{aligned} \psi(x_A, x_B, x_C) \approx \frac{\beta_C^2 h_1(x_A, x_B)}{\beta_A(\beta_C - \beta_A) h_2^2(x_B, x_C)} \Bigg(\left(1 - \frac{\beta_A}{\beta_C}\right) h_2(x_B, x_C) - h_1^{\frac{\beta_A}{\beta_C}}(x_A, x_B) \\ + \left[\left(\frac{\beta_C - \beta_A}{\beta_C} h_2(x_B, x_C) - h_1^{\frac{\beta_A}{\beta_C}}(x_A, x_B) \right)^2 + \frac{2\beta_A(\beta_C - \beta_A)}{\beta_C^2} h_2^2(x_B, x_C) \right]^{\frac{1}{2}} \Bigg) \text{ with} \\ h_1(x_A, x_B) = \frac{x_A}{\rho_1} + \left(\frac{\rho_0}{\rho_1}\right)^{\frac{\beta_A}{\beta_B}} \frac{\beta_A(\beta_C - \beta_B)}{\beta_B(\beta_C - \beta_A)} x_B, \\ h_2(x_B, x_C) = \left(\frac{\rho_0}{\rho_1}\right)^{\frac{\beta_A}{\beta_B}} \frac{\beta_C(\beta_B - \beta_A)}{\beta_B(\beta_C - \beta_A)} x_B + x_C, \end{aligned} \tag{19}$$

Given this solution for ψ, the additive model (17) can finally be re-written as an additive structure in three separate logarithms. Analogous to the constant relative potency case, one obtains

$$\begin{aligned} y = &\alpha_A + \beta_A \log(z_1) + \beta_A \log((1 - \rho_0)z_2 + \rho_0) + \\ &\beta_A \log\left(\left(1 - \frac{\rho_1}{\rho_0}\right)(1 - z_3) + \frac{\rho_1}{\rho_0} \right) \end{aligned} \tag{20}$$

but the expressions for z_1, z_2 and z_3 are now given by

$$
\begin{aligned}
z_1 &= x_A + \left(\frac{\rho_0}{\rho_1}\right)^{\frac{\beta_A}{\beta_B}-1} [\psi(x_A, x_B, x_C)]^{\frac{\beta_A(\beta_C-\beta_B)}{\beta_B(\beta_C-\beta_A)}} x_B + \psi(x_A, x_B, x_C) x_C \\
z_2 &= \frac{x_A}{x_A + \left(\frac{\rho_0}{\rho_1}\right)^{\beta_A/\beta_B-1} [\psi(x_A, x_B, x_C)]^{\frac{\beta_A(\beta_C-\beta_B)}{\beta_B(\beta_C-\beta_A)}} x_B + \frac{\rho_1}{\rho_0}\psi(x_A, x_B, x_C) x_C} \\
z_3 &= \frac{\psi(x_A, x_B, x_C) x_C}{z_1}
\end{aligned}
\tag{21}
$$

This again facilitates uniform sampling of (z_1^j, z_2^j, z_3^j) on the domain S. The actual doses (x_A^j, x_B^j, x_C^j) of interest can then be determined by numerically solving system (21). Altogether, the logic and mechanics are still similar to the constant relative potency case, but the conversions between the three drugs in the model formulation become more involved. Two important comments apply to both the case with constant and non-constant relative potencies: First, as in the uniform ray design, uniformity is not achieved by random sampling from a uniform distribution, but instead using uniform design theory by K.-T. Fang et al. (2018) and K.-T. Fang et al. (2000): One can make use of existing U-type tables with three columns (one for each drug), containing perumtations of numbers 1,...,n with n being the number of unique design points here. Each row in the table can then be mapped into a point on a $[0, 1]^3$ cube. In doing so, U-type tables minimize the resulting deviation from uniformity on the $[0, 1]^3$ cube and typically achieve very small discrepancy values. Ultimately, the points on the $[0, 1]^3$ cube need to be mapped to points on the triangular prism in the design space S, which is straight-forward here but is not generally straight-forward for more complex design spaces (H. B. Fang et al., 2017; Tan et al., 2003, 2009; Tian et al., 2009). Once uniform points on S have been obtained, they have to be transformed to points on the actual dose space S_0 according to systems (16) resp. (21). Yet, since the total dose is restricted according to drug A in S, it is not enough to solve systems (16) resp. (21) for x_A, x_B and x_C: The resulting solutions for x_B and x_C still have to be scaled to drug B resp. drug C units by taking

$$
x_B^{\frac{\beta_A}{\beta_B}} \cdot \exp\left(\frac{\alpha_A - \alpha_B}{\beta_B}\right) \quad \text{resp.} \quad x_C^{\frac{\beta_A}{\beta_C}} \cdot \exp\left(\frac{\alpha_A - \alpha_C}{\beta_C}\right) \tag{22}
$$

(Tan et al., 2009). In general, the described approach builds on a mixture-amount logic while carefully balancing drug potencies. It restricts attention to an additive effect range of interest,

distributing combination doses such that each drug's dose may range between 0 and the maximum monotherapy dose compatible with the upper effect range limit (excluding these edge cases). The mixture amount reasoning is even more evident in the simple two-drug case with constant relative potencies where one can directly sample total amounts and mixture proportions uniformly based on a rectangular design space (Tan et al., 2003). Apart from monotherapies with log-linear dose-response curves, uniform maximum power designs are also developed for monotherapies with linear dose-response curves or combinations of linear and log-linear dose response curves (H. B. Fang et al., 2009, 2017). In terms of the intended analysis, it should be kept in mind that a lack-of-fit test based on the additive model used in the design generation requires not only very reliable prior information about the monotherapy dose-response curves but also identical conditions across the previously conducted monotherapy experiments and the current combination experiment. Otherwise, the procedure may lead to false synergy or antagonism calls in the presence of strong experiment-to-experiment variations.

Nonetheless, even without the intended analysis, the design has a merit on its own: By building on an additive model, it allows to systematically cover a particularly interesting area of the dose space - such as the area covering combinations with $20 - 80\%$ expected effect under additivity - with as few or as many points as desired. It should be kept in mind, though, that uniformity on the z-domain does not imply uniformity on the dose-domain. Work by Huang & Chen (2021) made advances towards balancing uniformity on the z-domain (with the aim of reaching maximum power for synergy testing) against uniformity on the dose-domain (with the aim of improving response surface modeling). Huang & Liu (2024) also work on bringing the two domains together and allowing uniform sampling on the actual dose space, instead of uniformly sampling on sub-spaces implied by particular transformations, while maintaining the maximum power property of uniform designs for detecting deviations from additivity. This is achieved by taking another approach at modeling additivity, but is only developed for the two-drug case. It overcomes the shortcoming of the presented approach by Tan et al. (2009), which effectively only samples one set of mixture proportions at any sampled total amount and thereby still does not allow to explore the dose space freely in a fully two-dimensional (two drugs) resp. three-dimensional (three drugs) fashion (Huang & Liu, 2024). Altogether, the work by Huang & Chen (2021) and Huang & Liu (2024) highlights that uniform designs constitute an active area of research and applications to multi-drug combinations may be expected to be extended further in the future.

3.5 Further Designs

Apart from the discussed uniform designs, there are also other space-filling fractional factorial design approaches, e. g. approaches that maximize the minimal distance between pairs of points. Discussions on different space-filling approaches in the context of drug combination studies are provided by Zhou & Xu (2014) or S. S. Liu et al. (2016). Moving away from space-filling approaches, Brennan et al. (2022) propose a normalized diagonal sampling design, which samples combinations along the diagonal of an appropriately normalized checkercube. In this design, one first chooses a normalization metric of interest, possibly the EC50, and places multiples of this metric along the checkercube axes, effectively yielding a serial dilution based on the normalization metric. One then samples the diagonal of the resulting normalized checkercube, i.e. points where all drugs are at their lowest concentration until points where all drugs are at their highest concentration. Altogether, the idea consists in using a newly created dose-effect based synergy metric and identifying the combination with the lowest total dose that is still effective with regard to a criterion of interest, requiring a binary endpoint. The approach furthermore aims at clarifying whether interactions in a triple combination are due to pairwise or three-way interactions. To do so, it also samples two-drug combinations, setting one component of any particular three-drug combination on the diagonal of the checkercube to zero at a time. A downside of the approach is that it samples few triple combination points, many of which purposefully constitute high dose combinations. Although not the focus of this master's thesis, some authors recommend sequential designs for multi-drug combination experiments, reviewed e.g. briefly by Novick & Peterson (2014). These authors note that their applicability may be limited if an assay's variability is too high, possibly creating spontaneous shifts in responses across sequential experimentation steps. Hence, the gain in knowledge that could be leveraged across subsequent design steps must be balanced against the inaccuracy introduced by an assay's inherent run-to-run and experiment-to-experiment variability.

4 Design Evaluation: Simulation Set-Up

This chapter proceeds towards an implementation and evaluation of selected experimental designs for triple combination therapies in a simulation study. It introduces the applied methodology.

4.1 Monotherapy Scenarios and Simulation Work-Flow

Monotherapy scenarios: The simulation study investigates seven different monotherapy scenarios. In all scenarios, monotherapies are assumed to follow a four-parameter logistic regression model of the following form (Holland-Letz & Kopp-Schneider, 2015; Huang & Chen, 2021):

$$\begin{aligned} y_i \equiv f_i(x_i) &= A_i + \frac{B_i - A_i}{1 + \exp\Big(\frac{1}{scal_i}(\log(EC50_i) - \log(x_i))\Big)} \\ &= A_i + \frac{B_i - A_i}{1 + \Big(\frac{x_i}{EC50_i}\Big)^{-\frac{1}{scal_i}}} = A_i + \frac{B_i - A_i}{1 + \Big(\frac{x_i}{EC50_i}\Big)^{-h_i}} \quad \text{with} \quad h_i \equiv \frac{1}{scal_i} \end{aligned} \tag{23}$$

where the effect y_i is subsequently taken to express the percentage of cell killing induced by dose x_i of drug $i \in \{1, 2, 3\}$ in the context of an oncological cell-based assay with cytotoxic drugs. The parameters A and B represent the lower and upper asymptote, whereas $EC50$ denotes the dose where the half-maximal effect $y = (A + B)/2$ is reached. The parameter $scal$ resp. h governs the slope of the curve. A positive value indicates a positive slope while a negative value indicates a negative slope. In the following, the parameterization in terms of $h \equiv \frac{1}{\text{scal}}$ is reported as it facilitates a more intuitive understanding with higher absolute values indicating steeper slopes[3].

The investigated parameter combinations in each scenario are listed in Table 1. The difference between the scenarios lies in the assumed slopes of the monotherapies. The EC50 values are the same throughout all scenarios and placed at $x_1 = 1$ for drug 1, $x_2 = 2.5$ for drug 2 and $x_3 = 10$ for drug 3, making drug 1 the most potent and drug 3 the least potent drug in the combination. The lower asymptotes also remain fixed at a value of 10% killing[4]. For the upper asymptotes, two sub-cases are considered within each scenario: In case a), all drugs are full agonists with an

[3] Yet, in the simulation, the parameter $scal$ was specified at particular values of interest, explaining why the transformed parameter h may subsequently take seemingly odd decimal values

[4] The lower asymptote was conveniently placed at 10% to avoid negative effects at low doses in the simulation.

upper asymptote at the maximal response value of 100% killing. In case b), drug 2 and drug 3 are partial agonists with upper asymptotes at 80% resp. 60% killing.

Table 1: Overview of Monotherapy Scenarios

Drug	Parameters	Scenarios						
		1	2	3	4	5	6	7
Drug 1	**Slope h**	**1.00**	**0.50**	**2.00**	**1.00**	**1.00**	**0.66**	**1.43**
	EC50	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Lower A	10	10	10	10	10	10	10
	Upper B	100	100	100	100	100	100	100
Drug 2	**Slope h**	**1.00**	**0.50**	**2.00**	**0.50**	**0.80**	**0.50**	**1.11**
	EC50	2.50	2.50	2.50	2.50	2.50	2.50	2.50
	Lower A	10	10	10	10	10	10	10
	Upper B, case a)	100	100	100	100	100	100	100
	Upper B, case b)	80	80	80	80	80	80	80
Drug 3	**Slope h**	**1.00**	**0.50**	**2.00**	**2.00**	**1.43**	**0.83**	**2.00**
	EC50	10.00	10.00	10.00	10.00	10.00	10.00	10.00
	Lower A	10	10	10	10	10	10	10
	Upper B, case a)	100	100	100	100	100	100	100
	Upper B, case b)	60	60	60	60	60	60	60

Altogether, scenarios with equal low, medium and high slopes as well as scenarios with mixed slopes - either with very different profiles or with only moderately different profiles - are explored to gain insights into the performance of experimental designs for triple combination therapies under different conditions. Notice that all subsequently tested experimental designs comprise $3 \times 7 = 21$ monotherapy doses, which form the basis for fitting monotherapies' dose-response curves and determining the expected additive effect at combination points for the purpose of synergy assessment. Hence, for each monotherapy scenario, suitable seven-point dilution series need to be developed: Based on graphical inspection of the dose-response curves, a maximal dose and an integer dilution factor are chosen such that the resulting doses are equally spaced on the log-scale and two doses are placed towards the lower asymptote, three doses on the linear segment

around the EC50 and two doses towards the upper asymptote (Sebaugh, 2011). A visualization of the dose-response curves, indicating the chosen dilution series, is provided in appendix A.1. The dilution series always remains fixed for cases a) and b). In addition to the 21 monotherapy doses, each subsequently tested design comprises four controls without any drug present and up to 71 triple combination doses, such that up to 96 unique design points (x_A^j, x_B^j, x_C^j) are obtained as motivated in section 2.2. The remainder of this section describes the general simulation work-flow. Details about the selection of the triple combination points are postponed to section 4.2.

Simulation Work-Flow: To begin with, monotherapy effects at the 21 monotherapy points and the four control points are generated based on the assumed four-parameter logistic regression models. Then, up to 71 combination effects under Bliss independence are generated for each design under investigation: To do so, for any combination (x_A^j, x_B^j, x_C^j), the effects of doses x_A^j, x_B^j and x_C^j are calculated based on the respective logistic regression model and then combined under the assumption of Bliss independence as described in section 2.1.2. Since the effect ranges are not the (0,1) interval but rather the (0.1, 1) interval or another sub-interval such as (0.1, 0.6) resp. (0.1, 0.8) for partial agonists, the rescaling approach suggested by Thas et al. (2022) is used during the calculation of Bliss null values to respect the model's probabilistic nature.

In the next step, synergy is added to certain combination points, resulting in a 15% increase in the effect relative to the null effect under Bliss independence, i.e. any synergistic effect is set equal to 1.15 times the null effect. At this point, the ground truth data is successfully generated and noise is added to both the monotherapy and combination points by drawing from a normal distribution with mean=0 and standard deviation=3, a value that is informed by real-life experiments. Notice that the generation of synergistic effects before adding noise reflects the understanding that - while there is a true synergy effect size - it may not be equally pronounced in all experiments, with fluctuations in experimental conditions sometimes favoring or hindering the emergence of synergy. Hence, allowing some strengthening or canceling out of the synergistic effects by the random noise in the simulation appears to be indicative of what may happen in real-world experimentation. After adding noise, it is checked whether the effects exceed the natural effect range of $0 - 100\%$ killing; if this is the case, they are brought back into this range by drawing from a truncated normal distribution with standard deviation=3. For each design point, a total of four replicates with random noise is generated in each simulation run. The simulated data, comprising up to $4 \times 96 = 384$ points for each design, is ultimately analyzed based on the Bliss null model as well. Each design is

evaluated based on 200 simulation runs; monotherapy effects are kept identical across all tested designs to not distort the analysis (see section 4.3 for further information on the analysis).

Synergy Patterns: While the synergy effect size is fixed at 15% relative effect, the synergy pattern varies during the simulations: Synergy is located in different tetrahedra in the dose space. Three out of the four vertices of the tetrahedra are fixed at $(0, 0, 0)$, $(x_1^{max}, 0, 0)$ and $(0, x_2^{max}, 0)$ with x_i^{max} representing the highest monotherapy dose of drug i, but the "center" vertex (x_1, x_2, x_3) varies across simulation runs. Before the simulation, 200 center coordinates are generated for each monotherapy scenario by the following procedure: First, a sequence of 150 values evenly spread on the log-scale in each drug's activity range is created. Second, a 150x150x150 combination grid is set up. Third, drugs' EC50 values are perturbed by adding random noise from a normal distribution with mean=0 and standard deviation=$0.4 \times$EC50; to avoid too low or negative values, a minimum value of $0.2 \times$EC50 is enforced. In the next step, the euclidean distance of each grid point to the perturbed EC50 coordinate is calculated and further noise is introduced by scaling the distances with random numbers from a uniform distribution between 0.1 and 1. The grid point resulting in the lowest distance is then chosen as center coordinate[5].

The described procedure places the center coordinate close to the EC50 coordinate $(\text{EC50}_1, \text{EC50}_2, \text{EC50}_3)$ with random variations. Yet, despite the variations, the resulting center coordinate always lies between 0.2 and 2 times the EC50 in any direction. A visualization of the center coordinates in relation to the EC50 coordinate is displayed in appendix A.1. Overall, synergy is often hypothesized to occur at low to medium dose dose combinations (Calzetta et al., 2024) and, consistent with this hypothesis, the synergy tetrahedra cover a wider range of low dose resp. low to medium effect combinations. However, by extending along the drug 1 and drug 2 axes, the tetrahedra also cover some areas in which high null effects are expected. In terms of integration into the simulation, in each simulation run it is checked which combination points lie in the tetrahedron designated to this run and synergy is then added to these points.

[5]Notice that the lower limit of 0.1 in the perturbation of the distances based on uniform random numbers avoids extreme perturbations that yield center coordinates very far from the EC50; this maintains consistency in the simulated synergy pattern and thereby improves interpretability of the results. The main source of variability in the center coordinates stems from the EC50 perturbations.

4.2 Implementation of Experimental Designs

For each monotherapy scenario, seven experimental designs are tested. The designs' implementation is described below. Visualizations of the design points are contained in appendix A.1.

Design 1: Uniform maximum-power design The first design is a uniform maximum power design following Tan et al. (2009). Since log-linear dose-response curves are needed to implement the uniform maximum power design, the following approach is taken: When all drugs are full agonists, sharing the same effect range, the logistic regression models are log-linearized before applying the uniform maximum power design as recommended by Huang & Chen (2021). Otherwise, the partial agonist with the lowest upper asymptote, in this case drug 3, is used as a reference drug and approximated by a log-linear model around its EC50. Then, the effect range where this approximation appears satisfactory is determined and well-fitting log-linear models are derived for the remaining two drugs in this same effect range (ensuring that residuals lie within a range of approximately $\pm 2\%$ and quartiles within a range of approximately $\pm 1\%$). For the case with full agonists, effect ranges of interest are always set to $20 - 80\%$ based on the additive model while for the case with partial agonists case-by-case decisions are made based on the effect range that can be well-fitted by the log-linear approximations.

In the design generation, a U-type table with 71 runs and 71 factor levels, obtained from the R package `UniDOE` (Yang, 2019), is used to derive 71 combination points on a $[0, 1]^3$ cube; these points are then mapped to points on a uniform prism in the design space S of the transformations (z_1, z_2, z_3) following the algorithm by Tan et al. (2009). Afterwards, actual drug doses (x_1, x_2, x_3) need to be derived by solving systems (16) resp. (21) stated in chapter 3.4. For the simple case with equal slopes on the log-linear scale (scenarios 1-3 a)), system (16) is solved analytically. For the general case with unequal slopes or partial agonists, system (21) is solved numerically using root-finding algorithms as residual-based optimization is not guaranteed to fit all equations equally well in a potentially unstable system. Achieving convergence may still be numerically challenging, depending on the parameters of the dose-response curves. It turns out to be notably more robust for negative slopes than for positive slopes, which is why curves with negative slopes are used for the design derivation. Generally, convergence is best achieved when $\beta_3 \leq \beta_2 \leq \beta_1$, but partially other permutations of the drugs' roles in system (21) have to be tested to achieve convergence (scenarios 4-7, case b) respectively). Different starting values are

explored and typically placed around the EC50 of the drugs for the case of full agonists resp. at a central value in the respective drug's range of interest for the case of partial agonists.

Design 2: Ray design with mixture proportions based on EC50 values The second design follows the ideas of i.a. Berenbaum (1978) and Dou et al. (2011), placing "equipartition" rays along an (approximate) isoocontour plane corresponding to a (relative) effect of 50% and thereby exploring different ray ratios based on the drugs' EC50 values. In the present application, ten rays are constructed based on mixture proportions determined by the inner points of a simplex lattice of degree six connecting the drugs' EC50 values. The ray ratios thus correspond to (1:1:1), (4:1:1), (1:4:1), (1:1:4), (3:2:1), (3:1:2), (2:3:1). (2:1:3), (1:3:2), (1:2:3) times $\left(\text{EC50}_1 : \text{EC50}_2 : \text{EC50}_3\right)$. For each ray, the maximum total amount is chosen such that the expected effect under Bliss independence equals 95% killing. A common, preferably integer dilution factor is then selected such that the minimum total amount corresponds to approx. 15% killing under Bliss independence and is reached within seven equally spaced steps on the log-dose scale. This decision is made to potentially be able to model an (almost) complete dose-response curve for each ray, which may be of practical interest in real-world applications. The approach is applied in the same way for scenarios with full and partial agonists, while acknowledging that especially in the latter case the resulting rays are not evenly distributed across any sensibly defined isocontour plane and may explore total doses that lie on the asymptote of partial agonists. Maximum total doses and dilution factors may differ between scenarios with full and partial agonists, but always adhere to the described logic.

Design 3: Uniform ray design using only the dose range The third design is a uniform ray design (S. S. Liu et al., 2016). However, unlike in many other practical applications, it does not work with an underlying factor-level table but freely explores directions in the dose space in the following sense: First, a U-type table with ten runs and ten levels is used to determine ten uniform points on a $[0, 1]^3$ cube, again relying on the R package `UniDOE` (Yang, 2019). Second. these points are then scaled to the actual dose range of interest, using each drug's maximum dose from the monotherapy dilution series, thereby yielding points in the cube $[0, x_1^{max}] \times [0, x_2^{max}] \times [0, x_3^{max}]$. Each point then determines one set of mixture proportions $p_i = \frac{x_i}{x_1+x_2+x_3}$ for $i \in \{1, 2, 3\}$, which is ultimately used for the construction of a ray. As in case of design 2, a total of ten rays is constructed and a seven-step dilution is used according to the same principles.

Design 4: Uniform ray design using factor-level table The fourth design is a uniform ray design, constructed based on an underlying factor-level table (S. S. Liu et al., 2016). Here, a U-

type table with ten runs and five levels is utilized to uniformly select factor-level combinations from a factor-level table containing each drug's EC20, EC40, EC50, EC60 and EC80[6]. The resulting factor-level combinations represent points in the dose space, which - in turn - determine mixture proportions and thus directions along which rays can be extended according to the same logic as described for design 3. In the present application, the factor levels around the EC50 were placed closer to each other because the slope is steeper in these parts of the dose-response curves.

Design 5: 4x4x4 checkercube with 4 dilutions around EC50 values The fifth design is a checkercube design. Using the monotherapy dilution series, the closest two dilution steps above and below each drug's EC50 are selected and combined in a full-factorial design.

Design 6: Uniform sampling from 5x5x5 checkercube with 5 lowest dilutions The sixth design uniformly samples from a checkercube based on the five lowest doses of each monotherapy.

Design 7: 4x4x4 checkercube with 4 effective concentrations The seventh design is a checkercube design, where each drug's EC10, EC30, EC50 and EC60 values are combined. Effective concentrations on the checkercube axes are recommended e.g. by C. Chen et al. (2018). The levels are chosen so as to fit scenarios with full and partial agonists here, attempting to explore relevant parts of the drugs' activity ranges while avoiding too many high dose combinations.

4.3 Synergy Assessment

The synergy assessment evaluates the deviation between the observed effect and the expected effect under Bliss independence at each combination point (x_A^j, x_B^j, x_C^j), naturally focusing on "true" combination points that do not lie on any monotherapy axis for interaction assessment (subsequently referred to as off-axis points for enhanced clarity). Overall, the analysis makes use of features of the BIGL R package (Makariadou et al., 2024; Thas et al., 2022; Vanderborght et al., 2017), extending it to accommodate triple combinations. It closely relies on the methodology described by Thas et al. (2022), who introduced a general bootstrapping procedure and thereby

[6]Unfortunately, however, the corresponding U-type table has one repeated row. To deal with this, a replacement row is created by testing all candidate rows and picking the one with the smallest uniformity deviation. By this modification, all rows in the resulting U-type table are unique. Still, by peculiarities of the dose-response curves in the scenarios with equal slopes, not all rows translate into distinct mixture proportions; this is not corrected - in the few cases where a mixture proportion repeats, it is tested only once, resulting in a lower number of design points.

allow to avoid parametric assumptions about the synergy pattern during estimation and inference.

In each simulation run, the following steps apply: Monotherapy axis points are first used to estimate dose-response curves for the three monotherapies based on four-parameter logistic regression models. The resulting parameter estimates are then used to predict the expected effect in terms of percentage cell killing under Bliss independence at each off-axis point (x_A^j, x_B^j, x_C^j), allowing to compute the deviation between the observed effect - averaged over all replicates - and the expected effect at point j, i.e. $E_j = \overline{y_j} - \hat{y}_j$ with $\hat{y}_j$ denoting the predicted effect under Bliss independence (Makariadou et al., 2024; Thas et al., 2022). These "residuals" E_j are an effect size measure, with values above (below) zero suggesting synergy (antagonism) at point j. To account for inherent variability in the data and limit the amount of false-positive conclusions, a formal inferential procedure is followed for ultimately calling synergy (antagonism) instead of additivity at point j: Bootstrap-based confidence intervals are constructed around each E_j with a simultaneous nominal coverage rate of 90% across all combination points, thus allowing for a family-wise error rate of 10%. This means that in 10% of the simulations at least one false synergy or antagonism call would be expected across all combination points under investigation. Details about the construction of such bootstrap-based simultaneous confidence intervals are provided by Thas et al. (2022); the present application uses 750 bootstrap samples for their computation. The reliance on confidence intervals allows to judge the practical relevance of findings in addition to their statistical significance, facilitating a more informed decision about whether detected interactions are meaningful or not compared to stand-alone hypothesis tests.

4.4 Performance Metrics

A key performance metric to be evaluated is the sensitivity of the synergy detection achieved by each design. Sensitivity is calculated for each simulation run as the ratio between the number of correctly identified synergistic points and the total number of simulated synergistic points covered by the respective design. It thus reflects the percentage of simulated synergistic points that are correctly identified in the analysis ("true positive tests"). In case a specific design misses the entire synergy area in a particular run, sensitivity cannot be calculated; these cases are discarded from sensitivity-related computations. To strengthen the practical relevance of the design evaluation, the power to detect at least five (ten) synergistic points is regarded as another key performance metric.

Even though these thresholds are arbitrary, they allow to gauge whether a design manages to send an interesting signal, recognizing that very few isolated synergy calls may not warrant further downstream investigation of a triple combination therapy in practical applications. These power metrics are computed as the number of simulation runs with at least five (ten) correct synergy calls divided by the total number of simulation runs[7]. For completeness, specificity and family-wise error rate (FWER) are also evaluated. Specificity is calculated for each simulation run as the ratio between the number of correctly classified non-synergistic points and the total number of simulated non-synergistic points. It thus reflects the percentage of additive points that is correctly classified as non-synergistic ("true negative tests"). Importantly, since the focus is on synergy detection, false antagonism calls are still counted as a correct "no synergy" calls in the specificity computation. The family-wise error rate, in contrast, is computed as the percentage of runs with at least one false call, taking into account both false synergy and antagonism calls.

Sensitivity and specificity are averaged over simulation runs and reported as means with standard deviations. The latter allow to judge the run-to-run variability and thus designs' robustness in the light of variations in the simulated synergy pattern. All results are interpreted descriptively. A formal comparison of performance metrics based on hypothesis tests is not conducted for two reasons: First, even if one were to assume a true "population" value for e.g. the sensitivity, it would not only depend on the design and the analysis but also on the run-specific synergy pattern, making inference with regard to population means less appealing. Second, statistical inference is of limited practical value here: Metrics such as sensitivity often do not have a stand-alone value. Depending on how large the synergistic area is and how well a design samples it, even a moderate sensitivity may suffice to send a practically interesting signal. Naturally, in practice it is unknown how well a design samples a synergistic area, but for this simulation study both the raw number and the proportion of combination points lying in the synergy area are reported to further contextualize the results (in addition to the power metrics).

[7]Runs in which a particular design covers less than five (ten) simulated synergistic points remain included in the power calculation. The rationale for this decision is that there is always a sufficiently large synergy area in the simulation such that the designs could detect at least five (ten) points, if they sampled in a large enough area of the dose space. Hence, this approach purposefully penalizes designs which explore too narrow segments of the dose space.

5 Design Evaluation: Results

In this chapter, the results of the simulation study examining seven experimental designs in seven different monotherapy scenarios with respect to several performance metrics are presented.

Sensitivity: Figure 1 displays a heatmap for the sensitivity achieved by each design in each scenario, averaged over simulation runs. The perhaps most striking finding is that design 7 - the checkercube with drugs' EC10, EC30, EC50 and EC60 on the axes - achieves a mean sensitivity of $\geq 93\%$ across all monotherapy scenarios and thereby notably outperforms the other designs in terms of mean sensitivity and robustness. The only design that achieves an even better performance in isolated monotherapy scenarios is design 1, the uniform maximum power design. Yet, its mean sensitivity varies more strongly by scenario, ranging between 58% and 100%. Of note, the uniform maximum power design tends to perform better than both the checkercube design with four dilution steps around each drugs' EC50 on the axes (design 5) and the uniform sampling from a checkercube design with each drugs' five lowest dilution steps on the axes (design 6) in almost all scenarios. The difference between the latter two designs is often small, with the exception of scenario 7 with a mix of steeper slopes, where design 6 appears to perform better than design 5[8]. All the designs mentioned thus far - i.e. designs 1,5,6,7 - appear to reach a higher mean sensitivity than the ray designs under investigation.

When compared among each other, the ray designs all show a similar performance; the only noteworthy exception may be that the uniform ray design - especially the one choosing directions freely in the dose space (design 3) - performs somewhat better than the EC50-based ray design (design 2) in the scenario where the three drugs have very different slopes (scenario 4 with $h_1 = 1$, $h_2 = 0.5$, $h_3 = 2$). In this particular scenario, however, all ray designs tend to achieve their best performance. They also perform well in scenario 7a) where the three drugs have different, but somewhat steeper slopes ($h_1 = 1.4$, $h_2 = 1.1$, $h_3 = 2$). When two drugs are partial agonists in scenario 7b) or all slopes are at a high value of $h = 2$ in scenario 3, this finding changes, and ray designs reach their worst mean sensitivity. Especially the difficulties in terms of synergy detection

[8] This may, however, be an artefact: In scenario 7, as well as in scenario 2, design 5 happens to include a dilution step reaching the upper asymptote of a monotherapy, thus sampling in a rather high effect region. It appears plausible that this has a more detrimental impact on sensitivity in the high slope scenario 7 than in the low slope scenario 2 because high slopes can complicate the synergy detection as discussed below.

in the high slope scenario 3 are not surprising and are shared by the remaining designs, except for the checkercube design with effective concentrations on the axes.

Note that the sensitivity heatmap can only give a partial impression of the designs' comparative performance as it ignores two important aspects: First, the designs did not cover the simulated synergy area equally well and partially even missed the synergy area in some simulation runs, leading to possibly different practical implications of the same average sensitivity values. Second, the average sensitivity summarizes each design's performance across a range of different simulated synergy patterns. Focusing merely on this point estimate may thus hide a notable run-to-run variation in the sensitivity values, depending on the simulated synergy pattern and possibly also the experimental design. These two aspects are therefore investigated further.

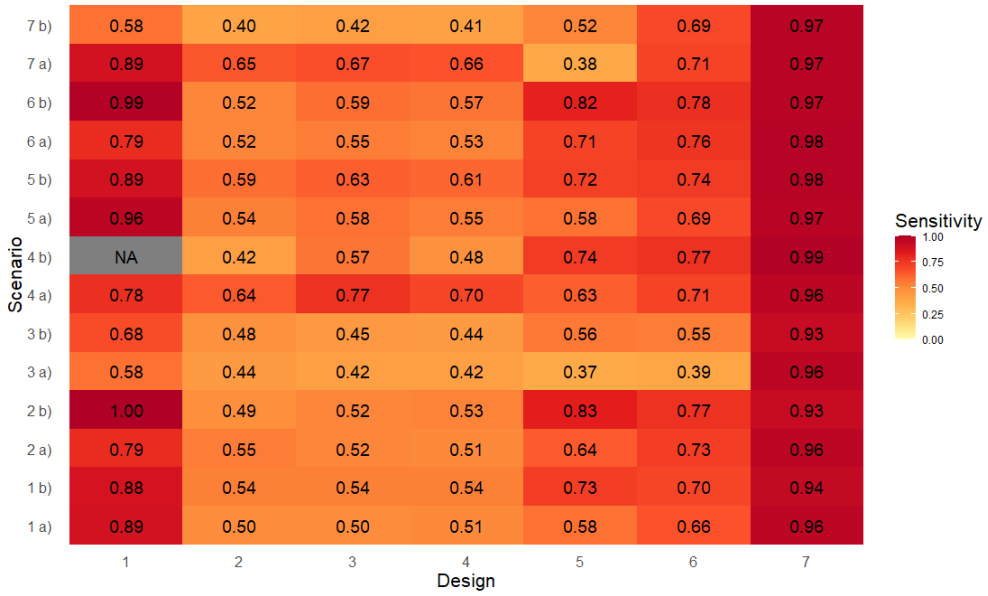


Figure 1: Heatmap displaying the mean sensitivity across simulation runs by experimental design (x-axis) and monotherapy scenario (y-axis).

The heatmap in Figure 2 displays the proportion of combination points that lie inside the simulated synergy area by design and monotherapy scenario, averaged over simulation runs. This indicates how well the designs covered the simulated synergy area, which assumed the shape of a tetrahedron with varying "center" coordinate in the dose space and emphasized synergy in low to moderate effect regions but also allowed for at least some synergy in higher effect regions (see

chapter 4.1).

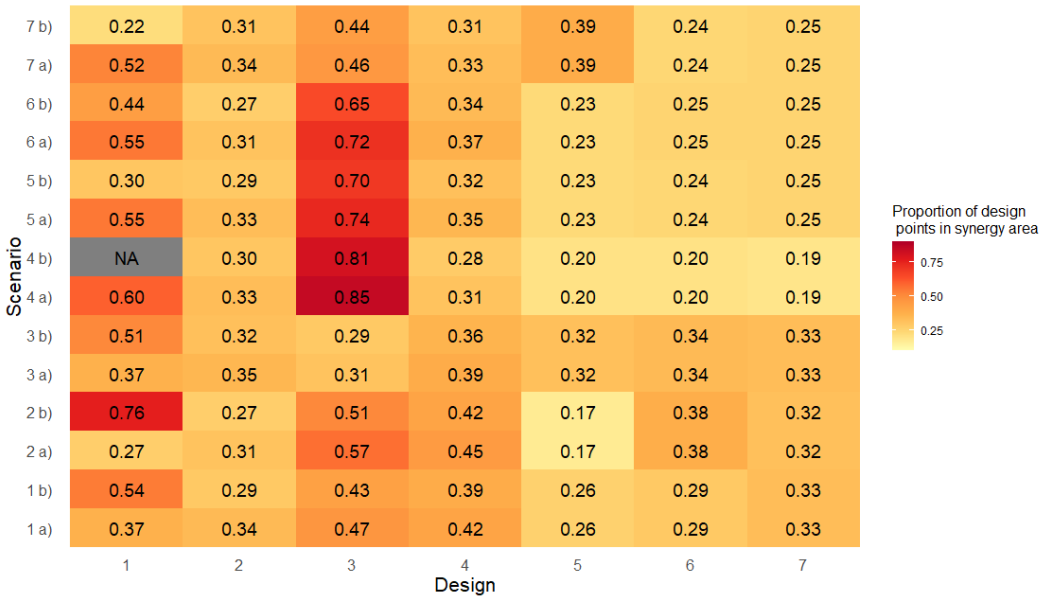


Figure 2: Heatmap displaying the mean proportion of design points lying in the synergy area across all simulation runs by experimental design (x-axis) and monotherapy scenario (y-axis). Notice that the color range is adjusted to the data range with 0.1<proportion<0.9

With regard to coverage, design 1 again appears to perform rather well, albeit with notable variations in the proportion of design points that lie within the synergy area across monotherapy scenarios and within monotherapy scenarios between the cases with full versus partial agonists. The difference between the latter two cases is not seen for any other design and is generally not expected because the synergy pattern was identical in these two cases and the design construction always followed the same premises. However, design 1 constitutes an exception in this regard: In the special case of the uniform maximum power design, there was a difference in the design construction between full and partial agonist cases. For the full agonist cases, the effect range of interest was specified to $20 - 80\%$ effect under Loewe additivity. For the partial agonist case, in contrast, it depended on the overlap in the log-linear segments of the monotherapy dose response curves, which was often small. Nonetheless, the erratic variation with sometimes a notably higher and sometimes a notably lower proportion of synergistic points in any of the two cases also signals a problem: In the presence of partial agonists, with limited overlap in the log-linear segments, there is reduced control over where the uniform maximum power design actually samples.

One may enforce a restriction on the additive effect to lie within a joint log-linear segment, but model-based extrapolations to lower individual drug doses where the respective monotherapy's log-linearity does not hold anymore always happen, possibly leading to deviations from the effect range of interest based on the true four-parameter logistic regression models. The reason for these extrapolations is that no restrictions on the mixture proportions are enforced in the design generation, hence any additive effect of interest can be reached with a particular agent either almost entirely dominating the mixture or being almost entirely absent. In case 4b), it was even impossible to achieve convergence to a sensible solution, which is why results for this scenario are not presented. In scenario 3b), convergence was impossible for a subset of the points, indicating that caution is required here as well[9].

Among the remaining designs, the uniform ray design with ray directions varying freely in the dose space (design 3) appears to perform strikingly well in terms of coverage of the synergy area, especially in cases with differences in slopes (scenarios 4-7). A reason for this may be that design samples mixtures with rather low doses of drug 3 in these scenarios, which may be seen as an unintended feature that happens to align well with the simulated synergy tetrahedra but generally calls for further potency alignments based on factor-level tables in the use of uniform ray designs (see appendix A.1). Nonetheless, also the remaining ray designs appear to sample the simulated synergy area adequately, often placing a higher proportion of design points in the simulated synergy area than the checkercube designs. Among the checkercube designs it is interesting to note that the design with the highest sensitivity, design 7 with effective concentrations on the axes, neither over- nor underperforms in terms of sampling the simulated synergy area[10].

[9] In full agonist cases, there were generally less problems with the implementation of the uniform maximum power design. However, in case 2a) with shallow slopes, the design happens to frequently sample in a rather high effect region, at least when considering Bliss independence as a null model (see appendix A.2). Of note, the model does not exceed its admissible range based on the Loewe-additivity model formulated in terms of the log-transformation of the dose response curves as recommended by Huang & Chen (2021); yet, it does place very few points in the low dose region in this case. A similar, albeit slightly less extreme observation holds for scenario 6a), where slopes are again low. The usage of a log-linear approximation around the EC50 instead of a log-transformation does not solve this issue of infrequent sampling in the low dose region in the presence of low slopes; the only solution appears to be to further restrict the effect range of interest in these scenarios

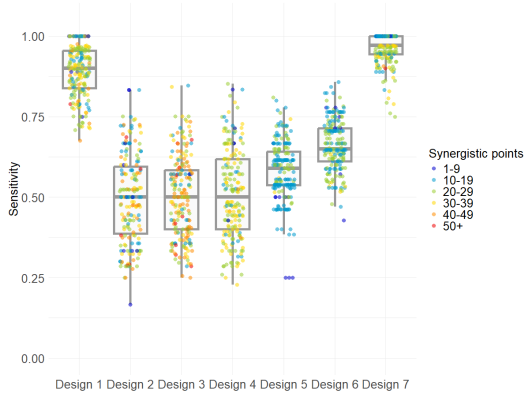
[10] This observation holds at least when excluding scenarios 2 and 7 for design 5 from the comparison, which - as mentioned before - may be artefacts due to the inclusion of a high monotherapy dose on the upper asymptote in these cases. See appendix A.2 for counts of design points with null effects $\geq 90\%$.

Altogether, when considering both Figure 1 and Figure 2, the slightly to notably higher proportion of simulated synergistic points tested by ray designs does not translate into enhanced sensitivities when compared to other designs with lower synergy coverage, such as the checkercube designs. The uniform maximum power design, in contrast, appears to frequently sample within the synergistic area and achieve a high sensitivity. In this regard, the question arises whether the high proportion of synergistic points in many scenarios results from the exploration of a wide dose range or simply a good alignment with the synergy pattern examined in this simulation study, which places particular emphasis on the low dose resp. low to medium effect regions. Especially when considering the checkercube designs as reference, it appears that the uniform maximum power design does not explore a particularly wide area in the dose space. However, it dedicates notably more points to low dose combinations, pointing towards a better alignment with the simulated synergy pattern in this study (see appendix A.1). This agrees with the fact that the effect range of interest was restricted during the construction of the design, allowing for a more thorough exploration of a somewhat narrower area in the dose space.

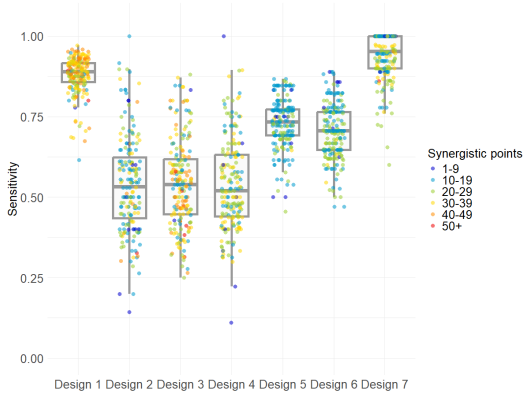
Apart from the average proportion of combination points lying in the synergy area, it is also of interest how often a design misses the synergy area completely. Detailed information hereto is provided in appendix A.2, but a clear trend is that design 2 - the EC50-based ray design - is least robust in this regard and tends to consistently miss the synergy area in almost 10% of the runs. Upon closer examination, the affected runs seem to have a rather small, but still non-empty synergy area: Illustrating for scenario 1, in 3/17 (9/17) cases where design 2 fails, the other designs cover less than five (ten) synergistic points on average. Designs 4 and 7 tend to have some problems predominantly in high slope scenarios, where they miss synergy in up to approx. 10% resp. 5% of the runs. The remaining designs always cover at least one simulated synergistic point.

After focusing on point estimates thus far, Figure 3 finally illustrates the variability in sensitivity values across simulation runs. It also continues to relate the achieved sensitivity values to the number of synergistic points tested by the respective design: The jitter on top of the boxplot indicates the run-specific sensitivity values, colored by the run-specific number of synergistic points among the combination points explored by a particular design. The first observation is that the variability in sensitivity is often large, illustrating that the synergy pattern simulated in a particular run has a large influence on how well synergy can be detected by a particular design. However, based on the boxplots' color coding as well as further graphical and correlation-based assessments,

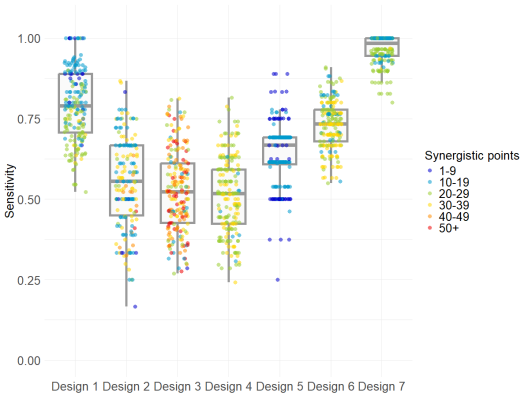
there does not seem to be a consistent relationship between the mere number of synergistic points and the sensitivity achieved by a particular design. This suggests that the placement of the synergistic points in the dose space may matter more than the mere number or proportion of synergistic points that exists. The second observation is that the variability may depend on both the monotherapy scenario and the design under investigation. On average across all scenarios, it appears that ray designs have the highest variability in sensitivity. In fact, their sensitivities' standard deviation is always higher than the standard deviation of checkercube designs (see appendix A.2). Yet, for both the ray and the checkercube designs, the variability appears rather consistent across all investigated monotherapy scenarios. This does not apply to the uniform maximum power design, where the variability is highly dependent on the monotherapy scenario. The latter design exhibits the highest variability among all designs in scenario 3 with equally high slopes of $h = 2$. Also in scenario 7b) with different, somewhat steeper slopes ($h_1 = 1.4$, $h_2 = 1.1$, $h_3 = 2$) and partial agonists it exhibits an extremely large variability, reaching any possible sensitivity value between (almost) 0% and 100%. On the other hand, it sometimes achieves a mean sensitivity close to 100%, which is only possible when the variability in sensitivity is small. The appendix A.2 provides numerical results for the mean sensitivity and the mean number as well as proportion of combination points lying in the synergy area, including the respective standard deviations.



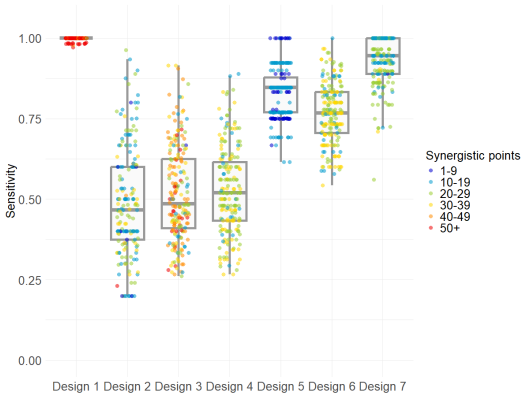
Scenario 1a)



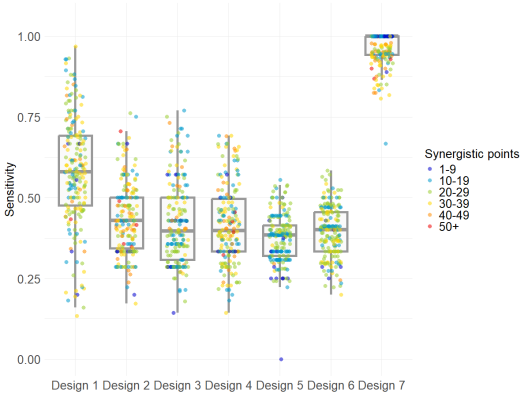
Scenario 1b)



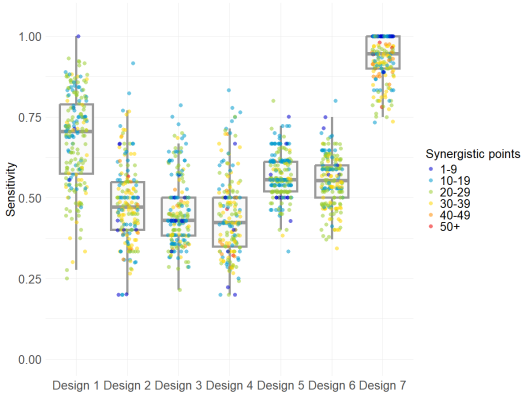
Scenario 2a)



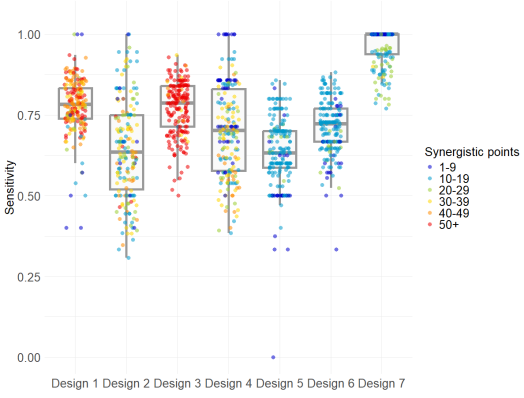
Scenario 2b)



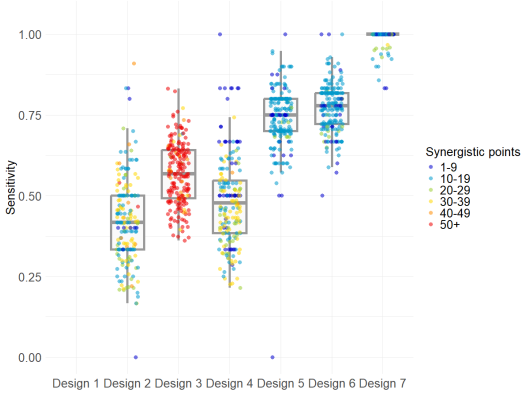
Scenario 3a)



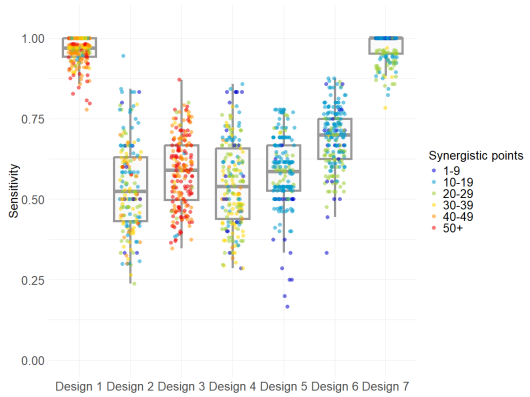
Scenario 3b)



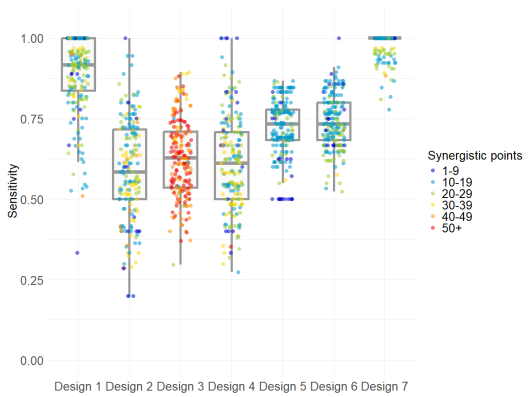
Scenario 4a)



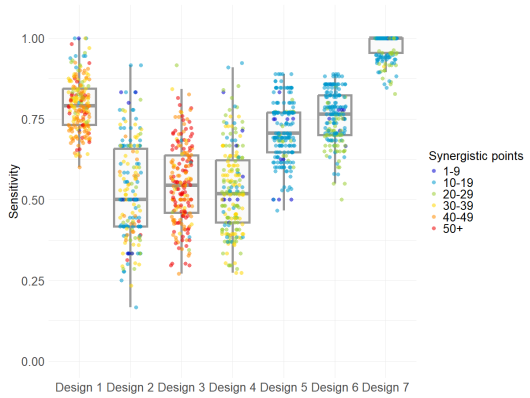
Scenario 4b)



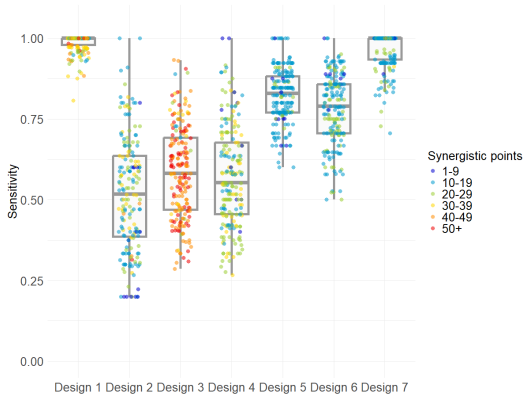
Scenario 5a)



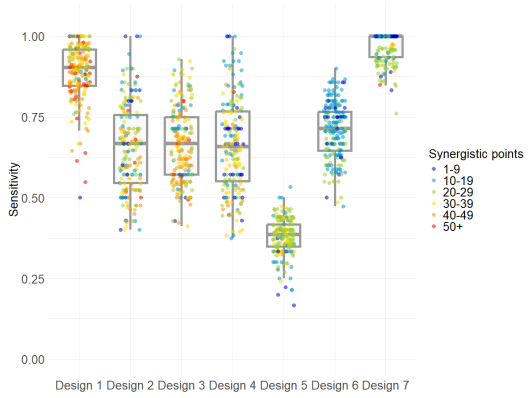
Scenario 5b)



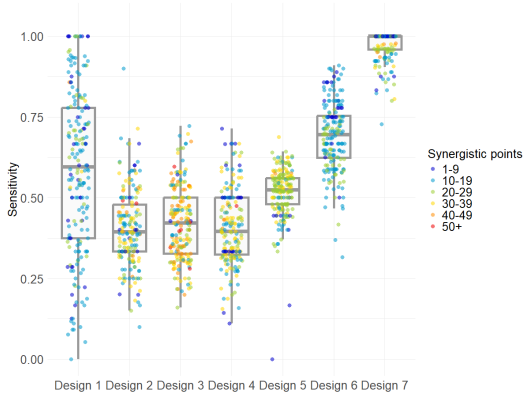
Scenario 6a)



Scenario 6b)



Scenario 7a)



Scenario 7b)

Figure 3: Box plots for sensitivity by scenario and design across 200 simulation runs. Points are colored by the number of synergistic points covered by a particular design in a particular run.

Power: Especially in the light of an often high variability in the sensitivity, it is of interest to assess how well designs perform with respect to a practically relevant metric, such as the power to detect at least five simulated synergistic points. Figure 4 displays the corresponding results in a heatmap[11]. The heatmap shows that the EC50-based ray design (design 2) has the lowest power, which is at least partially explained by the fact that it misses the simulated area of synergy more often than any other design. The uniform maximum power design (design 1), the uniform ray design without underlying factor level table (design 3) and the uniform sampling from a 5x5x5 checkercube (design 6), on the other hand, have a particularly high power, often reaching almost 100%. For the former two designs, this may be expected since they align well with the simulated synergy pattern. The latter design, however, did not appear to examine a particularly large proportion of synergistic points; nonetheless, it seems to sample wide enough with high enough sensitivity to almost consistently detect at least five synergistic points. Of interest, design 7 - the checkercube with effective concentrations on the axes, showing almost perfect sensitivity - does not manage to maintain its exceptional performance with regard to the power to detect at least five synergistic points. In this case, this owes exclusively to sometimes missing the simulated area of synergy; if only the runs where the area of synergy is not missed were considered, this design would have a power of 100%.

Also with regard to power it is noticeable that scenarios involving at least one very steep slope with h $= 2$, i.e. scenarios 3, 4 and 7, are more challenging for synergy detection, as already seen before with respect to sensitivity. Scenarios with at least one very shallow slope with h $= 0.5$, i.e. scenarios 2 and 6, on the other hand still achieve (almost) 100% power for the majority of the tested designs. An analogous heatmap for the power to detect at least ten simulated synergistic points is contained in appendix A.2. It shows that design 2 and design 5 perform notably worse across all monotherapy scenarios with respect to this stricter power metric. The remaining designs are more stable but also exhibit weaknesses in isolated scenarios, especially with steep slopes in scenarios 3, 4 and 7b).

[11] Here, it is important to remember that all runs were considered in the power calculation, irrespective of whether a particular design covered at least five synergistic points in a particular run or not. This is justified as the synergy areas were generally large enough such that designs had the chance to detect at least five synergistic points if they explored a wide enough area of the dose space: Overall, there were no instances where none of the tested designs covered at least five synergistic points.

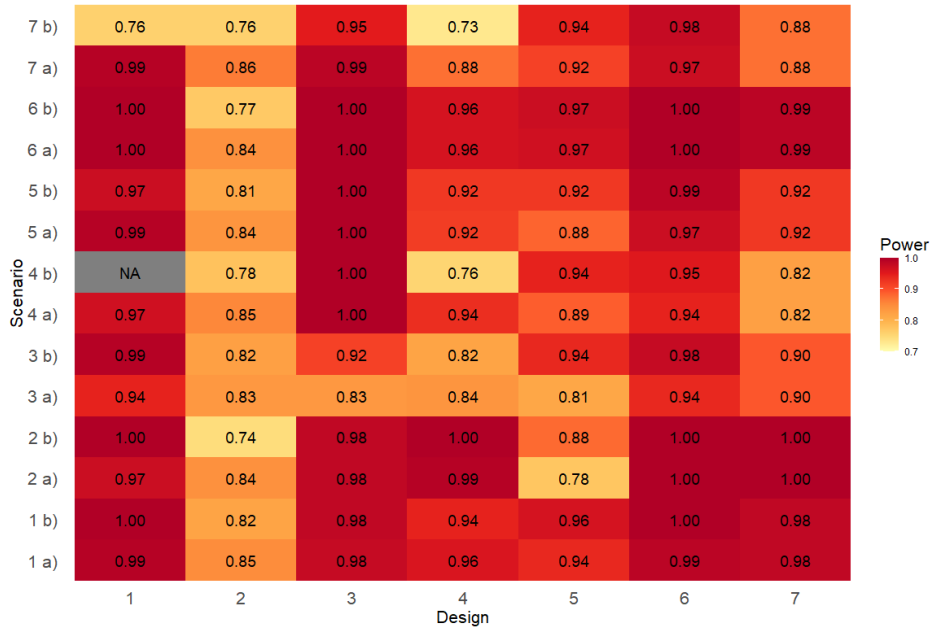


Figure 4: Heatmap displaying the power for detecting at least 5 synergistic points across all simulation runs by experimental design (x-axis) and monotherapy scenario (y-axis). Notice that the color range is adjusted to the data range with 0.7<power<1.

Specificity: Specificity was high, with a mean above 99.6% and a standard deviation below 2.2% across all designs and monotherapy scenarios, indicating that - irrespective of the design - only a small number of additive points was misclassified as synergistic. Results by scenario and design are contained in appendix A.2.

Error control: The family-wise error rate, reflecting the percentage of simulation runs where at least one false synergy or antagonism call was made, was generally controlled at approximately 10% as desired. However, there were exceptions in which the inferential procedure was either too strict or too liberal compared to the nominal level. The deviation patterns appear mostly random, with the following exceptions: Too strict decisions, leading to a lower FWER of approx. 6% on average across all designs, are systematically observed for the scenarios with either very shallow slopes of h $= 0.5$ or very steep slopes h $= 2$ or a mix thereof whenever partial agonists are present. When analyzed separately by design, uniform ray designs exhibited the most stringent error control with FWER $= 5.8\%$ (design 3) resp. FWER $= 8.4\%$ (design 4), averaged across all monotherapy scenarios. The EC50 based ray design reached FWER $= 9.1\%$ (design 2) and the

uniform maximum power design reached FWER $= 9.2\%$ (design 1). The checkercube designs achieved FWER $= 8.7\%$ (design 7), FWER $= 10.3\%$ (design 6) resp. FWER $= 12.3\%$ (design 5). The latter observation is mostly driven by one outlier, where the checkercube design with dilutions around the EC50 sampled very high effect combinations in a particular high slope scenario (scenario 7), leading to many additive effects being mistaken as antagonistic effects, probably due to the random procedure employed to ensure that effects remain in the $0 - 100\%$ killing range. In total, too liberal decisions occurred in 31/98 cases across all designs and all monotherapy scenarios, but were mostly moderate in nature; a threshold of FWER $= 15\%$ was e.g. only exceeded in 3/98 cases, pointing towards an overall satisfactory error control. The appendix A.2 provides further numerical results.

6 Discussion and Further Research Venues

The present simulation study explored seven experimental designs in seven different monotherapy scenarios with respect to their capacity to detect synergistic interactions among three drugs. During the 200 simulation runs, the location of synergies varied within the dose cube, following a tetrahedron shape with run-specific centerpoint. Altogether, synergy was created in such a way that it was most wide spread among low dose resp. low to medium effect combinations. The Bliss independence model was used both for data generation and data analysis.

The results of the simulation indicate that there may not be one single design that performs best with respect to every metric of interest and is robust across all scenarios. For instance, a 4x4x4 checkercube with effective concentrations EC10, EC30, EC50 and EC60 on the axes as suggested by C. Chen et al. (2018) showed almost perfect sensitivity across all instances. However, with respect to the power to detect at least five synergistic points it was ultimately outperformed by designs with lower sensitivities, such as a design sampling uniformly from a 5x5x5 checkercube with the five lowest serial dilution steps of the monotherapies on the axes. The main reason for this was that the 4x4x4 checkercube with effective concentrations sampled a rather narrow range in the dose space, thereby in some occasions missing the simulated synergy area. One step towards combining the advantages of these two designs could be to test uniform sampling from a checkercube with effective concentrations on the axes in future simulation studies, possibly even sampling from more than five effective concentrations to improve coverage of the design space.

Another promising design examined in the present simulation study was the uniform maximum power design by Tan et al. (2009). It tended to show high power and high sensitivy, but it is unclear whether these findings can generalize to different synergy patterns from the ones that were considered in this simulation study, as this design also covers a rather limited range in the dose space, which, however, happened to aligned well with the simulated synergy area. An advantage of this design is that the sampled region is thoughtfully chosen based on an effect range of interest under Loewe additivity. A disadvantage is that the uniform maximum power design can be difficult to implement in scenarios with partial agonists due to numerical instability and smaller shared log-linear segments on the dose-response curves. It also appears to exhibit a somewhat undesired point placement in scenarios with very low slopes. In scenarios with high slopes, it can possibly perform comparatively well, as long as there are no partial agonists and slopes are not too different.

Yet, its variability in sensitivities can become high when slopes are steep. Overall, the results suggest that this design may prove most useful for experiments involving full agonists with moderate slopes. From a practical standpoint, it must be kept in mind that this design involves considerably more laboratory effort as it does not follow any serial dilution or factorial combination scheme and thus requires the preparation of many distinct drug concentrations.

Regarding ray designs, it was found that they tend to exhibit lower sensitivities than the other tested designs, despite aligning adequately with the simulated synergy patterns. Moreover, on average, they showed the highest variability in sensitivities among all tested designs. Nonetheless, the power to detect at least five synergistic points appeared promising for uniform ray designs. Whenever the slopes vary among the drugs, it appears recommendable to base the construction of uniform rays on a suitable factor-level table to achieve a certain potency alignment. In general, optimizing the factor-level table and especially the dilution scheme further may allow to improve the performance of uniform ray designs with regard to sensitivity. For two-drug combinations, progress has been made with respect to the selection of dilution schemes for ray designs based on optimal design approaches, partially embedded in a Bayesian framework (Almohaimeed & Donev, 2014; Holland-Letz et al., 2020; Holland-Letz & Kopp-Schneider, 2015; Sperrin et al., 2015), suggesting future research venues for multi-drug combinations. Regarding the analysis, it is likely that the performance of ray designs would improve when moving from a pointwise to a raywise synergy assessment embedded in an appropriate analysis framework (Fan & Zhang, 2014). Yet, any modeling would require model checking, which is not easily feasible in a larger simulation study and potentially also resource-intensive in practical experimentation in early drug discovery.

Across all monotherapy scenarios and designs, the simulation study has confirmed that it is generally reasonable to conduct triple combination experiments with a budget of 384 runs, allowing for 96 design points and four replicates. This holds conditional on the simulated synergy pattern, the assumed synergy effect size of 15% relative to the null effect and the assumed experimental noise following a normal distribution with a standard deviation of 3. Strictly speaking, one may also condition on potency differences between the monotherapies, as the EC50 values were kept fixed at values suggesting that one of the drugs is notably less potent than the other two during the simulation. Yet, even under the somewhat simplified conditions explored in the simulation, it has also emerged that some specific monotherapy scenarios remain challenging, lowering the chances of detecting synergies. This applies especially to scenarios involving monotherapies with

high slopes: When all drugs had a slope of h $= 2$, power and sensitivity were reduced, but still reached levels of $\geq 80\%$ resp. $\geq 35\%$ in the present simulation. As long as theoretical progress regarding experimental designs for triple-combination studies continues to be largely lacking, it appears sensible to examine scenarios with high slopes in more detail in future simulation studies and investigate how to improve pragmatic checkercube and ray designs in these cases and/or test whether a higher replicate number can yield a more robust performance.

Finally, it must be kept in mind that the performance of the investigated experimental designs is also dependent on the applied analysis methodology. A strength of the implemented bootstrap-based approach is that the family-wise error rate appears, on average, well-controlled: Large deviations from the nominal level occur rarely and would possibly occur even more rarely if more simulation runs and more bootstrap samples were used. Yet, a satisfactory error control by bootstrapping methods such as the ones introduced by Thas et al. (2022) already constitutes notable progress in the area of drug interaction assessment, where especially early point-wise interaction screenings unaccompanied by modeling approaches are often conducted without any or without an adequate inferential procedure (Calzetta et al., 2024; Foucquier & Guedj, 2015; Thas et al., 2022). This is particularly dangerous as in-vitro experimentation is often faced with considerable experimental noise from various sources (Vlot et al., 2019), elevating the risk of false-positive findings in the absence of statistical error control.

The exclusive reliance on Bliss independence in the present simulation study may be considered a limitation. It primarily owes to the need for a null model that can handle both varying relative potencies and partial agonists without the need for complex adaptations. Yet, a re-examination of the simple monotherapy scenarios 1a)-3a), which are compatible with the standard Loewe additivity model, could contribute further insights into the robustness of the findings, especially when slopes are particularly shallow or particularly steep. Another way of extracting further insights from the present simulation study would be to analyze on a run-by-run level whether the implemented designs have difficulties detecting synergy in particular segments of the design space, such as particularly low or high effect areas in an attempt to better understand where the variability in sensitivities stems from. In the former case, synergy is less pronounced in terms of absolute effect and may thus be more easily swallowed by noise; in the latter case, however, the emergence of synergy may be challenged by the maximum possible effect of $100\ \%$ cell killing.

7 Ethical Considerations

Despite challenges inherent in the investigation of multi-drug combinations in in-vitro experiments and the translation of their findings into the clinic, such experiments appear crucial for gaining insights into the nature of drug interactions (Calzetta et al., 2024). In fact, as e.g. Chou (2008) argues, in-vitro studies contribute information about drug interactions that could never be obtained from clinical trials in humans: It would be ethically prohibitive to explore a wide enough range of monotherapy and combination doses in clinical trials to allow the assessment of interactions in their strict sense, defined as deviations from a suitable reference model for additive action at particular dose combinations (Chou, 2008; Foucquier & Guedj, 2015).

The exploration of multi-drug combinations already during early drug discovery may bear great potential both from a public health and an industry perspective: In oncology, for instance, curative cancer therapies (if they exist) almost always involve multi-drug combinations as single agents do not achieve a satisfactory benefit-risk profile (Palmer et al., 2019). Hence, systematically exploring a wider range of combinations therapies at earlier time points may improve success in drug discovery. Yet, the question remains how to prioritize combination therapies for testing and, subsequently, which criteria to apply when prioritizing among the tested combination therapies. Both aspects are tied to the question what main benefits one may expect from combination therapies in a particular therapeutic area. Palmer et al. (2019) discuss that focusing exclusively on combination therapies for which synergistic interactions in terms of efficacy are expected and possibly confirmed in experiments may turn out to be misleading: Currently established combination therapies in the area of oncology appear to mostly lack synergistic interactions but still prove clinically useful due to low cross-resistance, mitigating the early onset of drug resistance.

From an ethical perspective, synergistic efficacy may not be a suitable stand-alone decision criterion for the prioritization of combination therapies and should always viewed in the context of the overall efficacy, toxicity and resistance mechanism shown by a particular combination therapy (Palmer et al., 2019; Tallarida, 2011). Yet, establishing synergistic interactions at early stages in the discovery process may provide an incentive to further investigate a particular combination therapy with respect to the mentioned aspects and, ultimately, accumulate sufficient pre-clinical evidence on efficacy, toxicity and resistance mechanisms to make an informed decision about whether a combination therapy should be developed further or not. In general, before progressing to in-

vivo testing and eventually clinical trials in humans, it is crucial to ensure that all drugs involved in a combination therapy contribute to the combination's activity in a useful manner (Beppler et al., 2016; Brennan et al., 2022; Cokol et al., 2017). This aspect has not been addressed in the present master thesis as only an initial interaction screening setting, focused on detecting net interactions irrespective of how they originate, was considered. However, when synergies are detected in a triple combination context, the role of the constituent drugs in the combination should routinely be clarified in follow-up in-vitro experiments to avoid exposing animals and eventually humans to unnecessary risks from constituents that do not make a relevant contribution to efficacy but may still induce adverse effects in subsequent in-vivo experiments.

8 Conclusion

This master's thesis has demonstrated that the field of assessing interactions between drugs or, generally, chemicals has a long-standing history dating back roughly 100 years but has focused more on suitable analyses of drug interactions than on suitable experimental designs. Altogether, there appears to be a scarcity of principled design proposals for in-vitro experiments in the literature, especially for the case of three-drug or multi-drug combinations. Moreover, even existing methods such as the uniform maximum power design, are rarely used in practice and pragmatic checkercube or ray designs are preferred. The latter always involve choices, for which there is not yet any reliable general guidance, such as the specification of dose levels on the checkercube axes or ray ratios and accompanying dilution series. The absence of principled design proposals and generally valid guidance for checkercube and ray designs, however, is understandable as an optimization with respect to the mentioned aspects would mostly require knowledge about the expected interaction patterns, which is rarely available before the experimentation. This lack of knowledge creates the need to efficiently cover relevant areas of the dose space in an attempt to reduce the risk of missing interactions if they are truly present; whether this is achieved by dedicated space-filling designs such as uniform designs or, pragmatically, with e.g. checkercubes spanning a wide enough area in the dose space may appear less critical. The results of a simulation study have indicated that uniform maximum power designs and checkercube designs tend to achieve a better sensitivity than ray designs, but that almost all designs achieve a promising performance with respect to measures of practical interest, such as the power to detect at least five simulated synergistic points, across the majority of scenarios. Future research should focus on monotherapy scenarios with steep slopes, which have been found to pose the greatest difficulties for synergy detection in the presented simulation study.

References

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

Beppler, C., Tekin, E., Mao, Z., White, C., McDiarmid, C., Vargas, E., Miller, J. H., Savage, V. M., & Yeh, P. J. (2016). Uncovering emergent interactions in three-way combinations of stressors. *Journal of the Royal Society Interface*, *13*. https://doi.org/10.1098/rsif.2016.0800

Berenbaum, M. C. (1977). Synergy, additivism, and antagonism in immunosuppression. *Clinical and Experimental Immunology*, *28*, 1–18.

Berenbaum, M. C. (1978). A method for testing for synergy with any number of agents. *The Journal of Infectious Diseases*, *137*, 122–130. https://www.jstor.org/stable/30081623

Berenbaum, M. C. (1989). What is synergy? *Pharmacological Reviews*, *41*, 93–141. https://doi.org/10.1016/S0031-6997(25)00026-2

Berenbaum, M. C., Yu, V. L., & Felegie, T. P. (1983). Synergy with double and triple antibiotic combinations compared. *Journal of Antimicrobial Chemotherapy*, *12*, 555–563. http://jac.oxfordjournals.org/

Bliss, C.-I. (1939). The toxicity of poisons applied jointly. *Ann Appl Biol*, *26*, 585–615.

Boshuizen, J., & Peeper, D. S. (2020). Rational cancer treatment combinations: An urgent clinical need. *Molecular Cell*, *78*, 1002–1018. https://doi.org/10.1016/j.molcel.2020.05.031

Brennan, J., Jain, L., Garman, S., Donnelly, A. E., Wright, E. S., & Jamieson, K. (2022). Sample-efficient identification of high-dimensional antibiotic synergy with a normalized diagonal sampling design. *PLoS Computational Biology*, *18*, e1010311. https://doi.org/10.1371/journal.pcbi.1010311

Calzetta, L., Page, C., Matera, M. G., Cazzola, M., & Rogliani, P. (2024). Drug-drug interactions and synergy: From pharmacological models to clinical application. *Pharmacological Reviews*, *76*, 1159–1220. https://doi.org/10.1124/pharmrev.124.000951

Chen, C., Wicha, S. G., Nordgren, R., & Simonsson, U. S. H. (2018). Comparisons of analysis methods for assessment of pharmacodynamic interactions including design recommendations. *AAPS Journal*, *20*, 77. https://doi.org/10.1208/s12248-018-0239-0

Chen, D., Liu, X., Yang, Y., Yang, H., & Lu, P. (2015). Systematic synergy modeling: Understanding

drug synergy from a systems biology perspective. *BMC Systems Biology*, *9*, 56. https://doi.org/10.1186/s12918-015-0202-y

Chen, J., Lin, A., Jiang, A., Qi, C., Liu, Z., Cheng, Q., Yuan, S., & Luo, P. (2025). Computational frameworks transform antagonism to synergy in optimizing combination therapies. *Npj Digital Medicine*, *8*, 44. https://doi.org/10.1038/s41746-025-01435-2

Chou, T. C. (2006). Theoretical basis, experimental design, and computerized simulation of synergism and antagonism in drug combination studies. *Pharmacological Reviews*, *58*, 621–681. https://doi.org/10.1124/pr.58.3.10

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

Chou, T. C. (2014). Frequently asked questions in drug combinations and the mass-action law-based answers. *Synergy*, *1*, 3–21. https://doi.org/10.1016/j.synres.2014.07.003

Chou, T. C., & Talalay, P. (1984). Quantitative analysis of dose-effect relationships: The combined effects of multiple drugs or enzyme inhibitors. *Advances in Enzyme Regulation*, *22*, 27–55. https://doi.org/10.1016/0065-2571(84)90007-4

Cokol, M., Kuru, N., Bicak, E., Larkins-Ford, J., & Aldridge, B. B. (2017). Efficient measurement and factorization of high-order drug interactions in mycobacterium tuberculosis. *Science Advances*, *3*, e1701881.

Cokol-Cakmak, M., Bakan, F., Cetiner, S., & Cokol, M. (2018). Diagonal method to measure synergy among any number of drugs. *Journal of Visualized Experiments*, *136*, e57713. https://doi.org/10.3791/57713

Donev, A. N., & Tobias, R. D. (2011). Optimal serial dilutions designs for drug discovery experiments. *Journal of Biopharmaceutical Statistics*, *21*, 484–497. https://doi.org/10.1080/10543406.2010.481801

Dou, R. N., Liu, S. S., Mo, L. Y., Liu, H. L., & Deng, F. C. (2011). A novel direct equipartition ray design (EquRay) procedure for toxicity interaction between ionic liquid and dichlorvos. *Environmental Science and Pollution Research*, *18*, 734–742. https://doi.org/10.1007/s11356-010-0419-7

Fan, C., & Zhang, D. (2014). Estimating the interaction index in drug combination experiments. *Statistics in Biopharmaceutical Research*, *6*, 144–153. https://doi.org/10.1080/19466315.2013.876446

Fang, H. B., Chen, X., Pei, X. Y., Grant, S., & Tan, M. (2017). Experimental design and statistical

analysis for three-drug combination studies. *Statistical Methods in Medical Research*, *26*, 1261–1280. https://doi.org/10.1177/0962280215574320

Fang, H. B., Ross, D. D., Sausville, E., & Tan, M. (2008). Experimental design and interaction analysis of combination studies of drugs with log-linear dose responses. *Statistics in Medicine*, *27*, 3071–3083. https://doi.org/10.1002/sim.3204

Fang, H. B., Tian, G. L., Li, W., & Tan, M. (2009). Design and sample size for evaluating combinations of drugs of linear and loglinear dose-response curves. *Journal of Biopharmaceutical Statistics*, *19*, 625–640. https://doi.org/10.1080/10543400902964019

Fang, H. B., Yu, T., & Tan, M. (2010). Efficient experiment design and nonparametric modeling of drug interaction. *Frontiers in Bioscience*, *2*, 258–265.

Fang, K.-T., Lin, D. K. J., Winker, P., Zhang, Y., Tong, K., & Kong, H. (2000). Uniform design: Theory and application. *Technometrics*, *42*, 237–248.

Fang, K.-T., Liu, M.-Q., Qin, H., & Zhou, Y.-D. (2018). *Theory and application of uniform experimental designs*. Springer. http://www.springer.com/series/694

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

Gaddum, J.-H. (1940). *Pharmacology*. Oxford University Press.

Geary, N. (2013). Understanding synergy. *American Journal of Physiology - Endocrinology and Metabolism*, *304*, E237–E253. https://doi.org/10.1152/ajpendo.00308.2012

Gennings, C. (1996). Economical designs for detecting and characterizing departure from additivity in mixtures of many chemicals. *Food and Chemical Toxicology*, *34*, 1053–1058.

Gonzales, P. R., Pesesky, M. W., Bouley, R., Ballard, A., Biddy, B. A., Suckow, M. A., Wolter, W. R., Schroeder, V. A., Burnham, C. A. D., Mobashery, S., Chang, M., & Dantas, G. (2015). Synergistic, collaterally sensitive β-lactam combinations suppress resistance in MRSA. *Nature Chemical Biology*, *11*, 855–861. https://doi.org/10.1038/nchembio.1911

Greco, W., Bravo, G., & Parsons, J. C. (1995). The search for synergy: A critical review from a response surface perspective. *Pharmacological Reviews*, *47*, 331–385. https://doi.org/10.1016/S0031-6997(25)06847-4

Greco, W., Unkelbach, H.-D., Pöch, G., Sühnel, J., Kundi, M., & Bödeker, W. (1992). Consensus on concepts and terminology for combined-action assessment: The saariselkä agreement. *Arch. Complex Environ. Stud.*, *4*, 65–69.

Harbron, C. (2010). A flexible unified approach to the analysis of pre-clinical combination studies. *Statistics in Medicine*, *29*, 1746–1756. https://doi.org/10.1002/sim.3916

Hather, G., Chen, H., & Liu, R. (2013). Experimental design for in vitro drug combination studies. In M. Hu, Y. Liu, & J. Lin (Eds.), *Topics in applied statistics: 2012 symposium of the international chinese statistical association* (Vol. 55, pp. 337–344). Springer Proceedings in Mathematics; Statistics. https://doi.org/10.1007/978-1-4614-7846-1

Holland-Letz, T., & Kopp-Schneider, A. (2015). Optimal experimental designs for dose–response studies with continuous endpoints. *Archives of Toxicology*, *89*, 2059–2068. https://doi.org/10.1007/s00204-014-1335-2

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., 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

Huang, H., & Chen, X. (2021). Compromise design for combination experiment of two drugs. *Computational Statistics and Data Analysis*, *157*, 107150. https://doi.org/10.1016/j.csda.2020.107150

Huang, H., & Liu, M. Q. (2024). Two-dimensional maximin power designs for combination experiments of drugs. *Communications in Mathematics and Statistics*. https://doi.org/10.1007/s40304-023-00388-w

Jonker, M. J., Svendsen, C., Bedaux, J. J. M., Bongers, M., & Kammenga, J. E. (2005). Significance testing of synergistic/antagonistic, dose level-dependent, or dose ratio-dependent effects in mixture dose-response analysis. *Environmental Toxicology and Chemistry*, *24*, 2701–2713. https://doi.org/10.1897/04-431R.1

Kortenkamp, A., & Altenburger, R. (2010). Toxicity from combined exposure to chemicals. In C. V. Gestel, M. Jonker, J. Kammenga, R. Laskowski, & C. Svendsen (Eds.), *Mixture toxicity: Linking approaches from ecological and human toxicology* (1st ed., pp. 96–119). CRC Press.

Lederer, S., Dijkstra, T. M. H., & Heskes, T. (2018). Additive dose response models: Explicit formulation and the loewe additivity consistency condition. *Frontiers in Pharmacology*, *9*, 31. https://doi.org/10.3389/fphar.2018.00031

Lederer, S., Dijkstra, T. M. H., & Heskes, T. (2019). Additive dose response models: Defining synergy. *Frontiers in Pharmacology*, *10*, 1384. https://doi.org/10.3389/fphar.2019.01384

Liu, J. Y., & Sayes, C. M. (2024). Modeling mixtures interactions in environmental toxicology. *Environmental Toxicology and Pharmacology*, *106*, 104380. https://doi.org/10.1016/j.etap.2024.104380

Liu, S. S., Xiao, Q. F., Zhang, J., & Yu, M. (2016). Uniform design ray in the assessment of combined toxicities of multi-component mixtures. *Science Bulletin*, *61*, 52–58. https://doi.org/10.1007/s11434-015-0925-6

Loewe, S., & Muischnek, H. (1926). Über kombinationswirkungen. 1. Mitteilung: Hilfsmittel der fragestellung. *Naunyn-Schmiedebergs Arch Pharmakol*, *114*, 313–326.

Makariadou, E., Wang, X., Hein, N., Deresa, N. W., Mutambanengwe, K., Verbist, B., & Thas, O. (2024). Synergy detection: A practical guide to statistical assessment of potential drug combinations. *Pharmaceutical Statistics*. https://doi.org/10.1002/pst.2383

Makhoba, X. H., Viegas, C., Mosa, R. A., Viegas, F. P. D., & Pooe, O. J. (2020). Potential impact of the multi-target drug approach in the treatment of some complex diseases. *Drug Design, Development and Therapy*, *14*, 3235–3249. https://doi.org/10.2147/DDDT.S257494

Malik, M. A., Wani, M. Y., & Hashmi, A. A. (2020). Combination therapy: Current status and future perspectives. In *Combination therapy against multidrug resistance* (pp. 1–38). Elsevier. https://doi.org/10.1016/B978-0-12-820576-1.00001-1

Mao, B., & Guo, S. (2023). Statistical assessment of drug synergy from in vivo combination studies using mouse tumor models. *Cancer Research Communications*, *3*, 2146–2157. https://doi.org/10.1158/2767-9764.CRC-23-0243

Martin, O., Scholze, M., Ermler, S., McPhie, J., Bopp, S. K., Kienzler, A., Parissis, N., & Kortenkamp, A. (2021). Ten years of research on synergisms and antagonisms in chemical mixtures: A systematic review and quantitative reappraisal of mixture studies. *Environment International*, *146*, 106206. https://doi.org/10.1016/j.envint.2020.106206

Martinez-Irujo, J. J., Villahermosa, M. L., Alberdi, E. A., & Santiago, E. (1996). A checkerboard method to evaluate interactions between drugs. *Biochemical Pharmacology*, *51*, 635–644.

Mayer, L., & Janoff, A. (2007). Optimizing combination chemotherapy by controlling drug ratios. *Molecular Interventions*, *7*, 216–223.

Meyer, C. T., Wooten, D. J., Lopez, C. F., & Quaranta, V. (2020). Charting the fragmented landscape

of drug synergy. *Trends in Pharmacological Sciences*, *41*, 266–280. https://doi.org/10.1016/j.tips.2020.01.011

Ning, S., Xu, H., Al-Shyoukh, I., Feng, J., & Sun, R. (2014). An application of a hill-based response surface model for a drug combination experiment on lung cancer. *Statistics in Medicine*, *33*, 4227–4236. https://doi.org/10.1002/sim.6229

Nørgaard, K. B., & Cedergreen, N. (2010). Pesticide cocktails can interact synergistically on aquatic crustaceans. *Environmental Science and Pollution Research*, *17*, 957–967. https://doi.org/10.1007/s11356-009-0284-4

Novick, S. J., & Peterson, J. J. (2014). Drug combination design strategies. In W. Zhao & H. Yang (Eds.), *Statistical methods in drug combination studies* (pp. 41–52). Taylor & Francis.

Palmer, A. C., Chidley, C., & Sorger, P. K. (2019). A curative combination cancer therapy achieves high fractional cell killing through low cross resistance and drug additivity. *eLife*, *8*. https://doi.org/10.7554/eLife.50036

Pearson, R. A., Wicha, S. G., & Okour, M. (2023). Drug combination modeling: Methods and applications in drug development. *Journal of Clinical Pharmacology*, *63*, 151–165. https://doi.org/10.1002/jcph.2128

Pemovska, T., Bigenzahn, J. W., & Superti-Furga, G. (2018). Recent advances in combinatorial drug screening and synergy scoring. *Current Opinion in Pharmacology*, *42*, 102–110. https://doi.org/10.1016/j.coph.2018.07.008

Prichard, M. N., & Shipman, C. (1990). A three-dimensional model to analyze drug-drug interactions. *Antiviral Research*, *14*, 181–206.

Rider, C. V., & Simmons, J. E. (2018). Chemical mixtures and combined chemical and nonchemical stressors: Exposure, toxicity, analysis, and risk. In *Chemical Mixtures and Combined Chemical and Nonchemical Stressors: Exposure, Toxicity, Analysis, and Risk* (pp. 1–556). Springer International Publishing. https://doi.org/10.1007/978-3-319-56234-6

Ritz, C., Streibig, J. C., & Kniss, A. (2021). How to use statistics to claim antagonism and synergism from binary mixture experiments. *Pest Management Science*, *77*, 3890–3899. https://doi.org/10.1002/ps.6348

Rønneberg, L., Cremaschi, A., Hanes, R., Enserink, J. M., & Zucknick, M. (2021). Bayesynergy: Flexible bayesian modelling of synergistic interaction effects in in vitro drug combination experiments. *Briefings in Bioinformatics*, *22*, 1–12. https://doi.org/10.1093/bib/bbab251

Sebaugh, J. L. (2011). Guidelines for accurate EC50/IC50 estimation. *Pharmaceutical Statistics*, *10*, 128–134. https://doi.org/10.1002/pst.426

Sperrin, M., Thygesen, H., Su, T. L., Harbron, C., & Whitehead, A. (2015). Experimental designs for detecting synergy and antagonism between two drugs in a pre-clinical study. *Pharmaceutical Statistics*, *14*, 216–225. https://doi.org/10.1002/pst.1676

Stein, C., Makarewicz, O., Bohnert, J. A., Pfeifer, Y., Kesselmeier, M., Hagel, S., & Pletz, M. W. (2015). Three dimensional checkerboard synergy analysis of colistin, meropenem, tigecycline against multidrug-resistant clinical klebsiella pneumonia isolates. *PLoS ONE*, *10*, e0126479. https://doi.org/10.1371/journal.pone.0126479

Straetemans, R., & Bijnens, L. (2010). Application and review of the separate ray model to investigate interaction effects. *Frontiers in Bioscience*, *2*, 266–278.

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

Sun, D., Gao, W., Hu, H., & Zhou, S. (2022). Why 90 percent of clinical drug development fails and how to improve it? *Acta Pharmaceutica Sinica B*, *12*, 3049–3062. https://doi.org/10.1016/j.apsb.2022.02.002

Tallarida, R. J. (2011). Quantitative methods for assessing drug synergism. *Genes and Cancer*, *2*, 1003–1008. https://doi.org/10.1177/1947601912440575

Tallarida, R. J. (2012). Revisiting the isobole and related quantitative methods for assessing drug synergism. *Journal of Pharmacology and Experimental Therapeutics*, *342*, 2–8. https://doi.org/10.1124/jpet.112.193474

Tallarida, R. J., Kimmel, H., & Holtzman, G. (1997). Theory and statistics of detecting synergism between two active drugs: Cocaine and buprenorphine. *Psychopharmacology*, 378–382.

Tan, M. T., Fang, H. B., Tian, G. L., & Houghton, P. J. (2003). Experimental design and sample size determination for testing synergism in drug combination studies based on uniform measures. *Statistics in Medicine*, *22*, 2091–2100. https://doi.org/10.1002/sim.1467

Tan, M. T., Fang, H. B., & Tian, G.-L. (2009). Dose and sample size determination for multi-drug combination studies. *Statistics in Biopharmaceutical Research*, *1*, 301–316. https://doi.org/10.1198/sbr.2009.0029

Tang, H. X., Liu, S. S., Li, K., & Feng, L. (2016). Combining the uniform design-based

ray procedure with combination index to investigate synergistic lethal toxicities of ternary mixtures on caenorhabditis elegans. *Analytical Methods*, *8*, 4466–4472. https://doi.org/10.1039/c6ay00582a

Tang, J., Wennerberg, K., & Aittokallio, T. (2015). What is synergy? The saariselkä agreement revisited. *Frontiers in Pharmacology*, *6*, 181. https://doi.org/10.3389/fphar.2015.00181

Tekin, E., Savage, V. M., & Yeh, P. J. (2017). Measuring higher-order drug interactions: A review of recent approaches. *Current Opinion in Systems Biology*, *4*, 16–23. https://doi.org/10.1016/j.coisb.2017.05.015

Thas, O., Tourny, A., Verbist, B., Hawinkel, S., Nazarov, M., Mutambanengwe, K., & Bijnens, L. (2022). Statistical detection of synergy: New methods and a comparative study. *Pharmaceutical Statistics*, *21*, 345–360. https://doi.org/10.1002/pst.2173

Tian, G. L., Fang, H. B., Tan, M., Qin, H., & Tang, M. L. (2009). Uniform distributions in a class of convex polyhedrons with applications to drug combination studies. *Journal of Multivariate Analysis*, *100*, 1854–1865. https://doi.org/10.1016/j.jmva.2009.02.011

Vanderborght, K., Tourny, A., Bagdziunas, R., Thas, O., Nazarov, M., Turner, H., Verbist, B., & Ceulemans, H. (2017). BIGL: Biochemically intuitive generalized loewe null model for prediction of the expected combined effect compatible with partial agonism and antagonism. *Scientific Reports*, *7*, 17935. https://doi.org/10.1038/s41598-017-18068-5

Vlot, A. H. C., Aniceto, N., Menden, M. P., Ulrich-Merzenich, G., & Bender, A. (2019). Applying synergy metrics to combination screening data: Agreements, disagreements and pitfalls. *Drug Discovery Today*, *24*, 2286–2298. https://doi.org/10.1016/j.drudis.2019.09.002

White, D., Slocum, H., Brun, Y., Wrzosek, C., & Greco, W. (2005). A new nonlinear mixture response surface paradigm for the study of synergism: A three drug example. *Current Drug Metabolism*, *4*, 399–409. https://doi.org/10.2174/1389200033489316

Wiens, D. P. (1991). Designs for approximately linear regression: Two optimality properties of uniform designs. *Statistics & Probability Letters*, *12*, 217–221.

Wooten, D. J., Meyer, C. T., Lubbock, A. L. R., Quaranta, V., & Lopez, C. F. (2021). MuSyC is a consensus framework that unifies multi-drug synergy metrics for combinatorial drug discovery. *Nature Communications*, *12*. https://doi.org/10.1038/s41467-021-24789-z

Wooten, D. J., Meyer, C. T., Quaranta, V., & Lopez, C. (2019). *A consensus framework unifies multi-drug synergy metrics*. https://doi.org/10.1101/683433

Yadav, B., Wennerberg, K., Aittokallio, T., & Tang, J. (2015). Searching for drug synergy in complex dose-response landscapes using an interaction potency model. *Computational and Structural Biotechnology Journal*, *13*, 504–513. https://doi.org/10.1016/j.csbj.2015.09.001

Yang, Z. (2019). *UniDOE: R package for generating uniform design of experiments*. https://doi.org/https://github.com/ZebinYang/UniDOE

Zhang, J., Liu, S. S., Xiao, Q. F., Huang, X. H., & Chen, Q. (2014). Identifying the component responsible for antagonism within ionic liquid mixtures using the up-to-down procedure integrated with a uniform design ray method. *Ecotoxicology and Environmental Safety*, *107*, 16–21. https://doi.org/10.1016/j.ecoenv.2014.02.024

Zhang, L., & Yang, H. (2016). Design and evaluation of drug combination studies. In L. Zhang (Ed.), *Nonclinical statistics for pharmaceutical and biotechnology* (pp. 349–363). Springer. https://doi.org/10.1007/978-3-319-23558-5_13

Zhang, Y. H., Liu, S. S., Liu, H. L., & Liu, Z. Z. (2010). Evaluation of the combined toxicity of 15 pesticides by uniform design. *Pest Management Science*, *66*, 879–887. https://doi.org/10.1002/ps.1957

Zhang, Y. H., Liu, S. S., Song, X. Q., & Ge, H. L. (2008). Prediction for the mixture toxicity of six organophosphorus pesticides to the luminescent bacterium Q67. *Ecotoxicology and Environmental Safety*, *71*, 880–888. https://doi.org/10.1016/j.ecoenv.2008.01.014

Zhou, Y. D., & Xu, H. (2014). Space-filling fractional factorial designs. *Journal of the American Statistical Association*, *109*, 1134–1144. https://doi.org/10.1080/01621459.2013.873367

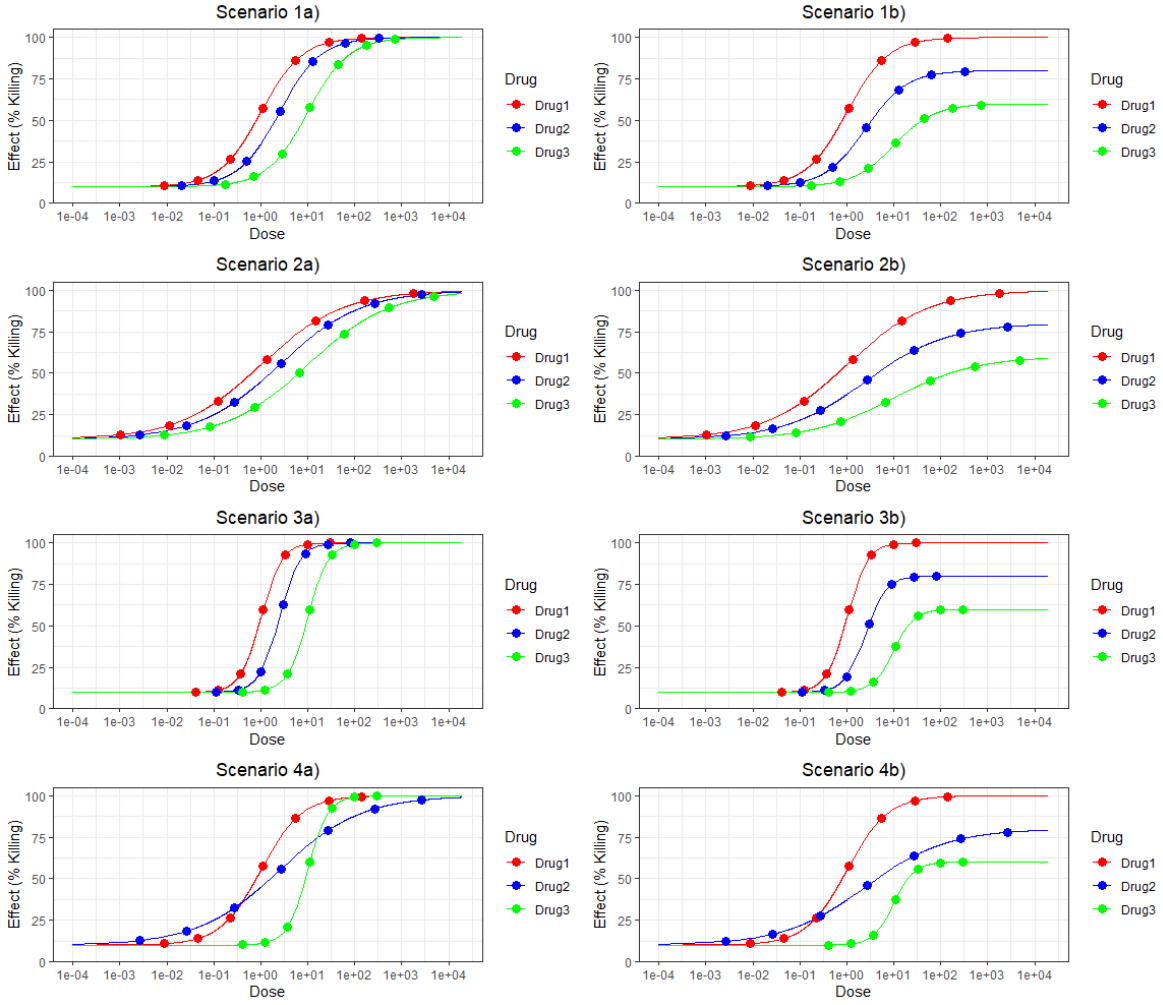
Zimmermann, G. R., Lehár, J., & Keith, C. T. (2007). Multi-target therapeutics: When the whole is greater than the sum of the parts. *Drug Discovery Today*, *12*, 34–42. https://doi.org/10.1016/j.drudis.2006.11.008

Zoli, W., Ricotti, L., Tesei, A., Barzanti, F., & Amadori, D. (2001). In vitro preclinical models for a rational design of chemotherapy combinations in human tumors. *Critical Reviews in Oncology/Hematology*, *37*, 69–82. www.elsevier.com/locate/critrevonc

A Appendix

A.1 Visualization of Simulation Settings

A.1.1 Monotherapy Dose-Response Curves



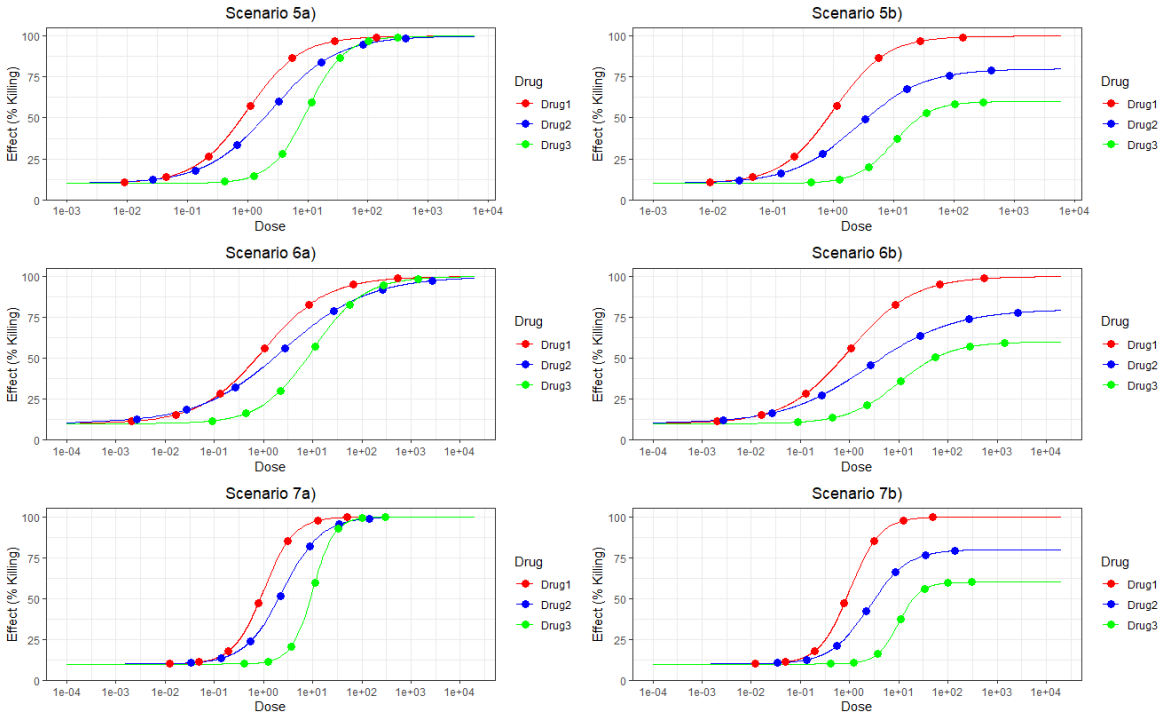


Figure A1: Monotherapy dose-response curves for all scenarios. (a) and (b) only differ in the upper asymptotes, the dilution series remains the same.

A.1.2 Centers of Synergy Tetrahedra

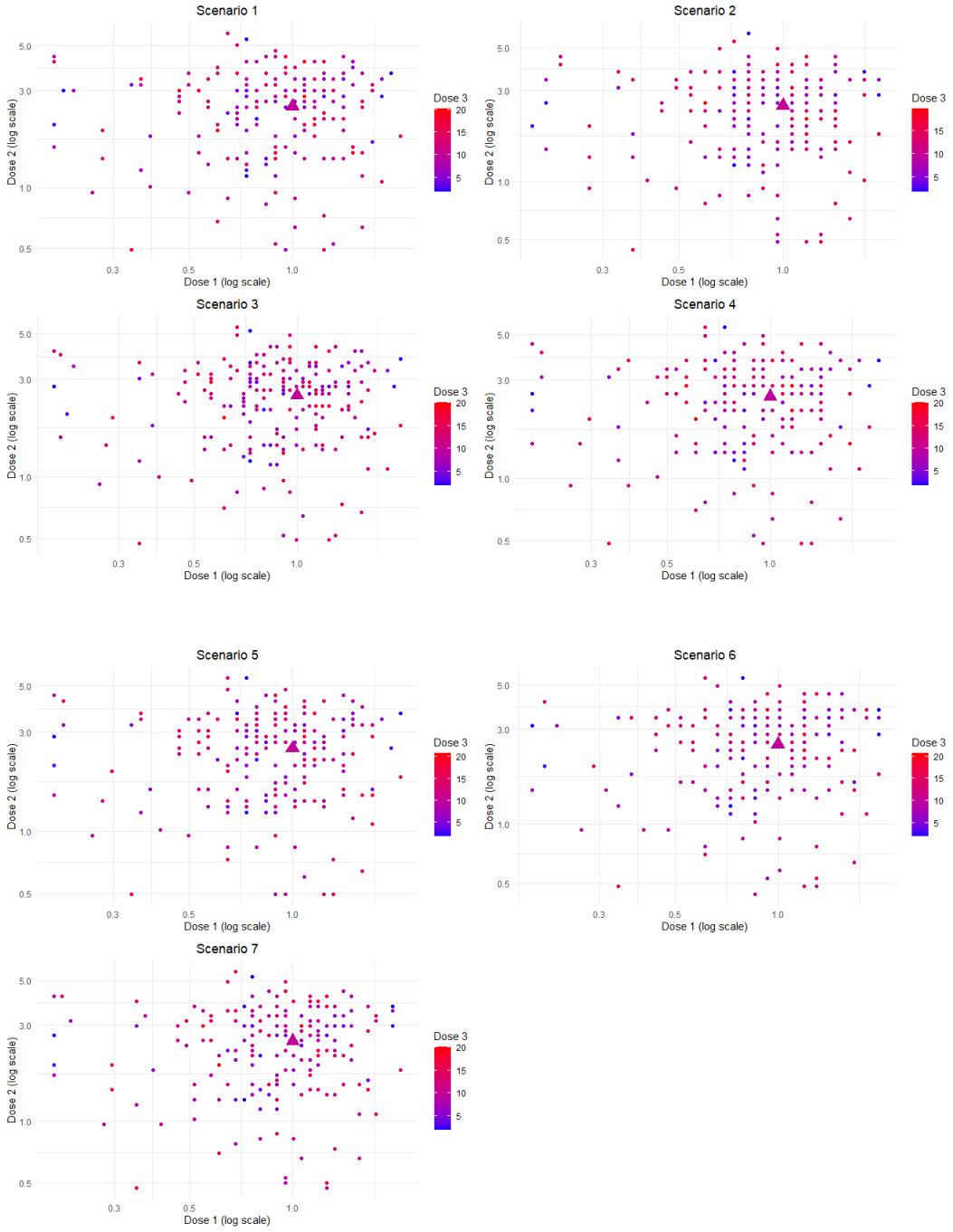
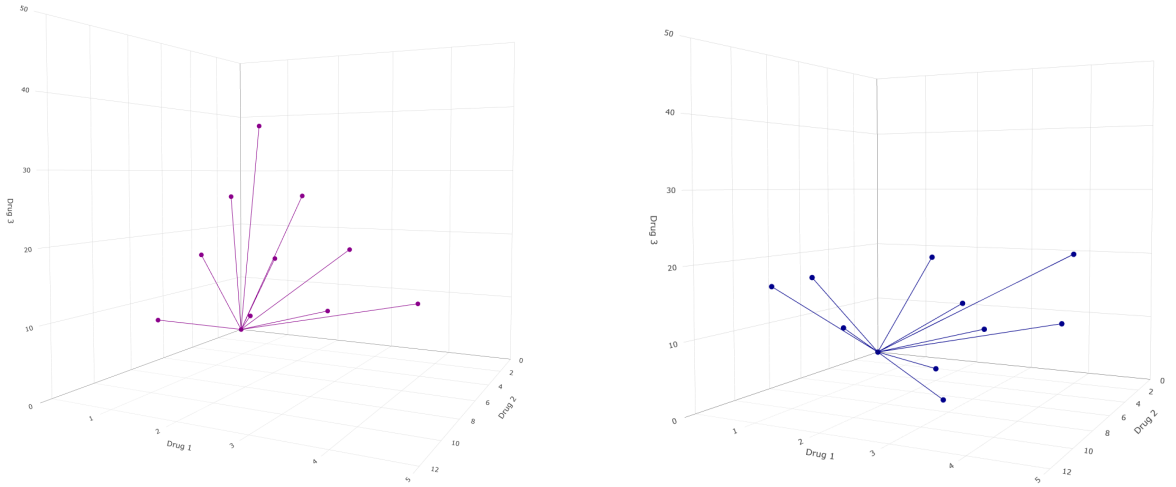


Figure A2: Center coordinates for the synergy tetrahedra for all scenarios. The triangle highlights the EC50 coordinate, around which centers vary.

A.1.3 Designs for Scenario 1a)



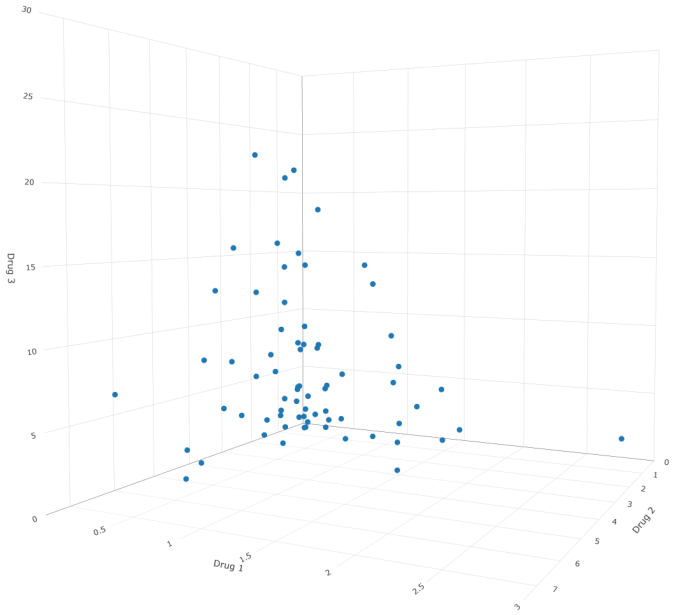
(a) EC50 ray design

(b) Uniform ray design using dose ranges

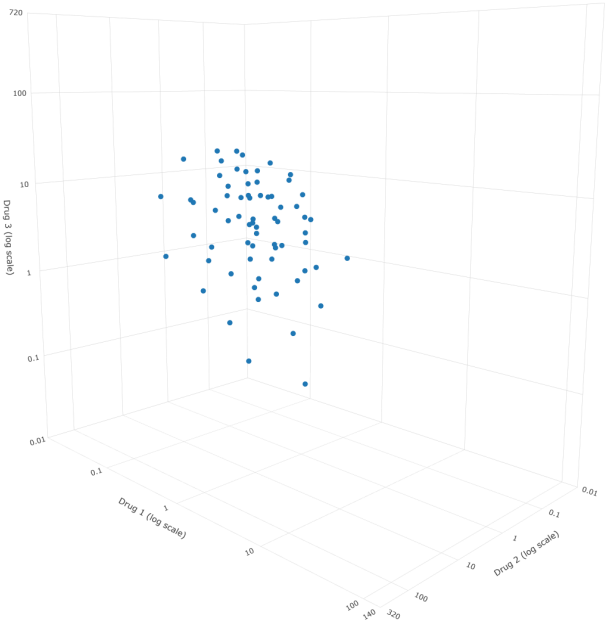
(c) Uniform ray design using factor levels

(d) Combined display of ray designs

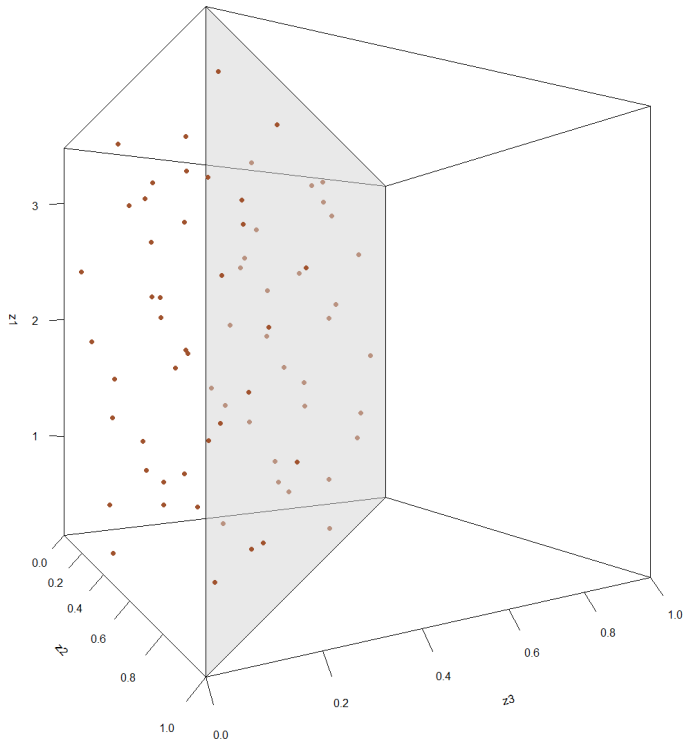
Figure A3: Ray designs for Scenario 1a). The plots use linearly scaled dose axes. Each axis spans 5 times the EC50 of the respective drug for enhanced comparability. Plot **(a)** uses mixture proportions based on the EC50 values. Plot **(b)** uses mixture proportions based on a uniform table, scaled to the dose range. Plot **(c)** uses mixture proportions based on a uniform table, scaled based on a factor-level table with EC20, EC40, EC50, EC60, EC80. Plot **(d)** combines plots **(a)**-**(c)**



(a) Uniform maximum power design on linear dose scale

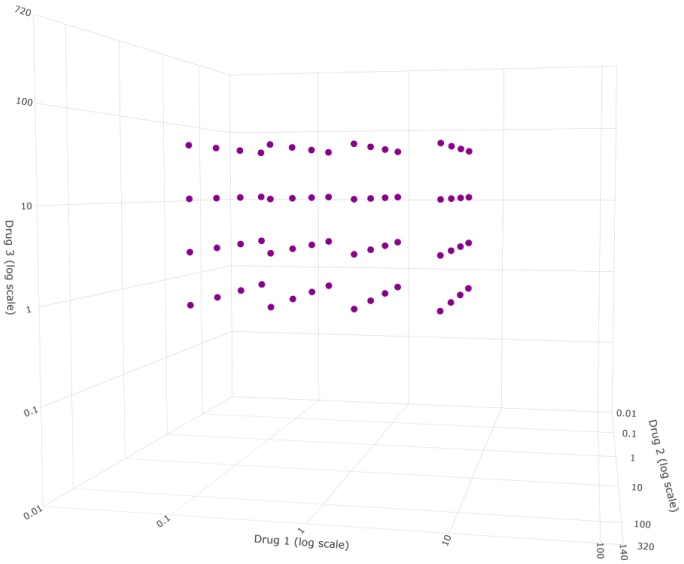


(b) Uniform maximum power design on log dose scale

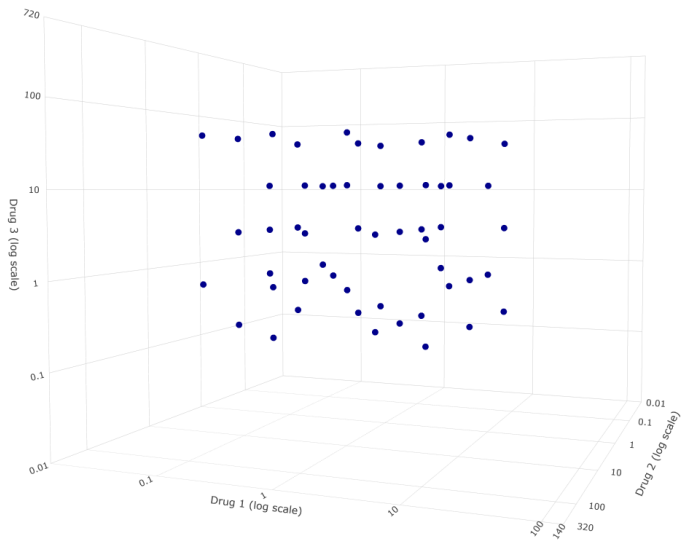


(c) Uniform maximum power design on z-space

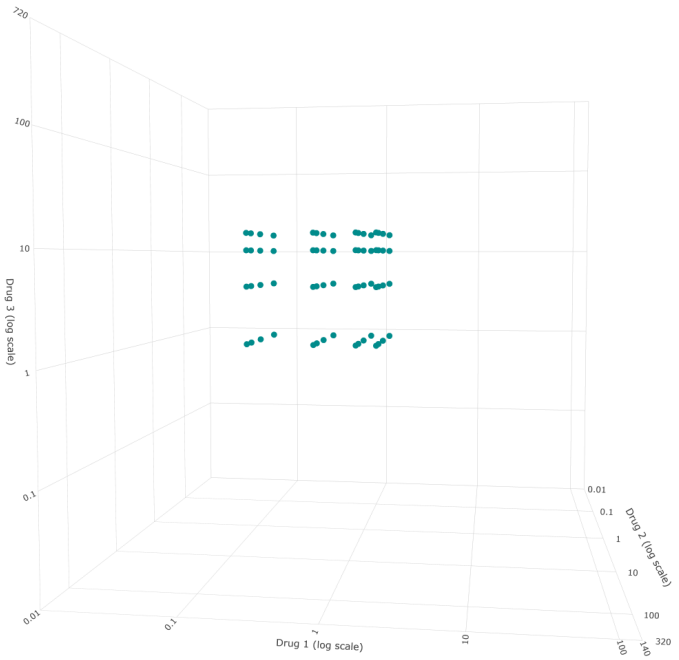
Figure A4: Uniform maximum power design for Scenario 1a). Plot **(a)** uses linearly scaled dose axes, spanning 3 times the EC50 of the respective drug. Plot **(b)** uses log scaled dose axes, spanning the entire dose range. Plot **(c)** visualizes the points on the design space for the transformed variables z, which is a triangular prism defined by the constraint $z_2 + z_3 < 1$.



(a) Checkercube with 4 dose levels around EC50



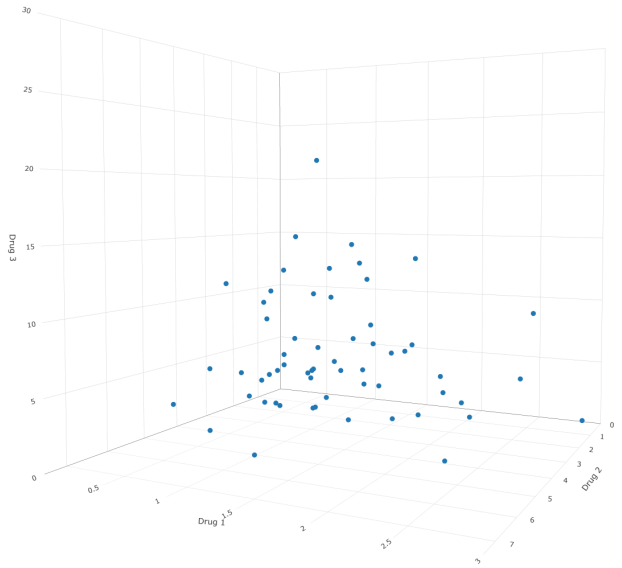
(b) Sampling from checkercube with the 5 lowest dose levels



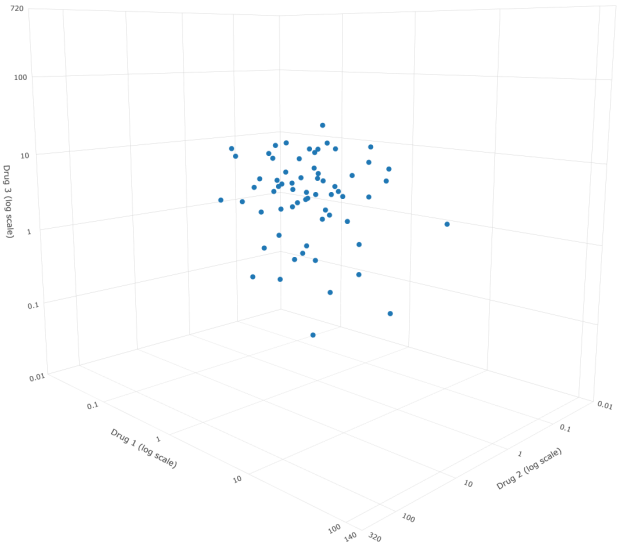
(c) Checkercube with 4 dose levels at EC10, EC30, EC50 and EC60

Figure A5: Checkercube designs for Scenario 1a). Plot **(a)** illustrates a 4x4x4 cube based on the closest 2 monotherapy dilutions above and the closest 2 monotherapy dilutions below the EC50 of each drug. Plot **(b)** samples uniformly from a 5x5x5 cube based on the lowest 5 monotherapy dilutions of each drug. Plot **(c)** is a 4x4x4 cube based on EC10, EC30, EC50, EC60 of each drug. The axes are log scaled and span the entire dose range.

A.1.4 Designs for Scenario 1b)



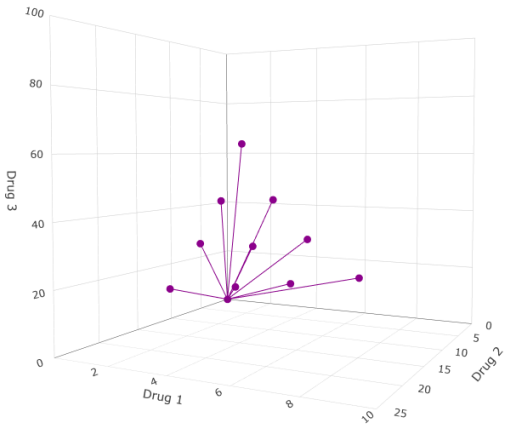
(a) Uniform maximum power design on linear dose scale



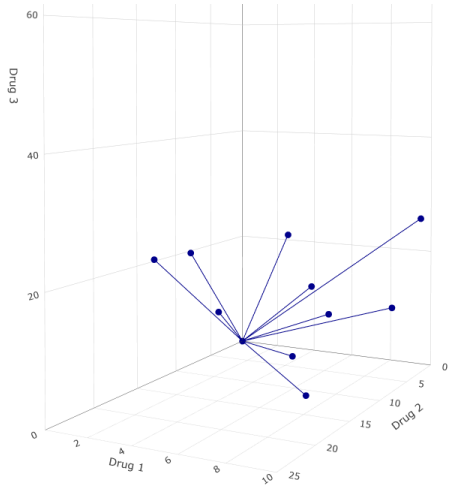
(b) Uniform maximum power design on log dose scale

Figure A6: Uniform maximum power design for Scenario 1b). Plot **(a)** uses linearly scaled dose axes, spanning 3 times the EC50 of the respective drug. Plot **(b)** uses log scaled dose axes, spanning the entire dose range.

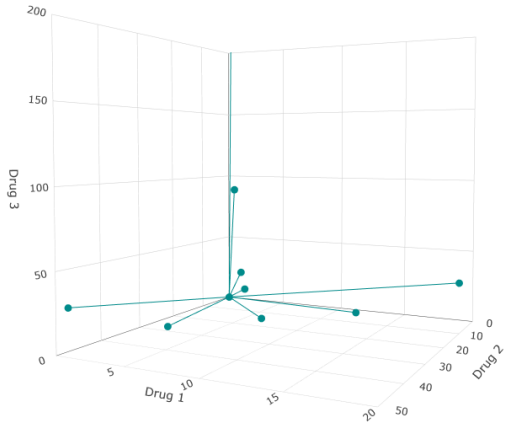
A.1.5 Designs for Scenario 2a)



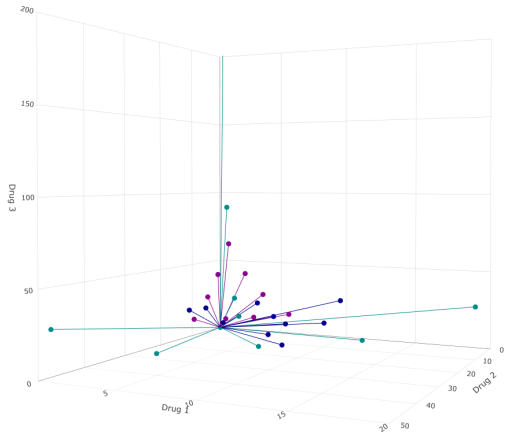
(a) EC50 ray design



(a) Uniform ray design using dose ranges

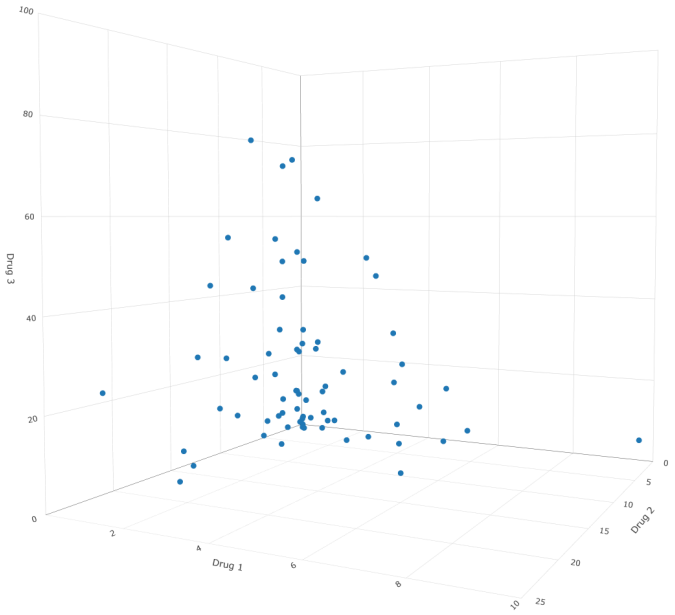


(b) Uniform ray design using factor levels

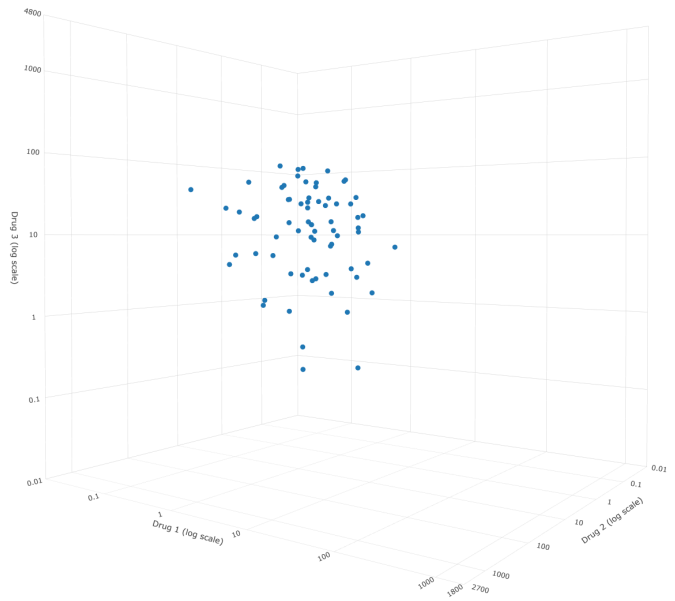


(c) Combined display of ray designs

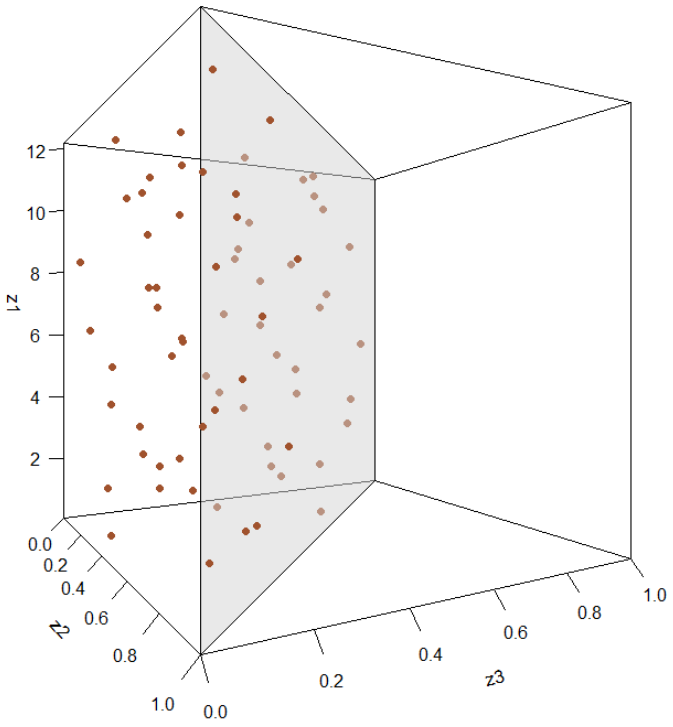
Figure A7: Ray designs for Scenario 2a). The plots use linearly scaled dose axes. The axes span up to 20 times the EC50 of the respective drug. Plot **(a)** uses mixture proportions based on the EC50 values. Plot **(b)** uses mixture proportions based on a uniform table, scaled to the dose range. Plot **(c)** uses mixture proportions based on a uniform table, scaled based on a factor-level table with EC20, EC40, EC50, EC60, EC80. Plot **(d)** combines plots **(a)**-**(c)**



(a) Uniform maximum power design on linear dose scale

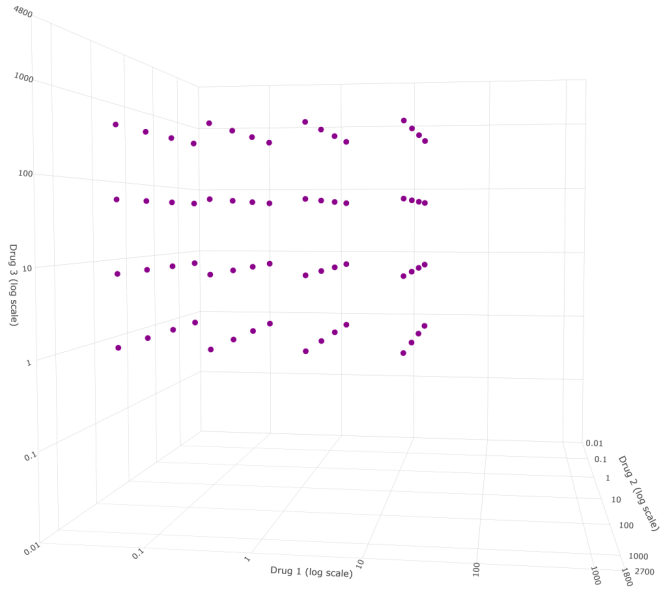


(b) Uniform maximum power design on log dose scale

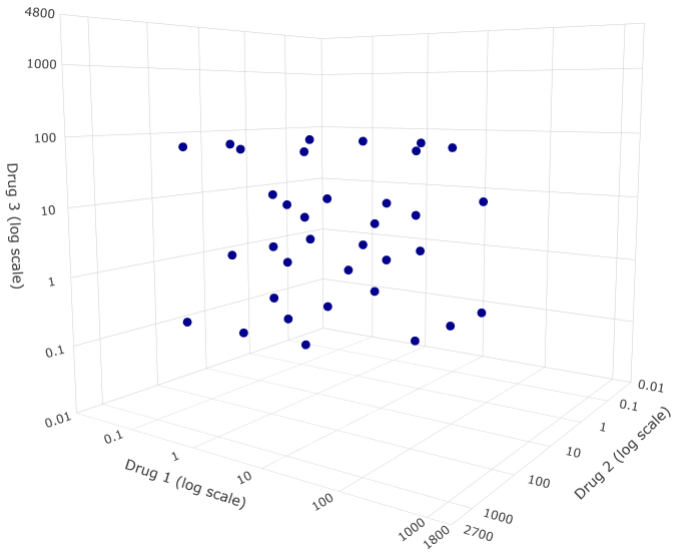


(c) Uniform maximum power design on z-space

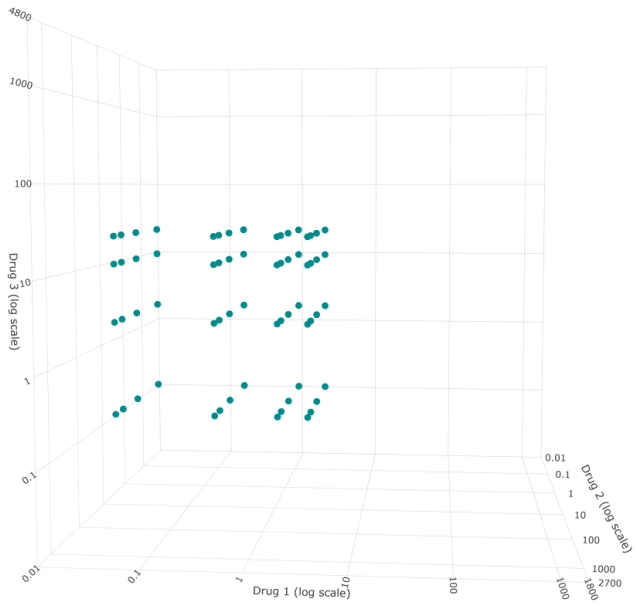
Figure A8: Uniform maximum power design for Scenario 2a). Plot **(a)** uses linearly scaled dose axes, spanning 10 times the EC50 of the respective drug. Plot **(b)** uses log scaled dose axes, spanning the entire dose range. Plot **(c)** visualizes the points on the design space for the transformed variables z, which is a triangular prism defined by the constraint $z_2 + z_3 < 1$.



(a) Checkercube with 4 dose levels around EC50



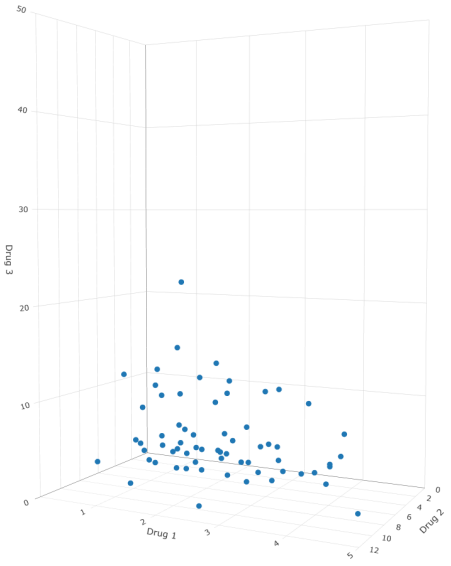
(b) Sampling from checkercube with the 5 lowest dose levels



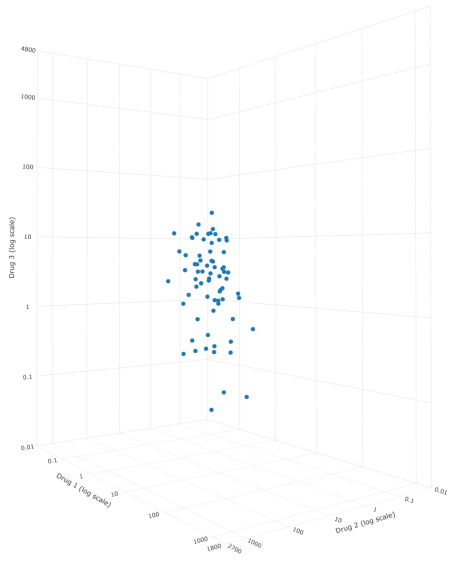
(c) Checkercube with 4 dose levels at EC10, EC30, EC50 and EC60

Figure A9: Checkercube designs for Scenario 2a). Plot **(a)** illustrates a 4x4x4 cube based on the closest 2 monotherapy dilutions above and the closest 2 monotherapy dilutions below the EC50 of each drug. Plot **(b)** samples uniformly from a 5x5x5 cube based on the lowest 5 monotherapy dilutions of each drug. Plot **(c)** is a 4x4x4 cube based on EC10, EC30, EC50, EC60 of each drug. The axes are log scaled and span the entire dose range.

A.1.6 Designs for Scenario 2b)



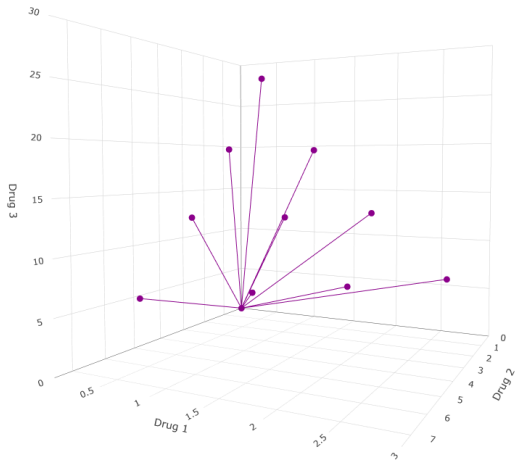
(a) Uniform maximum power design on linear dose scale



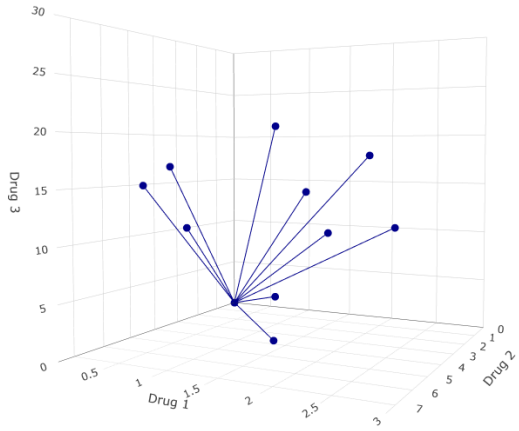
(b) Uniform maximum power design on log dose scale

Figure A10: Uniform maximum power design for Scenario 2b). Plot **(a)** uses linearly scaled dose axes, spanning 5 times the EC50 of the respective drug. Plot **(b)** uses log scaled dose axes, spanning the entire dose range.

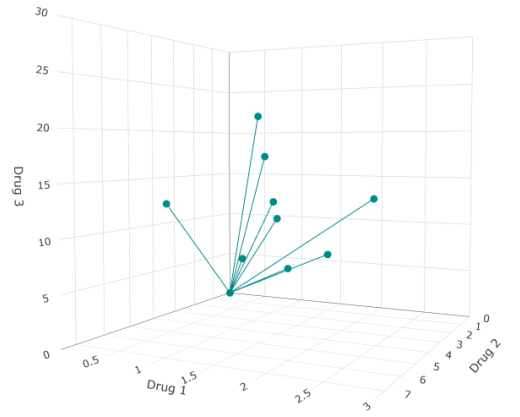
A.1.7 Designs for Scenario 3a)



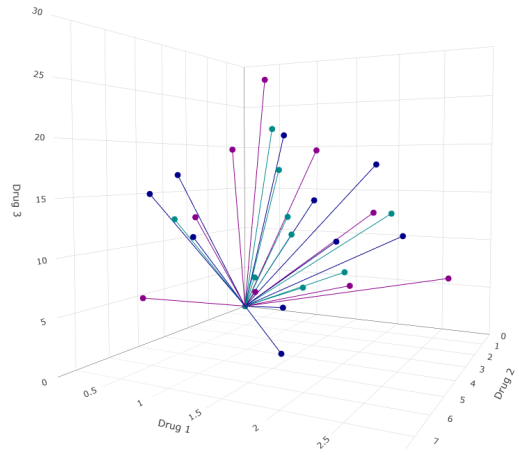
(a) EC50 ray design



(a) Uniform ray design using dose ranges

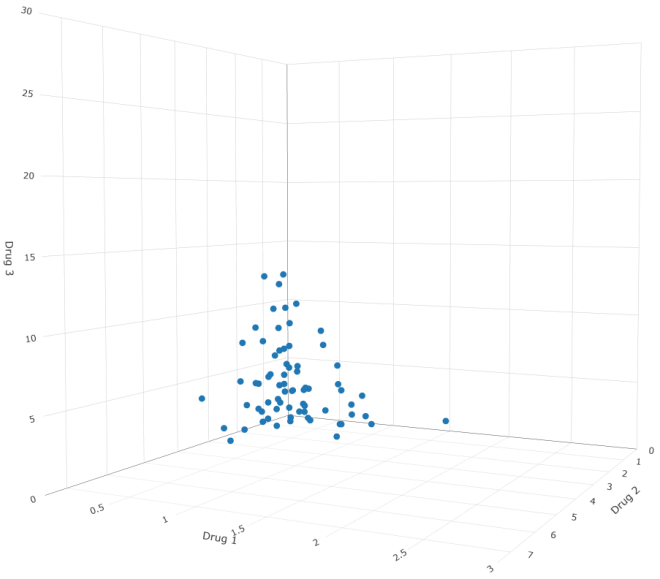


(b) Uniform ray design using factor levels

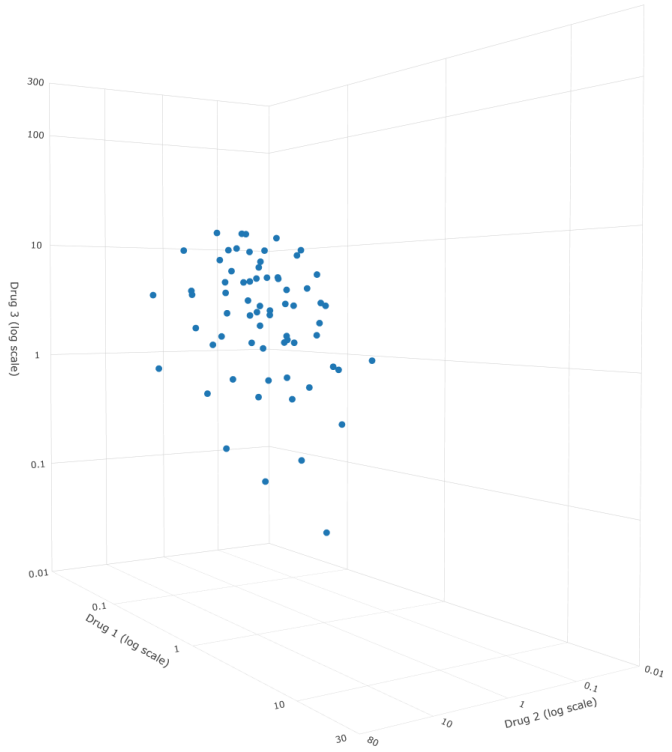


(c) Combined display of ray designs

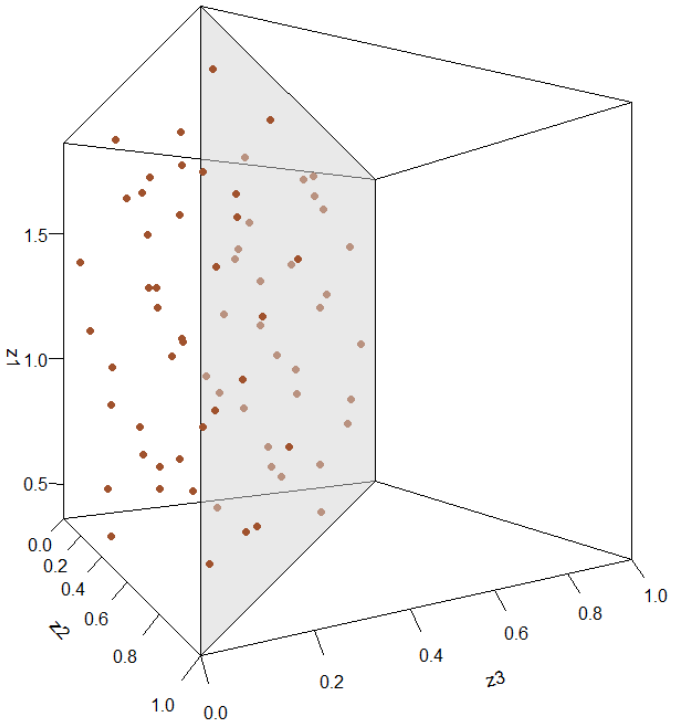
Figure A11: Ray designs for Scenario 3a). The plots use linearly scaled dose axes. Each axis spans 3 times the EC50 of the respective drug for enhanced comparability. Plot **(a)** uses mixture proportions based on the EC50 values. Plot **(b)** uses mixture proportions based on a uniform table, scaled to the dose range. Plot **(c)** uses mixture proportions based on a uniform table, scaled based on a factor-level table with EC20, EC40, EC50, EC60, EC80. Plot **(d)** combines plots **(a)**-**(c)**



(a) Uniform maximum power design on linear dose scale

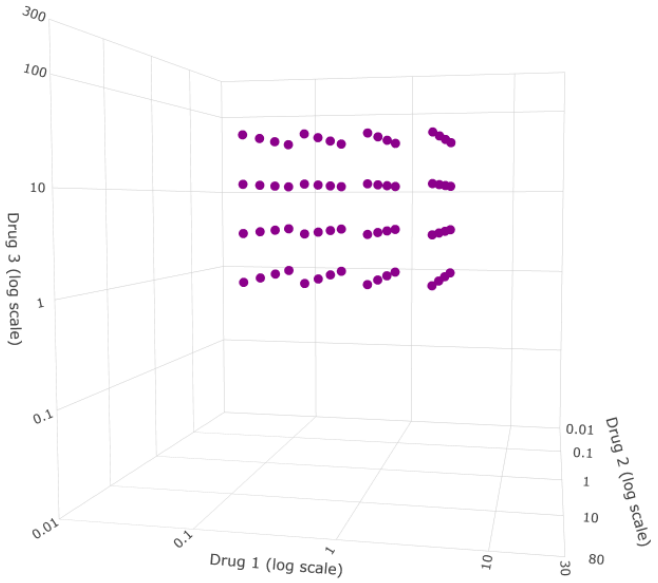


(b) Uniform maximum power design on log dose scale

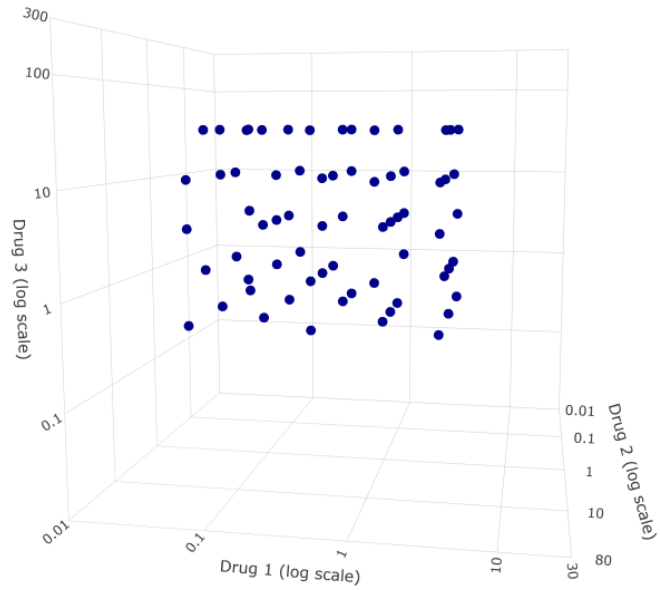


(c) Uniform maximum power design on z-space

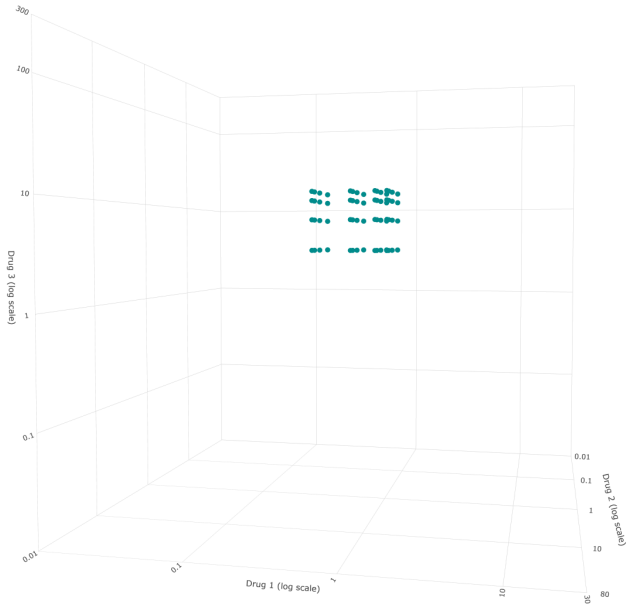
Figure A12: Uniform maximum power design for Scenario 3a). Plot **(a)** uses linearly scaled dose axes, spanning 3 times the EC50 of the respective drug. Plot **(b)** uses log scaled dose axes, spanning the entire dose range. Plot **(c)** visualizes the points on the design space for the transformed variables z, which is a triangular prism defined by the constraint $z_2 + z_3 < 1$.



(a) Checkercube with 4 dose levels around EC50



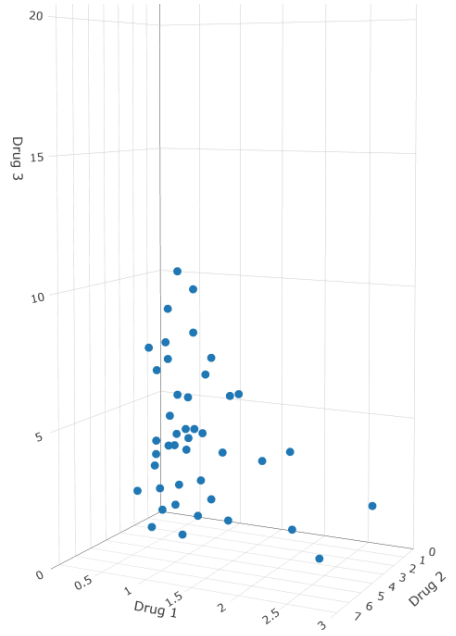
(b) Sampling from checkercube with the 5 lowest dose levels



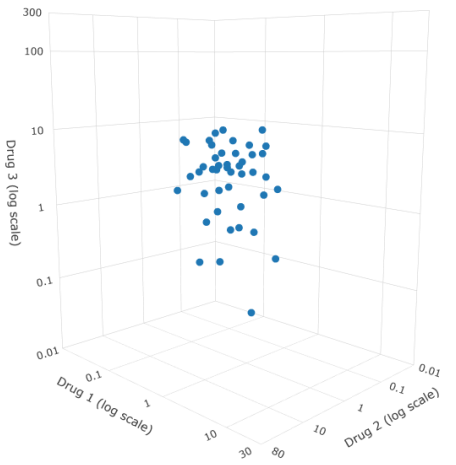
(c) Checkercube with 4 dose levels at EC10, EC30, EC50 and EC60

Figure A13: Checkercube designs for Scenario 3a). Plot **(a)** illustrates a 4x4x4 cube based on the closest 2 monotherapy dilutions above and the closest 2 monotherapy dilutions below the EC50 of each drug. Plot **(b)** samples uniformly from a 5x5x5 cube based on the lowest 5 monotherapy dilutions of each drug. Plot **(c)** is a 4x4x4 cube based on EC10, EC30, EC50, EC60 of each drug. The axes are log scaled and span the entire dose range.

A.1.8 Designs for Scenario 3b)



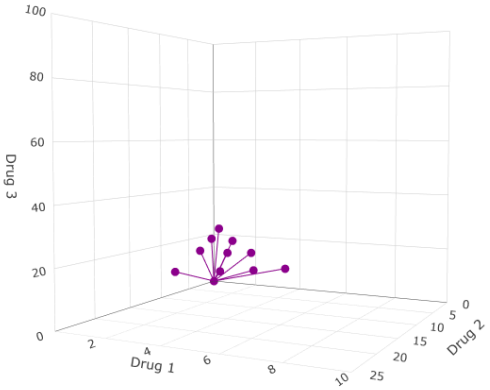
(a) Uniform maximum power design on linear dose scale



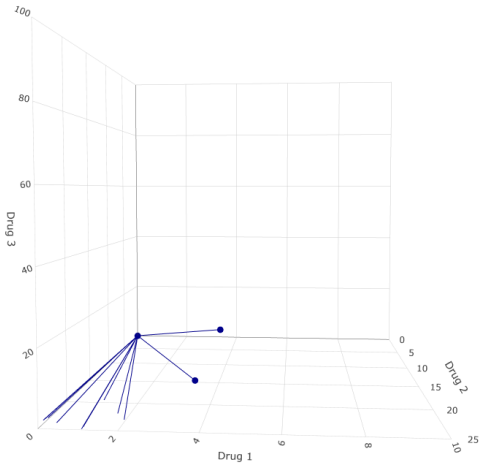
(b) Uniform maximum power design on log dose scale

Figure A14: Uniform maximum power design for Scenario 3b). Plot **(a)** uses linearly scaled dose axes, spanning up to 3 times the EC50 of the respective drug. Plot **(b)** uses log scaled dose axes, spanning the entire dose range.

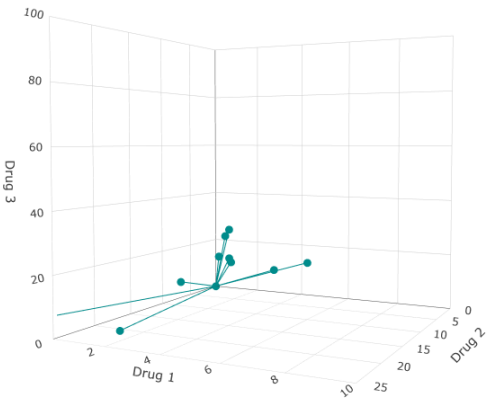
A.1.9 Designs for Scenario 4a)



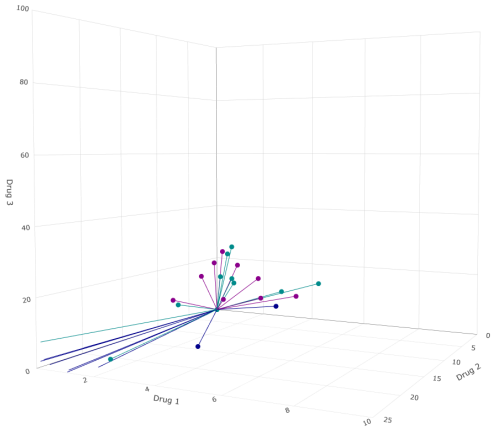
(a) EC50 ray design



(a) Uniform ray design using dose ranges

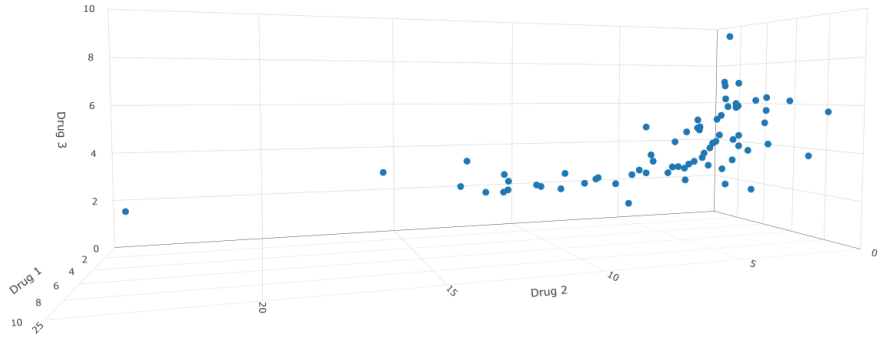


(b) Uniform ray design using factor levels

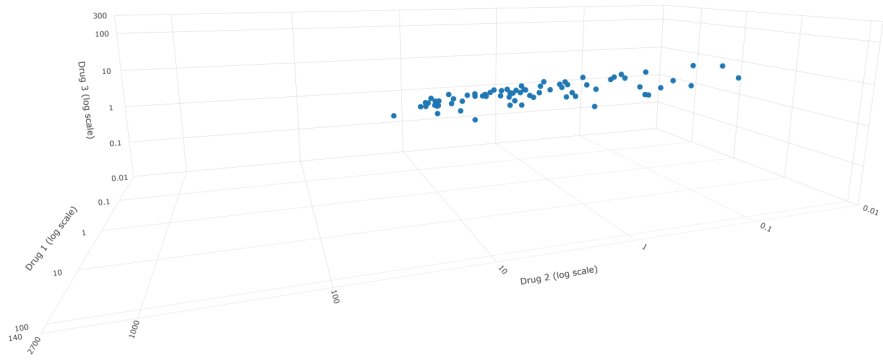


(c) Combined display of ray designs

Figure A15: Ray designs for Scenario 4a). The plots use linearly scaled dose axes. Each axis spans 10 times the EC50 of the respective drug for enhanced comparability. Plot **(a)** uses mixture proportions based on the EC50 values. Plot **(b)** uses mixture proportions based on a uniform table, scaled to the dose range. Plot **(c)** uses mixture proportions based on a uniform table, scaled based on a factor-level table with EC20, EC40, EC50, EC60, EC80. Plot **(d)** combines plots **(a)**-**(c)**

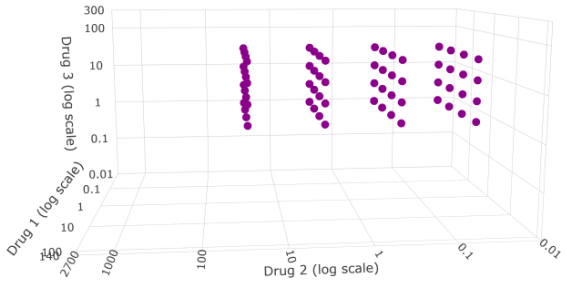


(a) Uniform maximum power design on linear dose scale

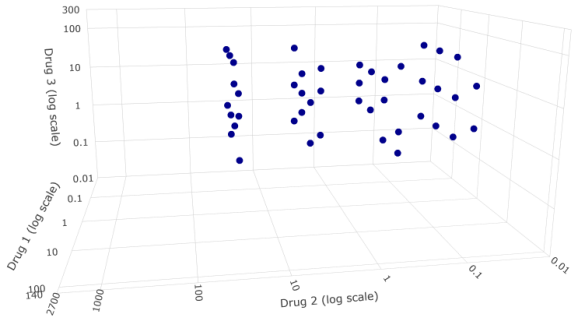


(b) Uniform maximum power design on log dose scale

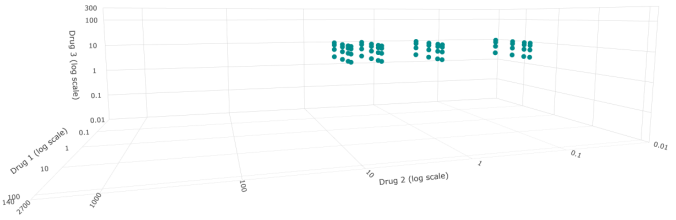
Figure A16: Uniform maximum power design for Scenario 4a). Plot **(a)** uses linearly scaled dose axes, spanning 10 times the EC50 of drugs 1 and 2 but only 1 time the EC50 of drug 3. Plot **(b)** uses log scaled dose axes, spanning the entire dose range.



(a) Checkercube with 4 dose levels around EC50



(b) Sampling from checkercube with the 5 lowest dose levels



(c) Checkercube with 4 dose levels at EC10, EC30, EC50 and EC60

Figure A17: Checkercube designs for Scenario 4a). Plot **(a)** illustrates a 4x4x4 cube based on the closest 2 monotherapy dilutions above and the closest 2 monotherapy dilutions below the EC50 of each drug. Plot **(b)** samples uniformly from a 5x5x5 cube based on the lowest 5 monotherapy dilutions of each drug. Plot **(c)** is a 4x4x4 cube based on EC10, EC30, EC50, EC60 of each drug. The axes are log scaled and span the entire dose range.

A.1.10 Designs for Scenario 4b)

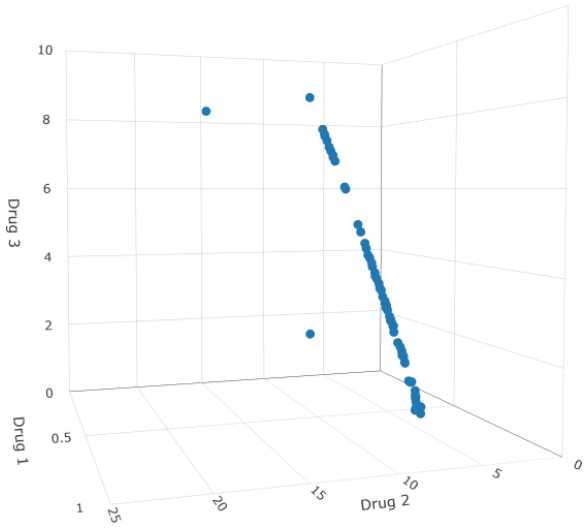
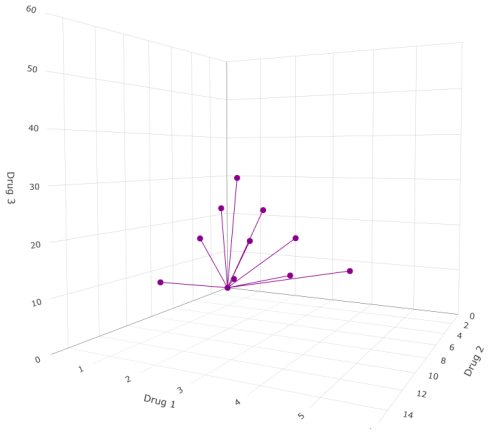
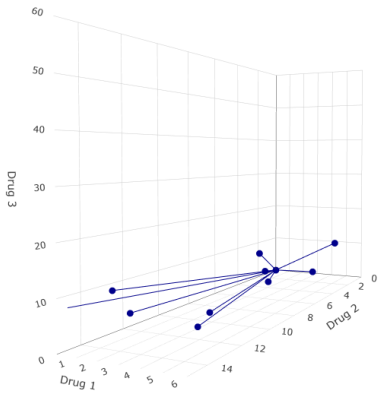


Figure A18: Uniform maximum power design for Scenario 4b). Plot uses linearly scaled dose axes.

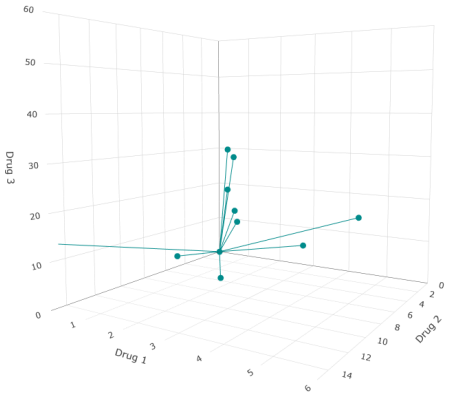
A.1.11 Designs for Scenario 5a)



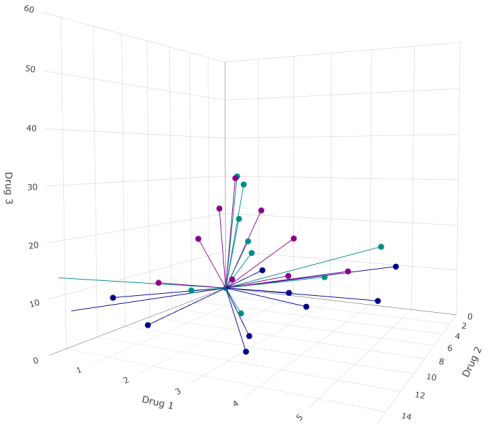
(a) EC50 ray design



(a) Uniform ray design using dose ranges

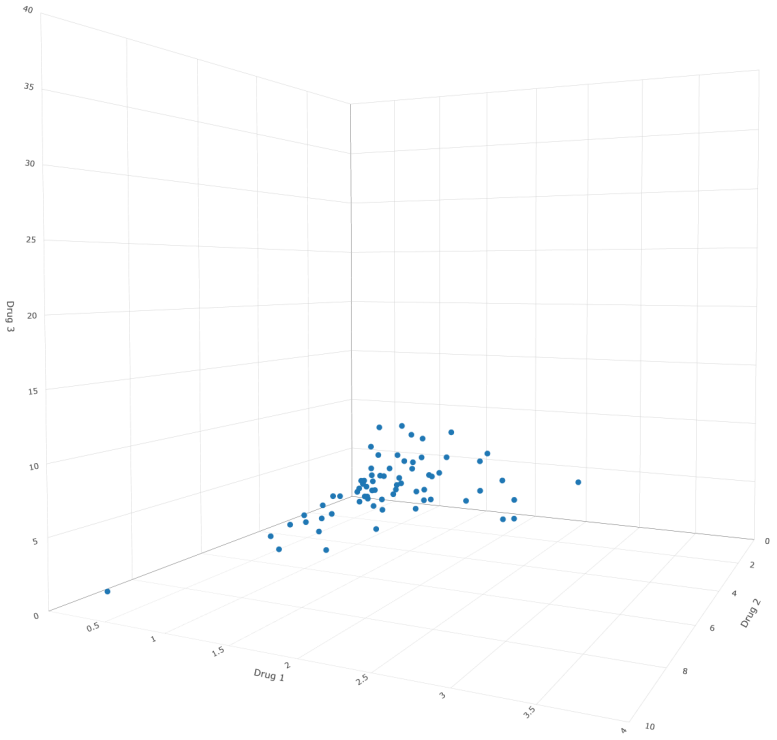


(b) Uniform ray design using factor levels

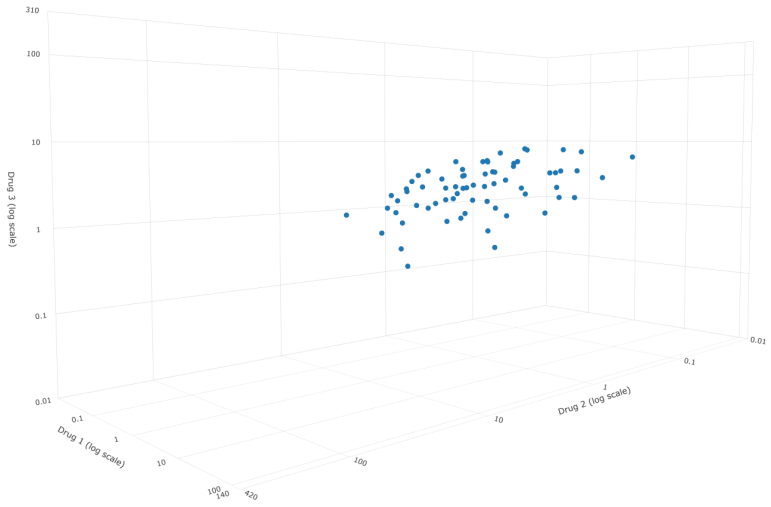


(c) Combined display of ray designs

Figure A19: Ray designs for Scenario 5a). The plots use linearly scaled dose axes. Each axis spans 6 times the EC50 of the respective drug for enhanced comparability. Plot **(a)** uses mixture proportions based on the EC50 values. Plot **(b)** uses mixture proportions based on a uniform table, scaled to the dose range. Plot **(c)** uses mixture proportions based on a uniform table, scaled based on a factor-level table with EC20, EC40, EC50, EC60, EC80. Plot **(d)** combines plots **(a)**-**(c)**

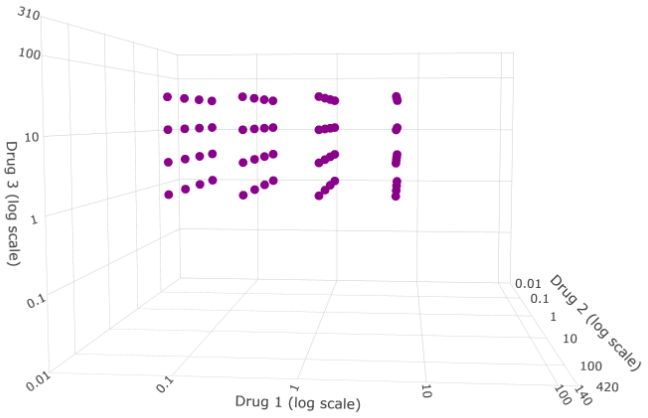


(a) Uniform maximum power design on linear dose scale

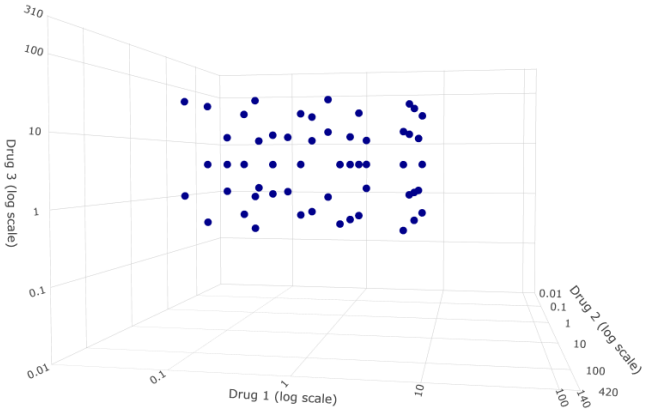


(b) Uniform maximum power design on log dose scale

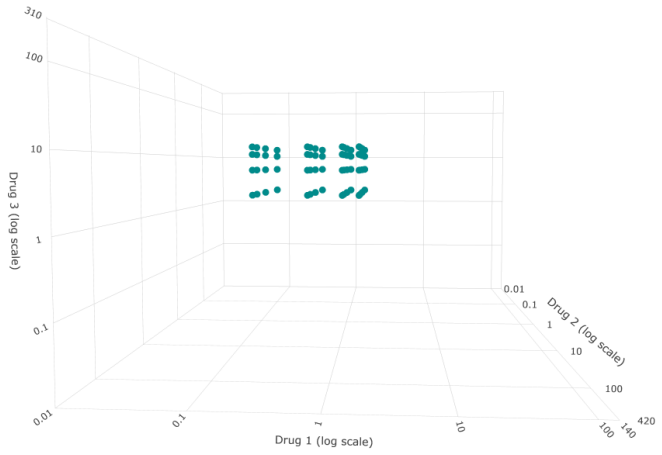
Figure A20: Uniform maximum power design for Scenario 5a). Plot **(a)** uses linearly scaled dose axes, spanning 4 times the EC50 of the respective drug. Plot **(b)** uses log scaled dose axes, spanning the entire dose range.



(a) Checkercube with 4 dose levels around EC50



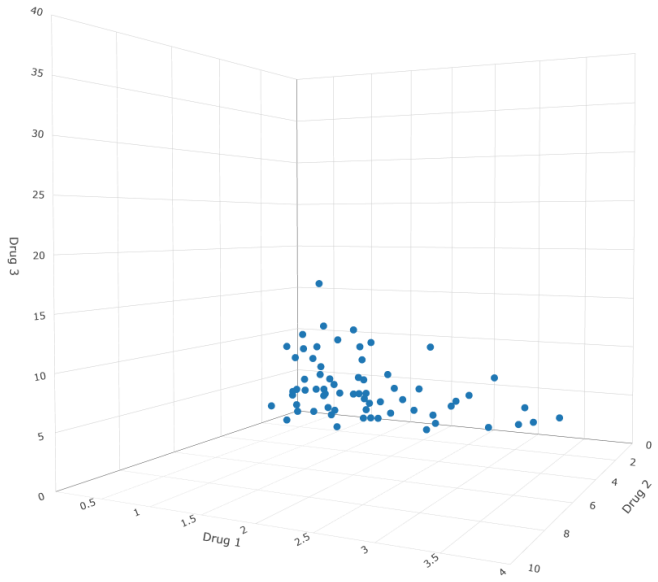
(b) Sampling from checkercube with the 5 lowest dose levels



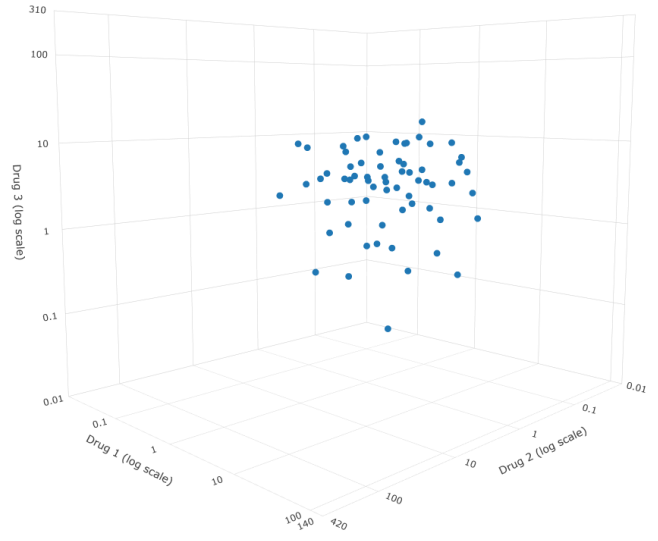
(c) Checkercube with 4 dose levels at EC10, EC30, EC50 and EC60

Figure A21: Checkercube designs for Scenario 5a). Plot **(a)** illustrates a 4x4x4 cube based on the closest 2 monotherapy dilutions above and the closest 2 monotherapy dilutions below the EC50 of each drug. Plot **(b)** samples uniformly from a 5x5x5 cube based on the lowest 5 monotherapy dilutions of each drug. Plot **(c)** is a 4x4x4 cube based on EC10, EC30, EC50, EC60 of each drug. The axes are log scaled and span the entire dose range.

A.1.12 Designs for Scenario 5b)



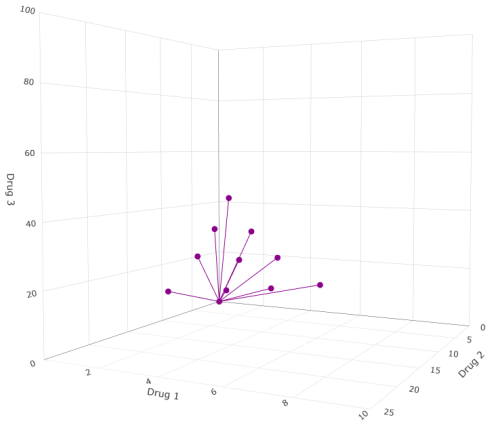
(a) Uniform maximum power design on linear dose scale



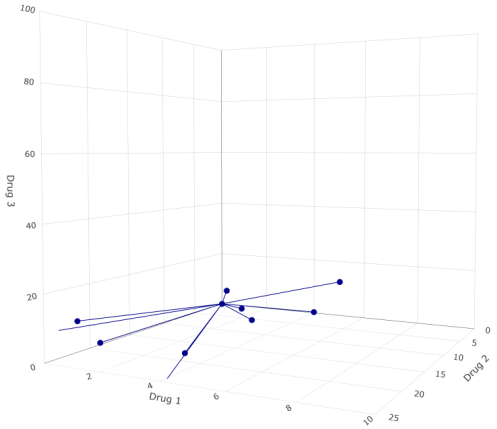
(b) Uniform maximum power design on log dose scale

Figure A22: Uniform maximum power design for Scenario 5b). Plot **(a)** uses linearly scaled dose axes, spanning 4 times the EC50 of the respective drug. Plot **(b)** uses log scaled dose axes, spanning the entire dose range.

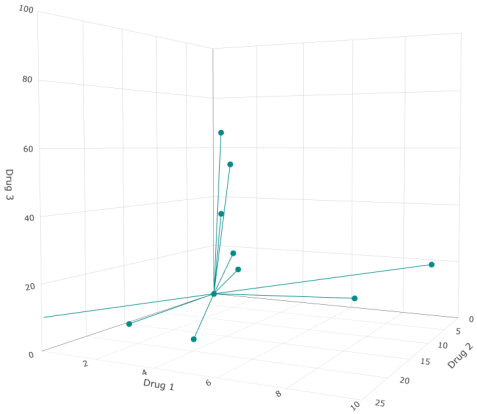
A.1.13 Designs for Scenario 6a)



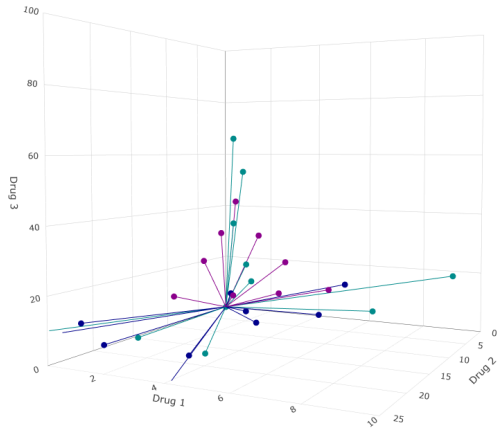
(a) EC50 ray design



(a) Uniform ray design using dose ranges

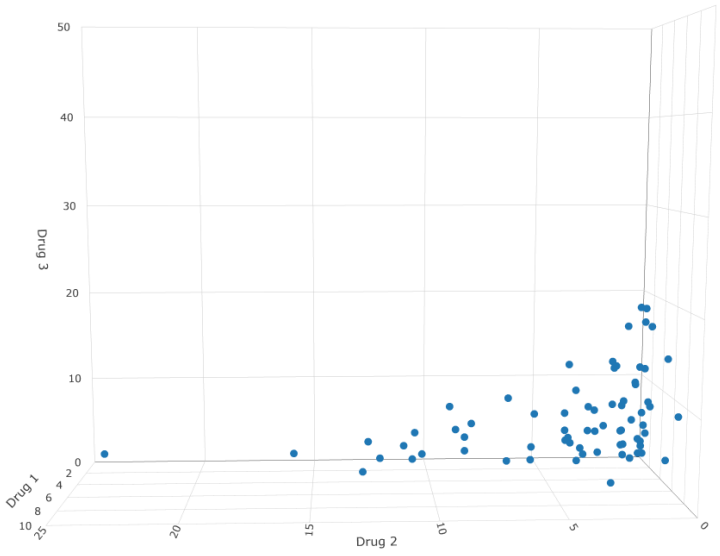


(b) Uniform ray design using factor levels

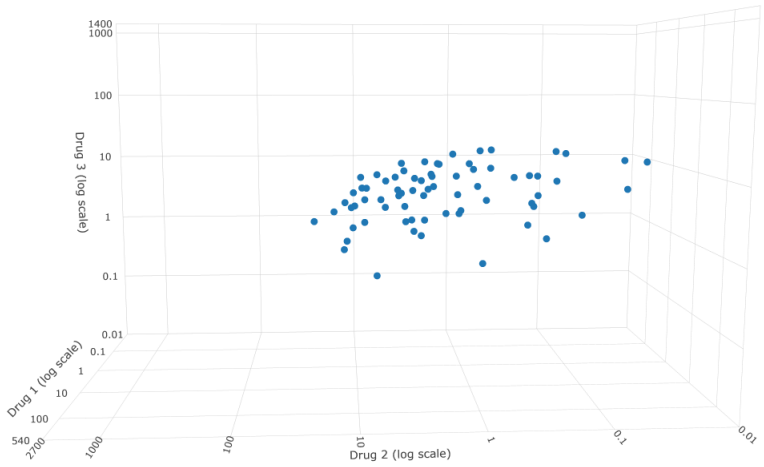


(c) Combined display of ray designs

Figure A23: Ray designs for Scenario 6a). The plots use linearly scaled dose axes. Each axis spans 10 times the EC50 of the respective drug for enhanced comparability. Plot **(a)** uses mixture proportions based on the EC50 values. Plot **(b)** uses mixture proportions based on a uniform table, scaled to the dose range. Plot **(c)** uses mixture proportions based on a uniform table, scaled based on a factor-level table with EC20, EC40, EC50, EC60, EC80. Plot **(d)** combines plots **(a)**-**(c)**

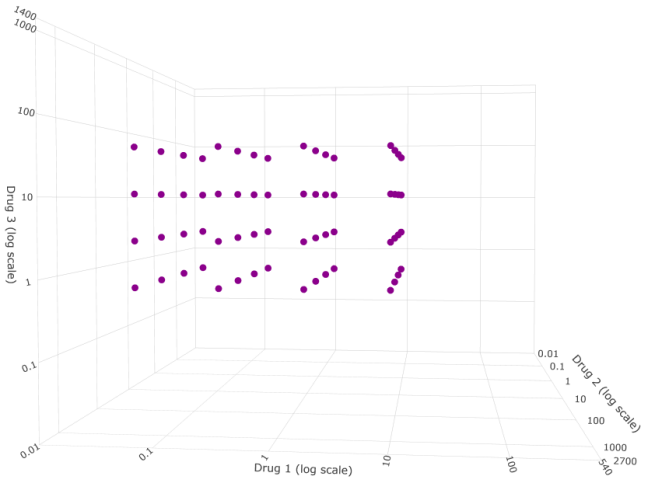


(a) Uniform maximum power design on linear dose scale

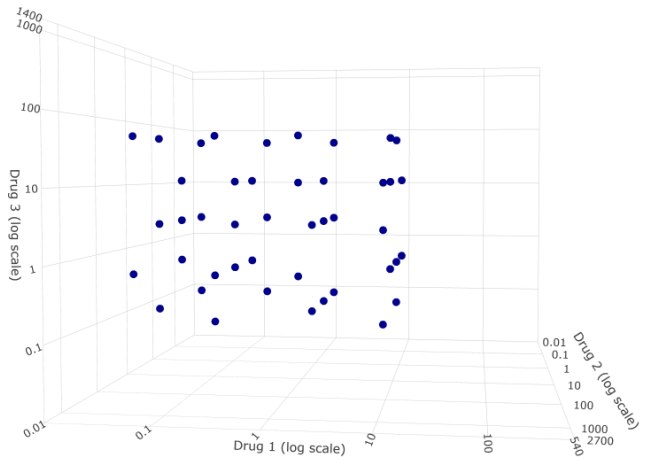


(b) Uniform maximum power design on log dose scale

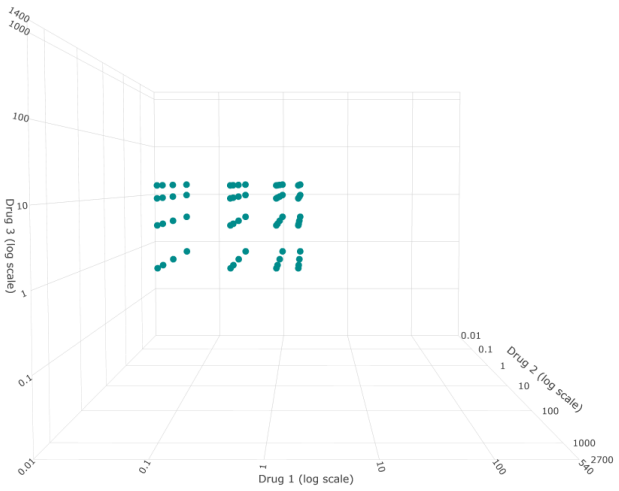
Figure A24: Uniform maximum power design for Scenario 6a). Plot **(a)** uses linearly scaled dose axes, spanning 10 times the EC50 of of drug 2 and drug 3 but only 5 times the EC50 of drug 3. Plot **(b)** uses log scaled dose axes, spanning the entire dose range.



(a) Checkercube with 4 dose levels around EC50



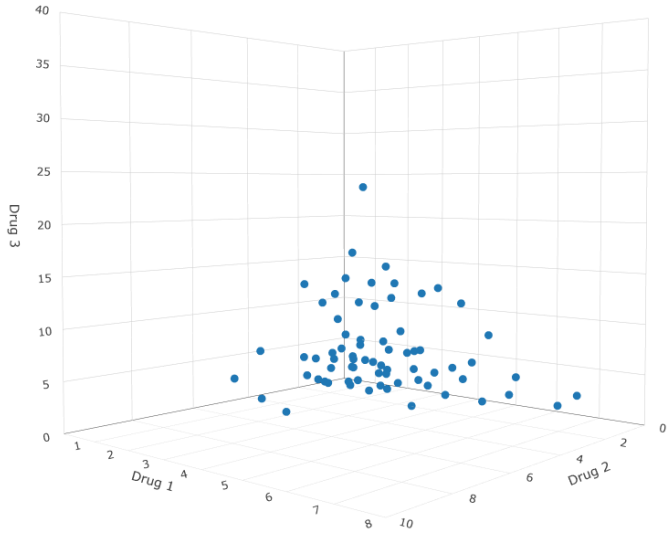
(b) Sampling from checkercube with the 5 lowest dose levels



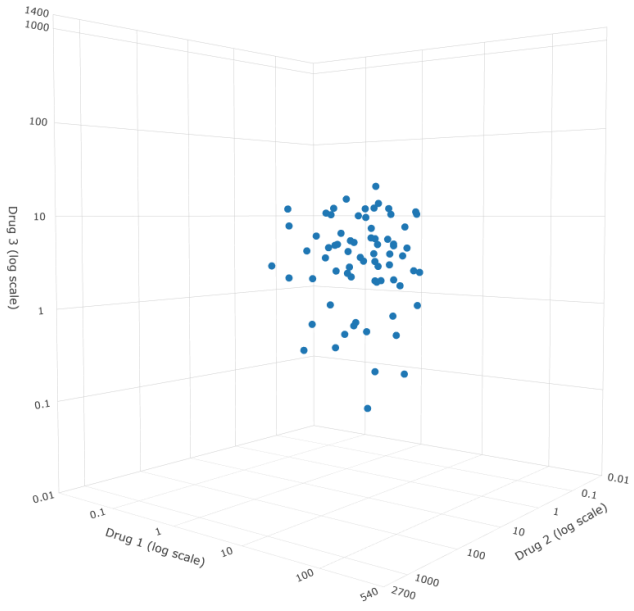
(c) Checkercube with 4 dose levels at EC10, EC30, EC50 and EC60

Figure A25: Checkercube designs for Scenario 6a). Plot **(a)** illustrates a 4x4x4 cube based on the closest 2 monotherapy dilutions above and the closest 2 monotherapy dilutions below the EC50 of each drug. Plot **(b)** samples uniformly from a 5x5x5 cube based on the lowest 5 monotherapy dilutions of each drug. Plot **(c)** is a 4x4x4 cube based on EC10, EC30, EC50, EC60 of each drug. The axes are log scaled and span the entire dose range.

A.1.14 Designs for Scenario 6b)



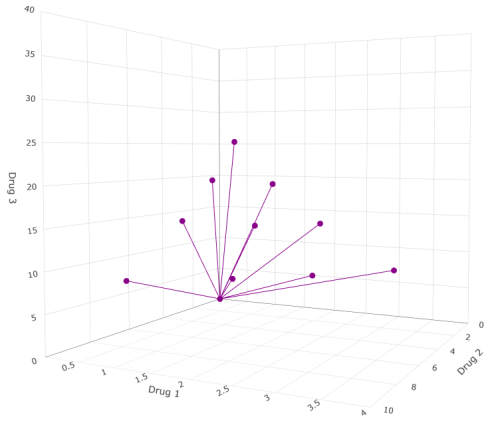
(a) Uniform maximum power design on linear dose scale



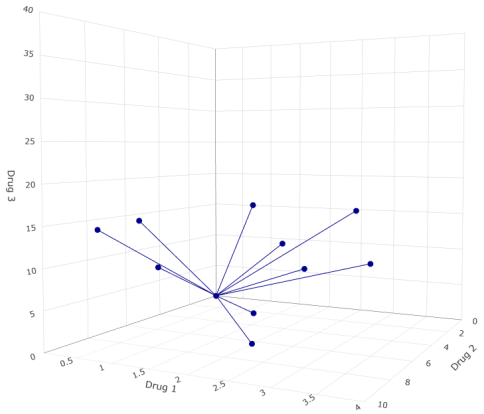
(b) Uniform maximum power design on log dose scale

Figure A26: Uniform maximum power design for Scenario 6b). Plot **(a)** uses linearly scaled dose axes, spanning 4 times the EC50 of drug 2 and drug 3 and 8 times the EC50 of drug 1. Plot **(b)** uses log scaled dose axes, spanning the entire dose range.

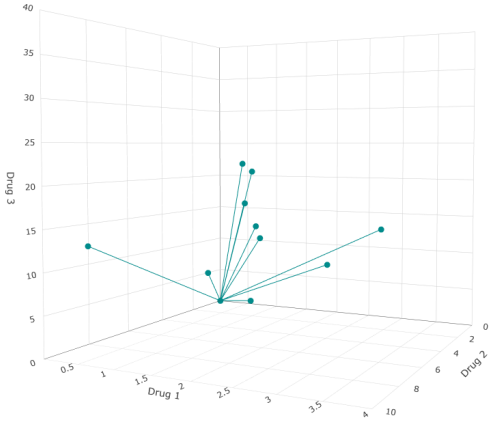
A.1.15 Designs for Scenario 7a)



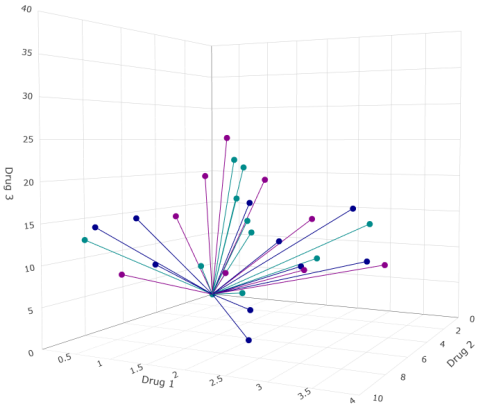
(a) EC50 ray design



(a) Uniform ray design using dose ranges

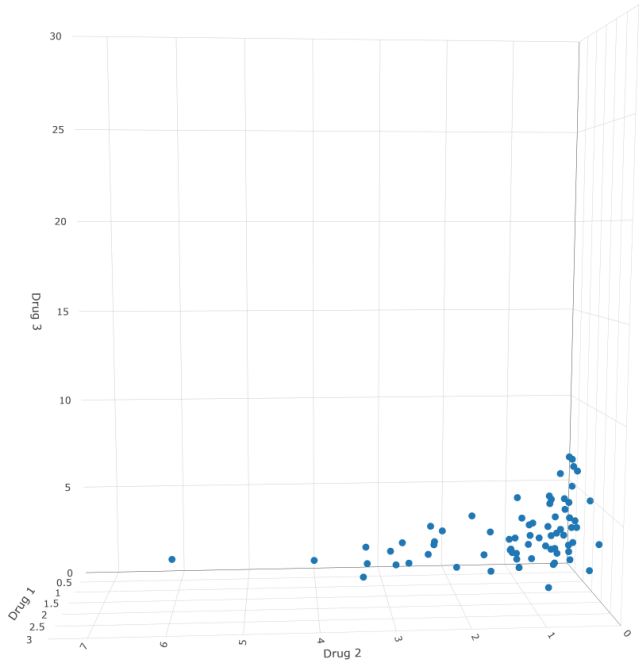


(b) Uniform ray design using factor levels

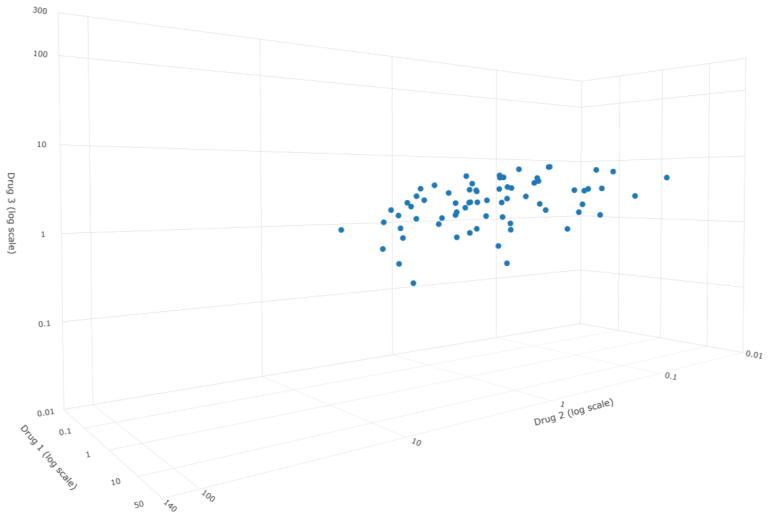


(c) Combined display of ray designs

Figure A27: Ray designs for Scenario 7a). The plots use linearly scaled dose axes. Each axis spans 4 times the EC50 of the respective drug for enhanced comparability. Plot **(a)** uses mixture proportions based on the EC50 values. Plot **(b)** uses mixture proportions based on a uniform table, scaled to the dose range. Plot **(c)** uses mixture proportions based on a uniform table, scaled based on a factor-level table with EC20, EC40, EC50, EC60, EC80. Plot **(d)** combines plots **(a)**-**(c)**

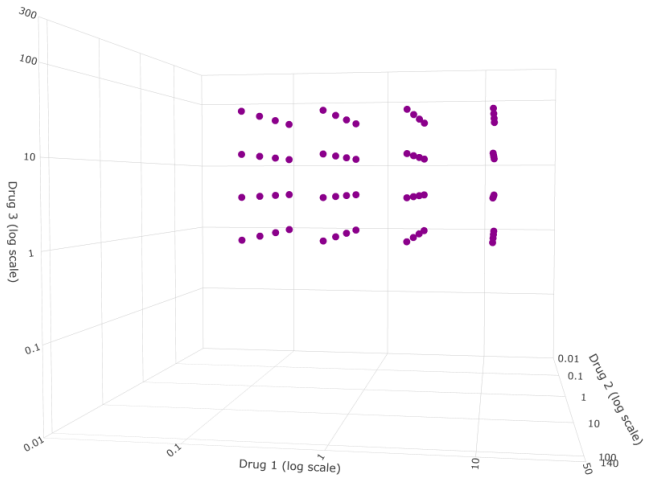


(a) Uniform maximum power design on linear dose scale

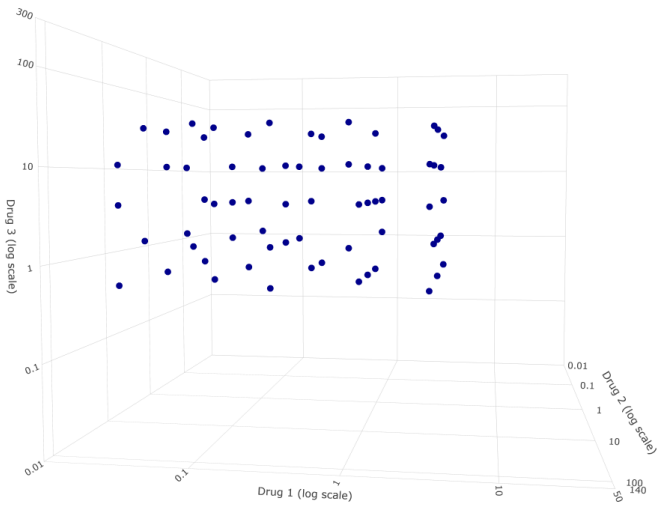


(b) Uniform maximum power design on log dose scale

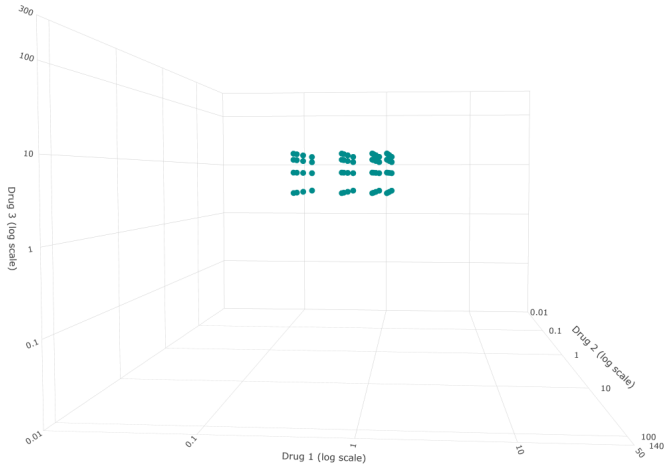
Figure A28: Uniform maximum power design for Scenario 6b). Plot **(a)** uses linearly scaled dose axes, spanning 3 times the EC50 of the respective drug. Plot **(b)** uses log scaled dose axes, spanning the entire dose range.



(a) Checkercube with 4 dose levels around EC50



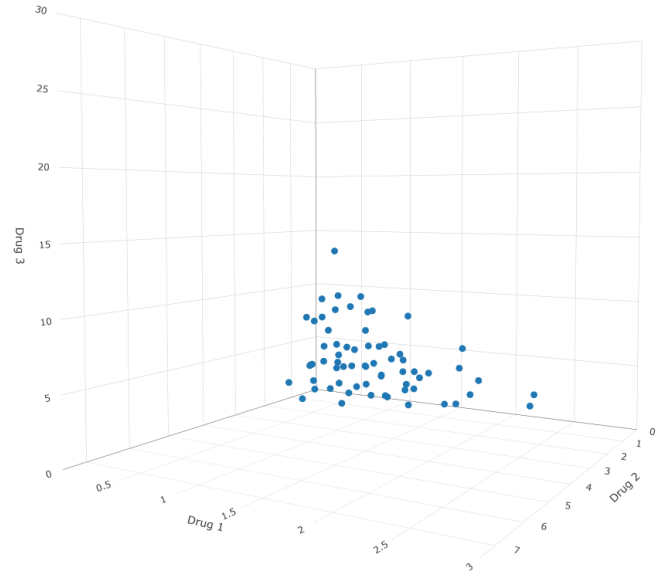
(b) Sampling from checkercube with the 5 lowest dose levels



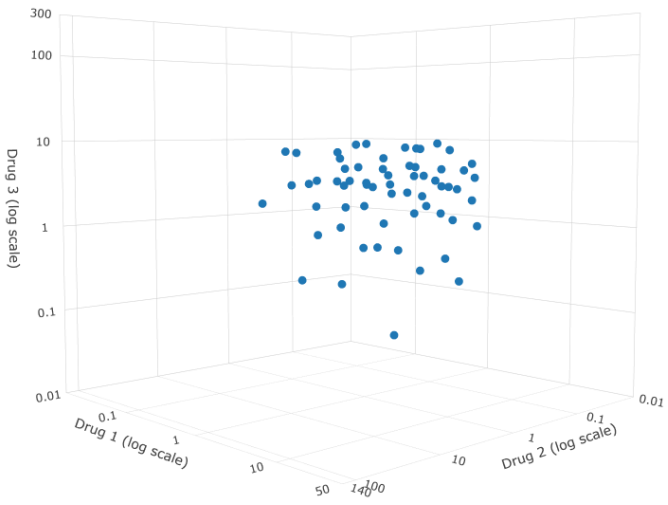
(c) Checkercube with 4 dose levels at EC10, EC30, EC50 and EC60

Figure A29: Checkercube designs for Scenario 7a). Plot **(a)** illustrates a 4x4x4 cube based on the closest 2 monotherapy dilutions above and the closest 2 monotherapy dilutions below the EC50 of each drug. Plot **(b)** samples uniformly from a 5x5x5 cube based on the lowest 5 monotherapy dilutions of each drug. Plot **(c)** is a 4x4x4 cube based on EC10, EC30, EC50, EC60 of each drug. The axes are log scaled and span the entire dose range.

A.1.16 Designs for Scenario 7b)



(a) Uniform maximum power design on linear dose scale



(b) Uniform maximum power design on log dose scale

Figure A30: Uniform maximum power design for Scenario 7b). Plot **(a)** uses linearly scaled dose axes, spanning 4 times the EC50 of drug 2 and drug 3 and 8 times the EC50 of drug 1. Plot **(b)** uses log scaled dose axes, spanning the entire dose range.

A.2 Further Simulation Results

A.2.1 Design Points and Coverage of Synergy Area

Table A1: Further Information about Implemented Designs

	Design 1	Design 2	Design 3	Design 4	Design 5	Design 6	Design 7
Scenario 1a)							
N_{total} (N_{90+})	71 (1)	70 (10)	70 (10)	63 (9)	64 (28)	68 (18)	64 (7)
$n_{\text{syn}<5}$ ($N_{\text{syn}=0}$)	1 (0)	17 (17)	0 (0)	4 (4)	4 (0)	0 (0)	3 (0)
n_{syn}: M (SD)	26.1 (9.2)	23.5 (13.6)	32.7 (12.7)	26.4 (8.1)	16.8 (5.5)	19.4 (4.5)	21.0 (8.5)
p_{syn}: M (SD)	0.37 (0.13)	0.34 (0.19)	0.47 (0.18)	0.42 (0.13)	0.26 (0.09)	0.28 (0.07)	0.33 (0.13)
Scenario 1b)							
N_{total} (N_{90+})	65 (5)	70 (10)	70 (11)	63 (10)	64 (13)	68 (10)	64 (0)
$n_{\text{syn}<5}$ ($N_{\text{syn}=0}$)	0 (0)	17 (17)	0 (0)	4 (4)	4 (0)	0 (0)	3 (0)
n_{syn}: M (SD)	34.9 (9.3)	20.3 (11.8)	30.2 (11.5)	24.4 (7.8)	16.8 (5.5)	19.4 (4.5)	21.0 (8.5)
p_{syn}: M (SD)	0.54 (0.14)	0.29 (0.17)	0.43 (0.16)	0.39 (0.12)	0.26 (0.09)	0.29 (0.07)	0.33 (0.13)
Scenario 2a)							
N_{total} (N_{90+})	71 (43)	70 (10)	70 (10)	63 (9)	64 (35)	68 (14)	64 (7)
$n_{\text{syn}<5}$ ($N_{\text{syn}=0}$)	3 (0)	18 (18)	0 (0)	0 (0)	23 (0)	0 (0)	0 (0)
n_{syn}: M (SD)	18.9 (7.0)	21.7 (12.6)	39.9 (13.0)	28.5 (4.5)	10.9 (3.4)	26.0 (4.8)	20.7 (6.6)
p_{syn}: M (SD)	0.27 (0.10)	0.31 (0.18)	0.57 (0.19)	0.45 (0.07)	0.17 (0.05)	0.38 (0.07)	0.32 (0.10)
Scenario 2b)							
N_{total} (N_{90+})	71 (0)	70 (10)	70 (10)	63 (10)	64 (12)	68 (6)	64 (0)
$n_{\text{syn}<5}$ ($N_{\text{syn}=0}$)	0 (0)	18 (18)	0 (0)	0 (0)	23 (0)	0 (0)	0 (0)
n_{syn}: M (SD)	54.0 (10.1)	18.7 (11.0)	35.7 (11.5)	26.3 (4.5)	10.9 (3.4)	26.0 (4.8)	20.7 (6.6)
p_{syn}: M (SD)	0.76 (0.14)	0.27 (0.16)	0.51 (0.16)	0.42 (0.07)	0.17 (0.05)	0.38 (0.07)	0.32 (0.10)
Scenario 3a)							
N_{total} (N_{90+})	71 (0)	70 (10)	70 (10)	63 (9)	64 (38)	68 (33)	64 (7)
$n_{\text{syn}<5}$ ($N_{\text{syn}=0}$)	1 (0)	18 (18)	0 (0)	18 (18)	5 (0)	1 (0)	21 (10)
n_{syn}: M (SD)	26.3 (9.3)	24.3 (13.9)	21.8 (9.4)	24.9 (13.4)	20.5 (7.6)	23.3 (6.6)	21.2 (13.2)
p_{syn}: M (SD)	0.37 (0.13)	0.35 (0.20)	0.31 (0.13)	0.39 (0.21)	0.32 (0.12)	0.34 (0.10)	0.33 (0.21)
Scenario 3b)							
N_{total} (N_{90+})	45 (4)	70 (10)	70 (12)	63 (10)	64 (18)	68 (16)	64 (0)
$n_{\text{syn}<5}$ ($N_{\text{syn}=0}$)	1 (0)	18 (18)	0 (0)	18 (18)	5 (0)	1 (0)	21 (10)
n_{syn}: M (SD)	22.9 (6.5)	22.4 (12.7)	20.7 (8.7)	22.9 (12.4)	20.5 (7.6)	23.3 (6.6)	21.2 (13.2)

	Design 1	Design 2	Design 3	Design 4	Design 5	Design 6	Design 7
p_{syn}: M (SD)	0.51 (0.14)	0.32 (0.18)	0.29 (0.12)	0.36 (0.20)	0.32 (0.12)	0.34 (0.10)	0.33 (0.21)
Scenario 4a)							
N_{total} (N_{90+})	71 (24)	70 (10)	70 (10)	70 (10)	64 (31)	68 (23)	64 (7)
$n_{syn<5}$ ($N_{syn=0}$)	2 (0)	17 (17)	0 (0)	2 (2)	8 (0)	3 (0)	35 (8)
n_{syn}: M (SD)	42.7 (12.4)	22.8 (13.8)	59.4 (9.5)	21.6 (13.0)	12.8 (4.3)	13.5 (3.8)	12.1 (7.1)
p_{syn}: M (SD)	0.60 (0.17)	0.33 (0.20)	0.85 (0.14)	0.31 (0.19)	0.20 (0.07)	0.20 (0.05)	0.19 (0.11)
Scenario 4b)							
N_{total} (N_{90+})	NA (NA)	70 (10)	70 (10)	70 (10)	64 (13)	68 (10)	64 (0)
$n_{syn<5}$ ($N_{syn=0}$)	NA (NA)	17 (17)	0 (0)	2 (2)	8 (0)	3 (0)	35 (8)
n_{syn}: M (SD)	NA (NA)	20.9 (12.2)	56.6 (9.2)	20.0 (11.9)	12.8 (4.3)	13.5 (3.8)	12.1 (7.1)
p_{syn}: M (SD)	NA (NA)	0.30 (0.17)	0.81 (0.13)	0.29 (0.17)	0.20 (0.07)	0.20 (0.05)	0.19 (0.11)
Scenario 5a)							
N_{total} (N_{90+})	71 (0)	70 (10)	70 (10)	70 (10)	64 (32)	68 (22)	64 (7)
$n_{syn<5}$ ($N_{syn=0}$)	1 (0)	17 (17)	0 (0)	3 (3)	8 (0)	2 (0)	15 (4)
n_{syn}: M (SD)	39.3 (10.7)	23.2 (13.7)	52.0 (10.9)	24.6 (10.5)	15.0 (5.2)	16.5 (4.7)	15.8 (7.4)
p_{syn}: M (SD)	0.55 (0.15)	0.33 (0.19)	0.74 (0.16)	0.35 (0.14)	0.23 (0.08)	0.24 (0.07)	0.25 (0.12)
Scenario 5b)							
N_{total} (N_{90+})	68 (0)	70 (10)	70 (10)	70 (11)	64 (14)	68 (10)	64 (0)
$n_{syn<5}$ ($N_{syn=0}$)	4 (0)	18 (17)	0 (0)	4 (3)	8 (0)	2 (0)	15 (4)
n_{syn}: M (SD)	20.2 (7.9)	20.2 (11.7)	48.9 (10.3)	22.1 (9.4)	15.0 (5.2)	16.5 (4.7)	15.8 (7.4)
p_{syn}: M (SD)	0.30 (0.12)	0.29 (0.17)	0.70 (0.15)	0.32 (0.13)	0.23 (0.08)	0.24 (0.07)	0.25 (0.12)
Scenario 6a)							
N_{total} (N_{90+})	71 (38)	70 (10)	70 (10)	70 (10)	64 (25)	68 (18)	64 (7)
$n_{syn<5}$ ($N_{syn=0}$)	0 (0)	17 (17)	0 (0)	0 (0)	3 (0)	0 (0)	2 (0)
n_{syn}: M (SD)	39.0 (10.3)	22.0 (13.0)	50.6 (10.1)	26.2 (7.4)	14.8 (3.8)	17.0 (3.7)	15.8 (4.8)
p_{syn}: M (SD)	0.55 (0.15)	0.31 (0.19)	0.72 (0.15)	0.37 (0.11)	0.23 (0.06)	0.25 (0.06)	0.25 (0.07)
Scenario 6b)							
N_{total} (N_{90+})	71 (0)	70 (10)	70 (10)	70 (11)	64 (10)	68 (8)	64 (0)
$n_{syn<5}$ ($N_{syn=0}$)	0 (0)	17 (17)	0 (0)	0 (0)	3 (0)	0 (0)	2 (0)
n_{syn}: M (SD)	31.0 (10.1)	19.0 (11.1)	45.8 (9.1)	23.6 (6.8)	14.8 (3.8)	17.0 (3.7)	15.8 (4.8)
p_{syn}: M (SD)	0.44 (0.14)	0.27 (0.16)	0.65 (0.13)	0.34 (0.10)	0.23 (0.06)	0.25 (0.06)	0.25 (0.07)
Scenario 7a)							
N_{total} (N_{90+})	71 (0)	70 (10)	70 (10)	70 (10)	64 (47)	68 (22)	64 (7)
$n_{syn<5}$ ($N_{syn=0}$)	1 (0)	17 (17)	0 (0)	10 (9)	1 (0)	2 (0)	25 (10)

	Design 1	Design 2	Design 3	Design 4	Design 5	Design 6	Design 7
n_{syn}: M (SD)	36.7 (10.7)	23.5 (13.8)	32.4 (12.4)	23.2 (12.9)	24.8 (6.9)	16.1 (5.0)	16.2 (10.1)
p_{syn}: M (SD)	0.52 (0.15)	0.34 (0.20)	0.46 (0.18)	0.33 (0.18)	0.39 (0.11)	0.24 (0.07)	0.25 (0.16)
Scenario 7b)							
N_{total} (N_{90+})	68 (1)	70 (10)	70 (10)	70 (10)	64 (33)	68 (10)	64 (0)
$n_{syn<5}$ ($N_{syn=0}$)	6 (0)	17 (17)	0 (0)	9 (9)	1 (0)	2 (0)	25 (10)
n_{syn}: M (SD)	15.1 (6.3)	21.8 (12.8)	30.9 (11.6)	21.4 (11.8)	24.8 (6.9)	16.1 (5.0)	16.2 (10.1)
p_{syn}: M (SD)	0.22 (0.09)	0.31 (0.18)	0.44 (0.17)	0.31 (0.17)	0.39 (0.11)	0.24 (0.07)	0.25 (0.16)

In Table A1, N_{total} refers to the total number of combination points tested in a particular design. Out of these N_{total} points, the subset N_{90+} is characterized by a null effect of $\geq 90\%$ under Bliss independence (before random noise was added). $n_{syn<5}$ is the number of simulation runs where less than five of the design points lie in the synergistic area. Out of these $n_{syn<5}$ runs, in the subset $n_{syn=0}$ there is no design point in the synergistic area, i.e. the respective design misses the simulated synergistic area completely. n_{syn} reflects the absolute number of design points lying in the synergistic area, whereas p_{syn} reflects the proportion of design points lying in the synergistic area. Means (M) and standard deviations (SD) across all runs are reported.

A.2.2 Sensitivity, Specificity and Family-Wise Error Rate

Table A2: Mean and Standard Deviation of Sensitivties

Sensitivity: Mean (SD)	Design 1	Design 2	Design 3	Design 4	Design 5	Design 6	Design 7
Scenario 1a)	0.89 (0.08)	0.50 (0.14)	0.50 (0.13)	0.51 (0.14)	0.58 (0.09)	0.66 (0.08)	0.96 (0.05)
Scenario 1b)	0.88 (0.06)	0.54 (0.15)	0.54 (0.14)	0.54 (0.14)	0.73 (0.08)	0.70 (0.09)	0.94 (0.07)
Scenario 2a)	0.79 (0.11)	0.55 (0.14)	0.52 (0.13)	0.51 (0.12)	0.64 (0.10)	0.73 (0.07)	0.96 (0.05)
Scenario 2b)	1.00 (0.01)	0.49 (0.17)	0.52 (0.15)	0.53 (0.14)	0.83 (0.08)	0.77 (0.09)	0.93 (0.07)
Scenario 3a)	0.58 (0.17)	0.44 (0.11)	0.42 (0.13)	0.42 (0.12)	0.37 (0.08)	0.39 (0.08)	0.96 (0.05)
Scenario 3b)	0.68 (0.14)	0.48 (0.12)	0.45 (0.11)	0.44 (0.13)	0.57 (0.08)	0.55 (0.08)	0.93 (0.07)
Scenario 4a)	0.78 (0.09)	0.64 (0.15)	0.77 (0.08)	0.70 (0.14)	0.63 (0.11)	0.71 (0.08)	0.96 (0.06)
Scenario 4b)	NA (NA)	0.42 (0.14)	0.57 (0.10)	0.48 (0.14)	0.74 (0.10)	0.77 (0.08)	0.99 (0.03)
Scenario 5a)	0.96 (0.04)	0.54 (0.14)	0.58 (0.11)	0.55 (0.14)	0.58 (0.12)	0.69 (0.09)	0.97 (0.04)
Scenario 5b)	0.89 (0.12)	0.59 (0.16)	0.63 (0.12)	0.61 (0.15)	0.72 (0.09)	0.74 (0.08)	0.98 (0.04)
Scenario 6a)	0.79 (0.08)	0.52 (0.15)	0.55 (0.12)	0.53 (0.13)	0.71 (0.09)	0.76 (0.08)	0.98 (0.04)
Scenario 6b)	0.99 (0.03)	0.52 (0.18)	0.59 (0.14)	0.57 (0.16)	0.82 (0.08)	0.78 (0.11)	0.96 (0.06)
Scenario 7a)	0.89 (0.09)	0.65 (0.13)	0.67 (0.12)	0.66 (0.14)	0.38 (0.06)	0.71 (0.09)	0.97 (0.05)
Scenario 7b)	0.58 (0.25)	0.40 (0.12)	0.42 (0.11)	0.41 (0.12)	0.52 (0.08)	0.69 (0.11)	0.97 (0.05)

Table A3: Mean and Standard Deviation of Specificities

Specificity: Mean (SD)	Design 1	Design 2	Design 3	Design 4	Design 5	Design 6	Design 7
Scenario 1a)	0.998 (0.01)	0.997 (0.02)	0.998 (0.01)	0.998 (0.01)	0.998 (0.01)	0.999 (0.01)	0.997 (0.01)
Scenario 1b)	0.998 (0.01)	0.997 (0.01)	0.997 (0.02)	0.998 (0.01)	0.998 (0.01)	0.998 (0.01)	0.998 (0.01)
Scenario 2a)	0.999 (0.01)	0,999 (0.00)	0.999 (0.01)	0.998 (0.01)	0.999 (0.01)	0.999 (0.01)	0.999 (0.01)
Scenario 2b)	0.997 (0.01)	0.998 (0.01)	0.999 (0.01)	0.997 (0.01)	0.999 (0.01)	0.998 (0.01)	0.999 (0.00)
Scenario 3a)	0.997 (0.01)	0.997 (0.01)	0.998 (0.01)	0.997 (0.01)	0.999 (0.01)	0.999 (0.01)	0.998 (0.01)
Scenario 3b)	0.999 (0.01)	0.998 (0.01)	0.999 (0.00)	0.999 (0.00)	0.999 (0.01)	0.999 (0.01)	0.999 (0.01)
Scenario 4a)	0.995 (0.02)	0.999 (0.00)	1.000 (0.00)	0.998 (0.01)	0.998 (0.01)	0.998 (0.01)	0.998 (0.01)
Scenario 4b)	NA (NA)	0.999 (0.01)	1.000 (0.00)	0.999 (0.00)	0.999 (0.00)	0.999 (0.00)	0.999 (0.01)
Scenario 5a)	0.999 (0.01)	0.998 (0.01)	0.998 (0.02)	0.998 (0.01)	0.998 (0.01)	0.998 (0.01)	0.999 (0.01)
Scenario 5b)	0.997 (0.02)	0.998 (0.02)	0.998 (0.01)	0.997 (0.02)	0.997 (0.01)	0.998 (0.01)	0.998 (0.01)
Scenario 6a)	0.996 (0.01)	0.997 (0.01)	0.998 (0.01)	0.997 (0.01)	0.997 (0.01)	0.998 (0.01)	0.998 (0.01)
Scenario 6b)	0.998 (0.01)	0.998 (0.01)	0.999 (0.01)	0.997 (0.01)	0.998 (0.01)	0.999 (0.01)	0.998 (0.01)
Scenario 7a)	0.998 (0.01)	0.999 (0.01)	0.998 (0.01)	0.998 (0.01)	0.998 (0.01)	0.998 (0.01)	0.998 (0.01)
Scenario 7b)	0.998 (0.01)	0.999 (0.01)	0.999 (0.01)	0.999 (0.01)	0.998 (0.01)	0.999 (0.01)	0.999 (0.01)

Table A4: Family-Wise Error Rates

FWER	Design 1	Design 2	Design 3	Design 4	Design 5	Design 6	Design 7
Scenario 1a)	0.11	0.12	0.09	0.08	0.12	0.14	0.09
Scenario 1b)	0.08	0.11	0.09	0.06	0.15	0.09	0.09
Scenario 2a)	0.16	0.08	0.06	0.10	0.11	0.06	0.06
Scenario 2b)	0.07	0.06	0.05	0.07	0.09	0.08	0.08
Scenario 3a)	0.10	0.11	0.10	0.08	0.08	0.08	0.08
Scenario 3b)	0.04	0.08	0.06	0.05	0.07	0.07	0.09
Scenario 4a)	0.12	0.07	0.01	0.12	0.12	0.14	0.11
Scenario 4b)	NA	0.05	0.01	0.07	0.09	0.06	0.10
Scenario 5a)	0.07	0.07	0.05	0.07	0.14	0.10	0.08
Scenario 5b)	0.11	0.10	0.06	0.11	0.13	0.16	0.09
Scenario 6a)	0.13	0.14	0.05	0.10	0.15	0.15	0.10
Scenario 6b)	0.09	0.13	0.07	0.11	0.11	0.11	0.10
Scenario 7a)	0.08	0.10	0.08	0.11	0.28	0.13	0.11
Scenario 7b)	0.08	0.09	0.06	0.07	0.10	0.11	0.07

A.2.3 Power for Detecting at Least 10 Synergistic Points

Figure A31: Heatmap displaying the power for detecting at least 10 synergistic points by experimental design (x-axis) and monotherapy scenario (y-axis). Power is calculated accross all simulation runs, irrespectively of whether at least 10 simulated synergistic points were covered by a particular design in a particular run. Nonetheless, synergy areas were generally large enough to give all designs the chance to detect at least 10 synergistic points, if they sampled in a relevant range: There were only 1-3 isolated runs in scenarios 1,2,3 and 7 where no design managed to cover at least 10 synergistic points.

A.2.4 Software Code and Raw Results

R code used in the simulations as well as files containing the raw simulation results for all scenarios and runs are provided via GitHub in the repository https://github.com/marielse/triplecombi.