Optimal designs for active controlled dose finding trials with efficacy-toxicity outcomes

Holger Dette, Katrin Kettelhake, Kirsten Schorning
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany
e-mail: holger.dette@ruhr-uni-bochum.de

Weng Kee Wong
Department of Biostatistics
UCLA School of Public Health
Los Angeles, CA 90095-1772
e-mail: wkwong@ucla.edu

Frank Bretz
Statistical Methodology
Novartis Pharma AG
4002 Basel, Switzerland
e-mail: frank.bretz@novartis.com

January 5, 2016

Abstract

Nonlinear regression models addressing both efficacy and toxicity outcomes are increasingly used in dose-finding trials, such as in pharmaceutical drug development. However, research on related experimental design problems for corresponding active controlled trials is still scarce. In this paper we derive optimal designs to estimate efficacy and toxicity in an active controlled clinical dose finding trial when the bivariate continuous outcomes are modeled either by polynomials up to degree 2, the Michaelis-Menten model, the Emax model, or a combination thereof. We determine upper bounds on the number of different doses levels required for the optimal design and provide conditions under which the boundary points of the design space are included in the optimal design. We also provide an analytical description of the minimally supported D-optimal designs and show that they do not depend on the correlation between the bivariate outcomes. We illustrate the proposed methods with numerical examples and demonstrate the advantages of the D-optimal design for a trial, which has recently been considered in the literature.

Keywords and Phrases: Active controlled trials, dose finding, optimal design, admissible design, Emax model, Equivalence theorem, Particle swarm optimization, Tchebycheff system.

1. Introduction

There is a vast literature on optimal design of experiments, with applications ranging across many disciplines. Atkinson (1996) showed the usefulness of optimal designs for real applications using examples from agriculture, animal breeding studies, accelerated life-testing experiments and computer experiments. Berger and Wong (2005) gave examples of the varied disciplines that increasingly use optimal design ideas for scientific investigations. Most of the literature focuses on optimal designs for models with a univariate outcome. In practice, however, drug trials are often conducted to measure multiple outcomes that are likely to be correlated. For instance, pharmaceutical dose-finding trials invariably have bivariate outcomes involving efficacy and toxicity. As a motivating example consider a randomized controlled clinical trials for hypertensive patients treated with an angiotensin-converting-enzyme (ACE) inhibitor. In such trials change from baseline in the sitting blood pressure is a frequent efficacy outcome. However, patients starting on an ACE inhibitor usually have a modest reduction in glomerular filtration rate (GFR) that stabilizes after several days. Because this decrease may be significant in conditions of decreased renal perfusion, the renal function should be closely monitored over the first few days in those patients [see Sidorenkov and Navis (2014) and Tao et al. (2015)]. Thus, the amount of decrease in GFR from baseline is a common outcome to assess unwanted side effects.

Several papers have addressed design problems that incorporate both efficacy and toxicity of a drug. Fan and Chaloner (2004) proposed using a continuous ratio model for a trinomial outcome, where the outcome of a patient may be classified as "no reaction" when neither toxicity nor efficacy occurs, "efficacy" for efficacy without toxicity, and "adverse reaction" for toxicity. Heise and Myers (1996) used the Gumbel bivariate binary quantal response model to study efficacy and toxicity. In their example, patients were continuously monitored whether they experienced a toxicity event or a treatment benefit. The authors investigated locally D- and Q-optimal designs, where D-optimal designs were constructed for estimating all model parameters and Q-optimal designs were obtained by maximizing the probability of efficacy without toxicity at a selected dose level. More recently, Magnusdottir (2013) applied c-optimal designs to the bivariate Emax model for continuous efficacy and toxicity outcomes. The author determined the dose level providing the best possible combination of efficacy and toxicity, based on a pre-specified clinical utility index [see Carrothers (2011)].

Adaptive dose finding trials incorporating both efficacy and safety have been investigated as well. For example, Thall and Cook (2004) and Dragalin and Fedorov (2006) found adaptive designs for dose-finding based on efficacy-toxicity outcome using a Gumbel bivariate logistic regression or a Cox bivariate model. Dragalin et al. (2008) proposed new designs for selecting drug combinations for a bivariate Probit correlated model based on an efficacy-toxicity outcome profile of a drug using Bayesian, minimax and adaptive methods. More recently, Tao et al. (2013) considered a joint model with mixed correlated toxicity and ef-

ficacy outcomes, with one discrete and the other continuous. Using outcomes constructed with Archimedean Copula, they extended the continual reassessment method to find the optimal dose for Phase III trials based on both efficacy and toxicity considerations. Some advantages and disadvantages of adaptive designs have been discussed in Dette et al. (2013). Recently, the use of active controls instead of placebo in clinical trials has received considerable attention in the literature [see for example Temple and Ellenberg (2000), Splawinski and Kuzniar (2004), Helms et al. (2015a,b) among many others], and design issues for univariate outcomes have been discussed in Dette et al. (2014, 2015). To our best knowledge, the problem of determining optimal designs for active controlled trials with bivariate mean outcomes has not been considered in the literature so far. In the present paper we derive optimal designs when the bivariate outcomes are efficacy and toxicity measures and are modeled using combinations of the linear, quadratic, Michaelis-Menten and Emax model as possible mean functions. We note that the Emax model is especially flexible in its shape and commonly applied in dose-finding trials. In particular, the Emax model can be justified through the relationship of drug-receptor interactions and therefore deduced from the chemical equilibrium equation [see e.g. Boroujerdi (2002)].

In Section 2 we introduce the model and provide some technical background. In Section 3 we consider various models for efficacy and toxicity outcomes without an active control and provide upper bounds on the required number of doses for each combination of the possible mean functions in the bivariate model. We also state sufficient conditions under which the boundary points of the design space are included as support points of the optimal design and determine the minimally supported optimal designs analytically. In Section 4 we apply these results to design active-controlled dose finding trials. In Section 5 we report the optimal designs for a real trial with bivariate outcome using particle swarm optimization. We conclude with a discussion in Section 6 and provide the technical proofs for our main results in Section 7.

2. Optimal designs for bivariate outcome

We consider a dose-finding trial investigating both the efficacy and toxicity of a new drug under investigation. Our goal is to find an optimal design for collecting data for the two outcomes at different dose levels. Given a statistical model defined on a given dose interval of interest, say $\mathcal{D} = [L, R] \subset \mathbb{R}_0^+$, and a given design criterion, the design problem is to determine the optimal number of doses, k, the dose levels $d_1, \ldots, d_k$ from the dose interval $\mathcal{D}$, and the number of patients assigned to each dose. In practice, the total sample size, say n_1, is determined either by standard power considerations or by requirements on the precision of estimating the dose-response curves. That is, for a given value of n_1, the optimal design needs to specify the number of patients n_{1i} at each dose level d_i subject to $\sum_{i=1}^k n_{1i} = n_1$. Note that we use the index "1" here in the notation (i.e. $n_1, n_{1i}, \ldots$) since in later sections

we consider active controlled clinical trials with a second sample denoted by the index "2". Let Y_{ij} be the two-dimensional outcome variable at dose level d_i from subject j and assume that

$$Y_{ij} = (Y_{ij}^e, Y_{ij}^t)^T \sim \mathcal{N}_2(\eta_1(d_i, \theta_1), \Sigma_1), \qquad j = 1, \ldots, n_{1i}, i = 1, \ldots, k. \tag{2.1}$$

The regression function

$$\eta_1(d, \theta_1) = (\eta_1^e(d, \theta_1^e), \eta_1^t(d, \theta_1^t))^T \in \mathbb{R}^2$$

describes the expected efficacy (η_1^e) and toxicity (η_1^t) at dose level $d \in \mathcal{D}$, where the $(s_1^e + 1)$- and $(s_1^t + 1)$-dimensional vectors θ_1^e and θ_1^t define the parameters of the model η_1^e and η_1^t, respectively. The parameter $\theta_1 = ((\theta_1^e)^T, (\theta_1^t)^T)^T$ varies in a compact parameter space, say $\Theta_1 \subset \mathbb{R}^{s_1}$, where $s_1 = s_1^e + s_1^t + 2$. The unknown covariance matrix is

$$\Sigma_1 = \mathrm{Cov}(Y) = \begin{pmatrix} \sigma_e^2 & \rho\sigma_e\sigma_t \\ \rho\sigma_e\sigma_t & \sigma_t^2 \end{pmatrix},$$

where $-1 < \rho < 1$ denotes the correlation between the two outcome variables and the variances of the random variables Y_{ij}^e and Y_{ij}^t are given by σ_e^2 and σ_t^2, respectively. We assume that η_1 is continuously differentiable with respect to the parameter θ_1 and denote by

$$f_e(d) \;=\; \frac{\partial}{\partial\theta_1^e}\eta_1^e(d, \theta_1) = (f_0^e(d), \ldots, f_{s_1^e}^e(d))^T \;,\; f_t(d) = \frac{\partial}{\partial\theta_1^t}\eta_1^t(d, \theta_1) = (f_0^t(d), \ldots, f_{s_1^t}^t(d))^T$$

the gradients of the two mean responses with respect to θ_1^e and θ_1^t, respectively. Following Fahrmeir and Tutz (2001), the Fisher information matrix is given by the $s_1 \times s_1$-matrix

$$\begin{aligned} \mathcal{I}_1(d, \theta_1) &= (\tfrac{\partial}{\partial\theta_1}\eta_1(d, \theta_1))^T\Sigma_1^{-1}(\tfrac{\partial}{\partial\theta_1}\eta_1(d, \theta_1)) = \begin{pmatrix} f_e(d) & \mathbf{0}_{s_1^e+1} \\ \mathbf{0}_{s_1^t+1} & f_t(d) \end{pmatrix} \Sigma_1^{-1} \begin{pmatrix} f_e^T(d) & \mathbf{0}_{s_1^t+1}^T \\ \mathbf{0}_{s_1^e+1}^T & f_t^T(d) \end{pmatrix} \\ &= \frac{1}{\sigma_e^2\sigma_t^2(1-\rho^2)} F(d). \end{aligned}$$

Here, $\mathbf{0}_d$ is the d-dimensional vector with all entries equal to 0 and

$$F(d) \;=\; \frac{1}{\sigma_e^2\sigma_t^2(1-\rho^2)} \begin{pmatrix} \sigma_t^2\mathcal{F}_1 & -\rho\sigma_e\sigma_t\mathcal{F}_2 \\ -\rho\sigma_e\sigma_t\mathcal{F}_2^T & \sigma_e^2\mathcal{F}_3 \end{pmatrix} \tag{2.2}$$

is defined through

$$\begin{aligned} \mathcal{F}_1 &= f_e(d)f_e^T(d) \in \mathbb{R}^{s_1^e+1 \times s_1^e+1}, \quad \mathcal{F}_3 = f_t(d)f_t^T(d) \in \mathbb{R}^{s_1^t+1 \times s_1^t+1}, \\ \mathcal{F}_2 &= f_e(d)f_t^T(d) \in \mathbb{R}^{s_1^e+1 \times s_1^t+1}. \end{aligned}$$

Note that we have suppressed the dependency of the matrices F, $\mathcal{F}_1$, $\mathcal{F}_2$ and $\mathcal{F}_3$ on the parameter θ_1 in our notation.

Throughout this paper we consider approximate designs in the sense of Kiefer (1974), which are defined as probability measures with finite support on the design space $\mathcal{D}$. If an approximate design ξ has k support points, say $d_1, \ldots, d_k$, with corresponding positive weights $\omega_1, \ldots, \omega_k$, such that $\sum_{i=1}^k \omega_i = 1$, and n_1 observations can be taken, a rounding procedure is applied to obtain integers n_{1i}, $i = 1, \ldots, k$, from the possibly rational numbers $\omega_i n_1$ [see Pukelsheim and Rieder (1992)]. The information matrix $M_1(\xi, \theta_1)$ of a design ξ is defined by the $s_1 \times s_1$ matrix

$$M_1(\xi, \theta_1) = \int_{\mathcal{D}} \mathcal{I}_1(d, \theta_1) d\xi(d) = \sum_{i=1}^{k} \frac{\omega_i}{\sigma_e^2 \sigma_t^2 (1-\rho^2)} F(d_i), \tag{2.3}$$

where the matrix $F(d)$ is defined in (2.2).

If observations are taken according to an approximate design ξ it can be shown that, under standard regularity conditions, the maximum likelihood estimator $\hat{\theta}_1$ is asymptotically normally distributed, that is

$$\sqrt{n_1}(\hat{\theta}_1 - \theta_1) \xrightarrow{\mathcal{D}} \mathcal{N}_{s_1}(\mathbf{0}, M_1^{-1}(\xi, \theta_1)),$$

as $n_1 \to \infty$, where the symbol $\xrightarrow{\mathcal{D}}$ means convergence in distribution. Consequently designs that make the information matrix $M_1(\xi, \theta_1)$ large in some sense are appropriate. There are several design criteria used in practice. An important example is Kiefer's ϕ_p-criterion [see Kiefer (1974)]. To be precise, let $p \in [-\infty, 1)$ and let $K \in \mathbb{R}^{s_1 \times m}$ be a given matrix of full column rank. A design ξ^* is called locally ϕ_p-optimal for estimating the linear combination $K^T\theta_1$ if it maximizes the concave functional

$$\phi_p(\xi) = \Big(\frac{1}{m} \mathrm{tr}(K^T M_1^-(\xi, \theta_1) K)^{-p} \Big)^{\frac{1}{p}}$$

among all designs ξ satisfying $\mathrm{Range}(K) \subset \mathrm{Range}(M_1(\xi, \theta_1))$, i.e. $K^T\theta_1$ is estimable by the design ξ. Here, $\mathrm{tr}(A)$ and A^- denote the trace and a generalized inverse of the matrix A, respectively.

One key advantage of working with approximate designs is that convex optimization theory can be applied if the design criterion is a concave functional. As a consequence, a general equivalence theorem is available to verify whether a design is optimal among all designs. Any concave functional has its own equivalence theorem but collectively they all have a similar form. For each member of Kiefer's ϕ_p-criterion, a direct application of Theorem 7.14 in Pukelsheim (2006) yields the following result.

Theorem 2.1 *Let K be a $s_1 \times m$ matrix of full column rank. If $p \in (-\infty, 1)$, a design ξ^* with*

Range$(K) \subset$ *Range*$(M_1(\xi^*, \theta_1))$ *is locally* ϕ_p*-optimal for estimating the linear combination* $K^T\theta_1$ *if and only if there exists a generalized inverse* G *of the information matrix* $M_1(\xi^*, \theta_1)$*, such that*

$$\operatorname{tr}\big(\mathcal{I}_1(d, \theta_1) G K (C_K(\xi^*))^{p+1} K^T G^T\big) \; - \; \operatorname{tr}(C_K(\xi^*))^p \leq 0 \tag{2.4}$$

holds for all $d \in \mathcal{D}$*, where* $C_K(\xi^*) = (K^T M_1^-(\xi^*, \theta_1) K)^{-1}$*. If* $p = -\infty$*, a design* ξ^* *with* *Range*$(K) \subset$ *Range*$(M_1(\xi^*, \theta_1))$ *is locally* $\phi_{-\infty}$*-optimal for estimating the linear combination* $K^T\theta_1$ *if and only if there exists a generalized inverse* G *of the information matrix* $M_1(\xi^*, \theta_1)$ *and a non-negative definite matrix* $E \in \mathbb{R}^{m \times m}$ *with* $\operatorname{tr}(E) = 1$*, such that*

$$\operatorname{tr}\big(\mathcal{I}_1(d, \theta_1) G K C_K(\xi^*) E C_K(\xi^*) K^T G^T\big) - \lambda_{\min}(C_K(\xi^*)) \leq 0. \tag{2.5}$$

holds for all $d \in \mathcal{D}$*. Moreover, if the design* ξ^* *is* ϕ_p*-optimal, there is equality in the above inequalities.*

The function on the left hand side of (2.4) or (2.5) is a function of the dose d and is sometimes called the sensitivity function of the design ξ^*. In practice, one plots the sensitivity function over the entire dose range and checks whether it is bounded above by zero. If it does, the design ξ^* is optimal; otherwise it is not. In addition, the sensitivity function, along with the equivalence theorem, can be used to provide a lower bound on the efficiency of any design. For example, if $p > -\infty$, one can show that the ϕ_p-efficiency of a design ξ can be bounded from below, that is

$$\operatorname{eff}_p(\xi) = \frac{\phi_p(\xi)}{\sup_\eta \phi_p(\eta)} \geq \frac{\operatorname{tr}(C_K(\xi))^p}{\max_{d \in \mathcal{D}} \operatorname{tr}(\mathcal{I}_1(d, \theta_1) G K (C_K(\xi))^{p+1} K^T G^T)} \tag{2.6}$$

[see Dette (1996)]. Moreover, characterizations of the type (2.4) or (2.5) are also useful to find optimal designs analytically if the model is not too complicated. However, regression models with multivariate outcome are complex and in practice optimal designs have to be found numerically [see Chang (1997), Atashgah and Seifi (2007) or Sagnol (2011) among others].

For such calculations, sharp bounds on the number of support points of the optimal designs are useful, because they can substantially reduce the complexity of the optimization problem. In order to derive such upper bounds we follow Karlin and Studden (1966) and call a design ξ_1 admissible if there does not exist a design ξ_2, such that $M_1(\xi_1, \theta_1) \neq M_1(\xi_2, \theta_1)$ and

$$M_1(\xi_1, \theta_1) \leq M_1(\xi_2, \theta_1)$$

with respect to the Loewner ordering. In other words, the information matrix of an admissible design cannot be improved and numerical optimization can be restricted to the class of admissible designs. The characterization of the number of support points of admissible designs has found considerable attention in the recent literature [see Yang and Stufken (2009,

2012), Yang (2010), Dette and Melas (2011) or Dette and Schorning (2013)]. These authors obtained substantially smaller bounds on the number of support points of optimal designs than provided by the classical approach using Caratheodory's theorem [see Pukelsheim (2006) for example].

In the following we demonstrate that the results in the above references can in fact be proved under weaker assumptions than usually made in the literature. For this purpose we will make use of the theory of Tchebycheff systems [see Karlin and Studden (1966)]. A set of $k+1$ continuous functions $u_0, \ldots, u_k \colon [L, R] \to \mathbb{R}$ is called a Tchebycheff system on the interval $[L, R]$ if the inequality $\det(u_i(d_j))_{i,j=0}^{k} > 0$ holds for all $L \leq d_0 < d_1 < \ldots < d_k \leq R$. We define the index $I(\xi)$ of a design ξ on the interval $[L, R]$ as the number of support points, where interior support points are counted by one and the support points at the boundary of the interval $[L, R]$ are counted by one half.

Note that we can rewrite the information matrix $M_1(\xi, \theta_1)$ in the form

$$M_1(\xi, \theta_1) = \begin{pmatrix} \int_L^R \psi_{11}(d) d\xi(d) & \ldots & \int_L^R \psi_{1 s_1}(d) d\xi(d) \\ \vdots & & \vdots \\ \int_L^R \psi_{s_1 1}(d) d\xi(d) & \ldots & \int_L^R \psi_{s_1 s_1}(d) d\xi(d) \end{pmatrix}, \tag{2.7}$$

where we ignore the dependence of the functions ψ_{ij} on the parameter θ_1. We now define $\psi_0(d) \equiv 1$ and choose a basis, say $\{\psi_0, \ldots, \psi_k\}$, for the space $\operatorname{span}(\{\psi_{ij} | 1 \leq i, j \leq s_1\} \cup \{1\})$. We further assume that ψ_k is one of the diagonal elements of the matrix $M_1(\xi, \theta_1)$, does not coincide with any of the other elements ψ_{ij} and that $\{\psi_0, \ldots, \psi_{k-1}\}$ is a basis of the space

$$\operatorname{span}\big(\{\psi_{ij} \mid i, j \in \{1, \ldots, s_1\}; \quad \psi_{ij} \neq \psi_k\} \cup \{1\}\big).$$

Our next result, Theorem 2.2, is a more general version of Theorem 3.1 in Dette and Melas (2011) that is specific to our problem here. The proof is quite similar to the one given in this reference and is omitted for the sake of brevity. Theorem 2.2 yields better bounds on the number of support points of optimal designs obtained from the current literature; an example is given at the end of Section 7.1.

Theorem 2.2

(A) If $\{\psi_0, \psi_1, \ldots, \psi_{k-1}\}$ and $\{\psi_0, \psi_1, \ldots, \psi_k\}$ are Tchebycheff systems on the interval $\mathcal{D}$, then for any design ξ there exists a design ξ^+ with at most $\frac{k+2}{2}$ support points, such that $M_1(\xi^+, \theta_1) \geq M_1(\xi, \theta_1)$. If the index of the design ξ satisfies $I(\xi) < \frac{k}{2}$, then the design ξ^+ is uniquely determined in the class of all designs η satisfying

$$\int_L^R \psi_i(d) d\eta(d) = \int_L^R \psi_i(d) d\xi(d), \quad i = 0, \ldots, k-1 \tag{2.8}$$

and coincides with the design ξ. Otherwise, in the case $I(\xi) \geq \frac{k}{2}$, the following two assertions are valid.

(A1) If k is odd, then ξ^+ has at most $\frac{k+1}{2}$ support points and ξ^+ can be chosen such that its support contains the point R.

(A2) If k is even, then ξ^+ has at most $\frac{k}{2} + 1$ support points and ξ^+ can be chosen such that its support contains the points L and R.

(B) If $\{\psi_0, \psi_1, \ldots, \psi_{k-1}\}$ and $\{\psi_0, \psi_1, \ldots, -\psi_k\}$ are Tchebycheff systems, then for any design ξ there exists a design ξ^- with at most $\frac{k+2}{2}$ support points, such that $M_1(\xi^-, \theta_1) \geq M_1(\xi, \theta_1)$. If the index of the design ξ satisfies $I(\xi) < \frac{k}{2}$ then the design ξ^- is uniquely determined in the class of all designs η satisfying (2.8) and coincides with the design ξ. Otherwise, in the case $I(\xi) \geq \frac{k}{2}$, the following two assertions are valid.

(B1) If k is odd, then ξ^- has at most $\frac{k+1}{2}$ support points and ξ^- can be chosen such that its support contains the point L.

(B2) If k is even, then ξ^- has at most $\frac{k}{2} + 1$ support points.

We note that Theorem 2.2 provides information about the admissible designs. For example, consider the case $(A2)$ with $k = 2m$ for some $m \in \mathbb{N}$. Any design ξ with index $I(\xi) \geq m$ can be improved with respect to the Loewner ordering by a design with at most $m + 1$ support points that includes the boundary points L and R. It follows that admissible designs are designs with index $< m$ and designs with $m + 1$ support points that include the boundary points L and R of the design space.

3. Optimal designs for placebo-controlled dose finding trials

In this section we study optimal designs for several nonlinear regression models which are commonly used in placebo-controlled dose-finding trials with joint efficacy-toxicity outcomes. In particular we use Theorem 2.2 to derive bounds on the number of support points of optimal

designs and explicit expressions for minimally supported designs. The proofs of the results presented here can be found in the Appendix.

3.1 Bounds on the number of support points

In order to determine bounds for the number of support points of optimal designs we note that the mapping $M \to (K^T M^- K)^{-1}$ is increasing with respect to the Loewner ordering on the set of all $s_1 \times s_1$-matrices satisfying $\text{Range}(K) \subset \text{Range}(M)$ [see Pukelsheim (2006)]. That is, if

$$M_1 \geq M_2 \quad \Rightarrow \quad (K^T M_1^- K)^{-1} \geq (K^T M_2^- K)^{-1},$$

for all matrices M_1, M_2 satisfying the range inclusion. It therefore follows that the information matrix $(K^T M^-(\xi, \theta_1) K)^{-1}$ of a non-admissible design can be improved with respect to the Loewner ordering. Because the ϕ_p-criteria are monotone, we have $\phi_p(\xi) \leq \phi_p(\xi^*)$ for any design ξ, where ξ^* is either ξ^+ or ξ^- as given in Theorem 2.2. This conclusion is true, whenever the assumptions of Theorem 2.2 are satisfied. The following results show that this is in fact the case for many of the commonly used dose response models with a bivariate outcome and give upper bounds on the number of support points of such designs.

Theorem 3.1 *Assume that the model for efficacy is given by $\eta_1^e(d, \theta_1) = \vartheta_0^e + \vartheta_1^e d$ and that ξ is an arbitrary design on the dose range $\mathcal{D} = [L, R]$.*

(a) If $\eta_1^t(d, \theta_1) = \vartheta_0^t + \vartheta_1^t d$, there exists a design ξ^* with at most two support points, such that $M_1(\xi^*, \theta_1) \geq M_1(\xi, \theta_1)$. If the index of the design ξ satisfies $I(\xi) \geq 1$, ξ^* can be chosen such that the support of ξ^* contains the points L and R.

(b) If $\eta_1^t(d, \theta_1) = \vartheta_0^t + \vartheta_1^t d + \vartheta_2^t d^2$, there exists a design ξ^* with at most three support points, such that $M_1(\xi^*, \theta_1) \geq M_1(\xi, \theta_1)$. If the index of the design ξ satisfies $I(\xi) \geq 2$, ξ^* can be chosen such that the support of ξ^* contains the points L and R.

(c) If $\eta_1^t(d, \theta_1)$ is given by a Michaelis-Menten model, that is $\eta_1^t(d, \theta_1) = \frac{\vartheta_1^t d}{\vartheta_2^t + d}$, there exists a design ξ^* with at most four support points, such that $M_1(\xi^*, \theta_1) \geq M_1(\xi, \theta_1)$. If the index of the design ξ satisfies $I(\xi) \geq 3$, ξ^* can be chosen such that the support of ξ^* contains the points L and R.

(d) If $\eta_1^t(d, \theta_1)$ is given by an Emax-model, that is $\eta_1^t(d, \theta_1) = \vartheta_0^t + \frac{\vartheta_1^t d}{\vartheta_2^t + d}$, there exists a design ξ^* with at most four support points, such that $M_1(\xi^*, \theta_1) \geq M_1(\xi, \theta_1)$. If the index of the design ξ satisfies $I(\xi) \geq 3$, ξ^* can be chosen such that the support of ξ^* contains the points L and R.

Remark 3.1 Note that the bounds provided by Theorem 3.1 are not necessarily sharp. For example, if η_1^t is the Emax and η_1^e is the linear model, then by the first part of Theorem 3.1(d)

one does not decrease the information (with respect to the Loewner ordering) by considering only designs with at most four support points. Any design with four support points or three support points in the interior of the dose range has index ≥ 3 and can therefore be further improved by a design with at most four support points including the boundary points L and R. As one requires at least three different dose levels to estimate the parameters in the Emax model, it follows that one can restrict the search of optimal designs to three point designs with at least one boundary point as support point (as the index should be smaller than or equal to 5/2) or to four point designs containing both boundary points in its support.

Theorem 3.2 *Assume that the model for efficacy is given by* $\eta_1^e(d,\theta_1) = \vartheta_0^e + \vartheta_1^e d + \vartheta_2^e d^2$ *and let* ξ *denote an arbitrary design on the dose range* $\mathcal{D} = [L,R]$.

(a) If $\eta_1^t(d,\theta_1) = \vartheta_0^t + \vartheta_1^t d + \vartheta_2^t d^2$, *there exists a design* ξ^* *with at most three support points, such that* $M_1(\xi^*,\theta_1) \geq M_1(\xi,\theta_1)$. *If the index of the design* ξ *satisfies* $I(\xi) \geq 2$, ξ^* *can be chosen such that the support of* ξ^* *contains the points* L *and* R.

(b) If $\eta_1^t(d,\theta_1) = \frac{\vartheta_1^t d}{\vartheta_2^t + d}$, *there exists a design* ξ^* *with at most five support points, such that* $M_1(\xi^*,\theta_1) \geq M_1(\xi,\theta_1)$. *If the index of the design* ξ *satisfies* $I(\xi) \geq 4$, ξ^* *can be chosen such that the support of* ξ^* *contains the points* L *and* R.

(c) If $\eta_1^t(d,\theta_1) = \vartheta_0^t + \frac{\vartheta_1^t d}{\vartheta_2^t + d}$, *there exists a design* ξ^* *with at most five support points, such that* $M_1(\xi^*,\theta_1) \geq M_1(\xi,\theta_1)$. *If the index of the design* ξ *satisfies* $I(\xi) \geq 4$, ξ^* *can be chosen such that the support of* ξ^* *contains the points* L *and* R.

Theorem 3.3 *Assume that the model for efficacy is given by* $\eta_1^e(d,\theta_1) = \frac{\vartheta_1^e d}{\vartheta_2^e + d}$ *and let* ξ *denote an arbitrary design on the dose range* $\mathcal{D} = [L,R]$.

(a) If $\eta_1^t(d,\theta_1) = \frac{\vartheta_1^t d}{\vartheta_2^t + d}$ *with* $\vartheta_2^e \neq \vartheta_2^t$, *there exists a design* ξ^* *with at most five support points, such that* $M_1(\xi^*,\theta_1) \geq M_1(\xi,\theta_1)$. *If the index of the design* ξ *satisfies* $I(\xi) \geq 4$, ξ^* *can be chosen such that the support of* ξ^* *contains the point* R.

(b) If $\eta_1^t(d,\theta_1) = \vartheta_0^t + \frac{\vartheta_1^t d}{\vartheta_2^t + d}$ *with* $\vartheta_2^e \neq \vartheta_2^t$, *there exists a design* ξ^* *with at most five support points, such that* $M_1(\xi^*,\theta_1) \geq M_1(\xi,\theta_1)$. *If the index of the design* ξ *satisfies* $I(\xi) \geq 4$, ξ^* *can be chosen such that the support of* ξ^* *contains the points* L *and* R.

Theorem 3.4 *Assume that the model for efficacy is given by* $\eta_1^e(d,\theta_1) = \vartheta_0^e + \frac{\vartheta_1^e d}{\vartheta_2^e + d}$ *and let* ξ *denote an arbitrary design on the dose range* $\mathcal{D} = [L,R]$. *If* $\eta_1^t(d,\theta_1) = \vartheta_0^t + \frac{\vartheta_1^e d}{\vartheta_2^e + d}$ *with* $\vartheta_2^e \neq \vartheta_2^t$, *there exists a design* ξ^* *with at most five support points, such that* $M_1(\xi^*,\theta_1) \geq M_1(\xi,\theta_1)$. *If the index of the design* ξ *satisfies* $I(\xi) \geq 4$, ξ^* *can be chosen such that the support of* ξ^* *contains the points* L *and* R.

Remark 3.2 The remaining cases can be obtained by interchanging the roles of η^e and η^t in Theorem 3.1 - 3.4. For example, consider the case, where $\eta_1^e(d,\theta_1)$ is the Emax and $\eta_1^t(d,\theta_1)$ the Michaelis-Menten model with $\vartheta_2^e \neq \vartheta_2^t$. Then it follows from Theorem 3.3(b) that for any design ξ there exists a design ξ^* with at most five support points, such that $M_1(\xi^*,\theta_1) \geq M_1(\xi,\theta_1)$. Moreover, if the index of the design ξ satisfies $I(\xi) \geq 4$, then ξ^* can be chosen such that the support of ξ^* contains L and R. The other cases are obtained in the same way.

3.2 Minimally supported D-optimal designs

For a design ξ let $\#\,\mathrm{supp}(\xi)$ be the number of its support points and let

$$m^* = \min\{\#\,\mathrm{supp}(\eta) \mid \det(M_1(\eta,\theta_1)) > 0,\ \eta \text{ design on } \mathcal{D}\}$$

be the minimal number of support points required for a design with a non-singular information matrix in model (2.1). A design ξ is called minimally supported if $\det(M_1(\xi,\theta_1)) > 0$ and the number of support points is given by m^*. Minimally supported designs are useful if, for example, a drug under investigation may be only available at few dose levels.
In general, the optimal designs have to be found numerically for complex models and even then many of the current algorithms may not work well. However, if one restricts the search to minimally supported designs, the optimization problem can be greatly simplified which may then allow us to determine locally D-optimal designs. In some cases these minimally supported optimal designs may not be optimal among all designs [see Section 5 below for some examples] so that an equivalence theorem must be used to confirm its optimality among all designs or its efficiency should be evaluated using the estimate (2.6). Before we present analytically derived minimally supported designs for model (2.1) for different efficacy-toxicity regression models, we give a result about the general structure of these designs.

Theorem 3.5 *If the number of parameters in the mean function of the efficacy model is the same as the number of parameters in the mean function of the toxicity model, i.e.* $s_1^e = s_1^t$, *the minimally supported locally D-optimal design for model* (2.1) *is a uniform design. Moreover, its support points do not depend on the entries in the covariance matrix* Σ_1.

The following result provides minimally supported D-optimal designs for several commonly used dose-response models. Its proof makes use of Theorem 3.5, which reduces the optimization problem to the determination of the support points.

Theorem 3.6

(1) *Assume that the model for efficacy is given by* $\eta_1^e(d,\theta_1) = \vartheta_0^e + \vartheta_1^e d$.

(1a) If $\eta_1^t(d,\theta_1) = \vartheta_0^t + \vartheta_1^t d$, *the minimally supported D-optimal design is a two-point design with equal masses at the points L and R.*

(1b) *If* $\eta_1^t(d,\theta_1) = \frac{\vartheta_1^t d}{\vartheta_2^t + d}$, *the minimally supported D-optimal design is a two-point design with equal masses at the points* $L \vee \frac{1}{2}(\sqrt{R^2 + 10R\vartheta_2^t + 9(\vartheta_2^t)^2} - R - 3\vartheta_2^t)$ *and* R.

(2) *Assume that the model for efficacy is given by* $\eta_1^e(d,\theta_1) = \vartheta_0^t + \vartheta_1^t d + \vartheta_2^t d^2$.

(2a) *If* $\eta_1^t(d,\theta_1) = \vartheta_0^t + \vartheta_1^t d + \vartheta_2^t d^2$, *the minimally supported D-optimal design is a three-point design with equal masses at the points* L, $\frac{L+R}{2}$ *and* R.

(2b) *If* $\eta_1^t(d,\theta_1) = \vartheta_0^t + \frac{\vartheta_1^t d}{\vartheta_2^t + d}$, *the minimally supported D-optimal design is a three-point design with equal masses at the points* L, $\sqrt{(L+\vartheta_2^t)(R+\vartheta_2^t)} - \vartheta_2^t$ *and* R.

(3) *Assume that the model for efficacy is given by* $\eta_1^e(d,\theta_1) = \frac{\vartheta_1^e d}{\vartheta_2^e + d}$.

(3a) *If* $\eta_1^t(d,\theta_1) = \vartheta_0^t + \vartheta_1^t d$, *the minimally supported D-optimal design is a two-point design with equal masses at the points* $L \vee \frac{1}{2}(\sqrt{R^2 + 10R\vartheta_2^e + 9(\vartheta_2^e)^2} - R - 3\vartheta_2^e)$ *and* R.

(3b) *If* $\eta_1^t(d,\theta_1) = \frac{\vartheta_1^t d}{\vartheta_2^t + d}$, *the minimally supported D-optimal design is a two-point design with equal masses at the optimal points* $L \vee \frac{\sqrt{R\vartheta_2^e\vartheta_2^t(R+\vartheta_2^e+\vartheta_2^t)+(\vartheta_2^e\vartheta_2^t)^2} - \vartheta_2^e\vartheta_2^t}{(R+\vartheta_2^e+\vartheta_2^t)}$ *and* R.

(4) *Assume that the model for efficacy is given by* $\eta_1^e(d,\theta_1) = \vartheta_0^e + \frac{\vartheta_1^e d}{\vartheta_2^e + d}$.

(4a) *If* $\eta_1^t(d,\theta_1) = \vartheta_0^t + \vartheta_1^t d + \vartheta_2^t d^2$, *the minimally supported D-optimal design is a three-point design with equal masses at the points* L, $\sqrt{(L+\vartheta_2^e)(R+\vartheta_2^e)} - \vartheta_2^e$ *and* R.

(4b) *If* $\eta_1^t(d,\theta_1) = \vartheta_0^t + \frac{\vartheta_1^t d}{\vartheta_2^t + d}$, *the minimally supported D-optimal design is a three-point design with equal masses at the points* L, $\frac{\sqrt{(L+\vartheta_2^e)(L+\vartheta_2^t)(R+\vartheta_2^e)(R+\vartheta_2^t)} + LR - \vartheta_2^e\vartheta_2^t}{L+R+\vartheta_2^e+\vartheta_2^t}$ *and* R.

4. Active-controlled dose-finding trials

The use of active controls instead of placebo in clinical trials has received considerable attention in the literature [see Temple and Ellenberg (2000) and Splawinski and Kuzniar (2004) among many others]. In active controlled dose-finding trials patients are randomized to receive either one of several doses of the new drug or an active control (a marketed drug administered at a specific dose level). Inference issues for active-controlled dose-finding trials were investigated only more recently [see, for example, Helms et al. (2015a,b)]. Dette et al. (2014, 2015) investigated optimal design problems for such trials by determining the optimal number of different dose levels, the individual dose levels within the dose range under investigation and the allocation ratios of patients at each dose level and the active control. Despite the increasing importance of such trials [see Hasselblad and Kong (2001)], there is

virtually no work on developing optimal designs for active-controlled dose-finding trials with efficacy-toxicity outcomes, especially given the fact that designs for placebo controlled trials do not extend directly to active-controlled trials [see Dette et al. (2014)].
Our goal in this section is to design an active-controlled dose-finding trial with a pre-determined total number of patients N by determining the optimal number k of different dose levels for the new drug, their individual dose levels $d_1, \ldots, d_k$, and the optimal number n_1 of patients to be assigned to the new drug, along with the allocation scheme across the recommended doses. The remaining number $n_2 = N - n_1$ of patients are assigned to the the active control, which is assumed to be available at a fixed dose level C. In terms of approximate designs, we have designs of the form

$$\tilde{\xi} = \begin{pmatrix} (d_1, 0) & \ldots & (d_k, 0) & (C, 1) \\ \tilde{\omega}_1 & \ldots & \tilde{\omega}_k & \tilde{\omega}_{k+1} \end{pmatrix} , \tag{4.1}$$

where $\tilde{\omega}_i$ denotes the proportion of patients assigned treated at the i^{th} dose level of the new drug, $i = 1, \ldots, k$ and $\tilde{\omega}_{k+1}$ the proportion of patients treated with the active control, that is $n_2 \approx \tilde{\omega}_{k+1} N$. Here the second component of a design points in (4.1) specifies if patients receive the new drug ("0") or the active control ("1"). Note that the approximate design $\tilde{\xi}$ induces an approximate design of the form

$$\xi = \begin{pmatrix} d_1 & \ldots & d_k \\ \omega_1 & \ldots & \omega_k \end{pmatrix}, \tag{4.2}$$

for the new drug defining $\omega_i = \tilde{\omega}_i/(1 - \tilde{\omega}_{k+1})$. Extending the statistical model from Dette et al. (2014) to the efficacy-toxicity outcomes considered here, we have

$$Y_{ij} = (Y_{ij}^e, Y_{ij}^t)^T \sim \mathcal{N}_2(\eta_1(d_i, \theta_1), \Sigma_1) \; ; j = 1, \ldots, n_{1i}, \tag{4.3}$$

$$Z_j = (Z_j^e, Z_j^t)^T \sim \mathcal{N}_2(\eta_2(\theta_2), \Sigma_2) \; ; j = 1, \ldots, n_2, \tag{4.4}$$

where Y_{ij} denotes the outcome from the jth patient treated with the new drug at dose level d_i, and Z_j the outcome from the jth patient treated with the active control. The two-dimensional vector $\eta_2(\theta_2)$ is the expected outcome, and θ_2 a parameter which varies in a compact parameter space, say Θ_2, and Σ_2 is a 2×2 covariance matrix. The function $\eta_2 : \Theta_2 \to \mathbb{R}^2$ is assumed to be continuously differentiable. Assuming that all observations are independent, it can be shown that the information matrix of a design $\tilde{\xi}$ defined in (4.1) has a block-structure of the form

$$M(\tilde{\xi}, \theta) = \begin{pmatrix} (1 - \tilde{\omega}_{k+1}) M_1(\xi, \theta_1) & \mathbf{0} \\ \mathbf{0} & \tilde{\omega}_{k+1} \mathcal{I}_2(\theta_2) \end{pmatrix}, \tag{4.5}$$

where $\theta = (\theta_1^T, \theta_2^T)^T$ and

$$\mathcal{I}_2(\theta_2) = (\tfrac{\partial}{\partial\theta_2}\eta_2(\theta_2))^T \Sigma_2^{-1} (\tfrac{\partial}{\partial\theta_2}\eta_2(\theta_2))$$

is the Fisher information matrix corresponding to the active control. Following Dette et al. (2015) locally optimal designs for active-controlled dose-finding trials can be obtained from locally optimal designs for ordinary dose-finding trials. We extend this result to the class of admissible designs in Theorem 4.1, whose proof can be found in the Appendix.

Theorem 4.1 *If ξ is an admissible design of the form* (4.2) *in model* (2.1) *and $\tilde{\omega}_{k+1} \in (0,1)$, the design $\tilde{\xi}$ defined in* (4.1) *is an admissible design for the model* (4.3) *with an active control* (4.4).

We now characterize admissible designs for various regression functions in the model (4.3) with an active control. For this purpose we apply Theorem 4.1 to the results from Section 3. We illustrate the methodology in an example with the Michaelis-Menten and Emax model. The other models discussed in Section 3 can be considered in a similar way.

Example 4.1 Suppose that the mean outcome for toxicity is given by an Emax model. We consider two situations, where the efficacy outcome is first modeled by an Emax model and in the second case, is modeled by the Michaelis-Menten model. For the first case, it follows from Theorem 3.4 (d) that admissible designs in trials without an active control have at most five support points. By Theorem 4.1, we conclude that admissible designs in active-controlled trials are of the form (4.1) with at most six support points and a positive weight $\tilde{\omega}_6 \in (0,1)$ for the active control. Similarly, for the second case, it follows from Theorem 3.3 (b) that there exists an admissible design for the corresponding active-controlled trial with at most six support points with a positive weight $\tilde{\omega}_6 \in (0,1)$ for the active control. Moreover, the dose levels for the new drug include the boundary points L and R of the dose range.

In a similar way, ϕ_p-optimal designs for active-controlled trials with efficacy-toxicity outcomes can be obtained. For this purpose we state the following result which can be proved in a similar way as Theorem 1 in Dette et al. (2015) using the block-structure of the matrix $M(\tilde{\xi}, \theta)$ in (4.5).

Proposition 4.1 *Let ξ^* denote the locally ϕ_p-optimal design of the form* (4.2) *in the dose-response model* (4.3) *with masses $w_1^*, \ldots, w_k^*$, at the points $d_1^*, \ldots, d_k^*$, respectively. The design $\tilde{\xi}^*$ with masses $\tilde{w}_1^* = \rho_p(1+\rho_p)^{-1}w_1^*, \ldots, \tilde{w}_k^* = \rho_p(1+\rho_p)^{-1}w_k^*$, and $\tilde{w}_{k+1}^* = (1+\rho_p)^{-1}$ at the points $(d_1^*, 0), \ldots, (d_k^*, 0)$ and $(C, 1)$, respectively, is locally ϕ_p-optimal in the dose-response*

model (4.3) *with an active control* (4.4)*, where*

$$\rho_p = \begin{cases} \frac{(\mathrm{tr}[\{\mathcal{I}_2^{-1}(\theta_2)\}^{-p}])^{1/(p-1)}}{(\mathrm{tr}[\{M_1^{-1}(\tilde{\xi}^*,\theta_1)\}^{-p}])^{1/(p-1)}} & \text{if } p \in (-\infty,1)\setminus\{0\} \\ \frac{s_1}{2} & \text{if } p = 0 \\ \frac{\lambda_{\min}(\mathcal{I}_2(\theta_2))}{\lambda_{\min}(M_1(\tilde{\xi}^*,\theta_1))} & \text{if } p = -\infty \end{cases} . \tag{4.6}$$

We note that Proposition 4.1 can be extended to construct minimally supported designs. In particular, any minimally supported ϕ_p-optimal design of the form (4.2) for the dose response model (4.3) yields a minimally supported ϕ_p-optimal design for the dose response model (4.3) with an active control (4.4) by the transformation described in Proposition 4.1. We conclude this section by constructing minimally supported D-optimal designs for some of the models considered in Section 3.2.

Example 4.2 Assume that the effect of the drug on efficacy and toxicity are both studied using Emax models. The minimally supported D-optimal design for model (4.3) with an active control (4.4) can be obtained from Theorem 3.6 part (4b) and Proposition 4.1. We set $s_1 = 6$ and Theorem 3.6 provides the support points of the minimally supported D-optimal design for the dose-response model (4.3). Proposition 4.1 yields $\tilde{\omega}_4^* = 1/4$ for the proportion of patients treated with the active control. Additionally, the minimally supported D-optimal design for model (4.3) with an active control (4.4) allocates the rest of the patients equally to the new drug at 3 dose levels given by

$$L,\ \frac{\sqrt{(L+\vartheta_2^e)(L+\vartheta_2^t)(R+\vartheta_2^e)(R+\vartheta_2^t)}+LR-\vartheta_2^e\vartheta_2^t}{L+R+\vartheta_2^e+\vartheta_2^t} \text{ and } \ R.$$

In a similar manner explicit results for the other models considered in Section 3.2 can be obtained (and are omitted for space considerations).

5. Examples

We now apply our results from previous sections and construct optimal designs for active controlled trials for three examples. In the first one we determine the locally D-optimal design for a particular scenario of the motivating example in the introduction. The second example compares the D-optimal design with the E-optimal design, which is another type of optimal design sometimes used for making inference on the model parameters. The third example contrasts D-optimal designs with minimally supported D-optimal designs with recommendations on their use in practice from a statistical viewpoint.

If the optimal designs are not minimally supported they usually have to be determined numerically and several algorithms have been proposed in the literature for this purpose. The optimal designs presented in this section are found using particle swarm optimization

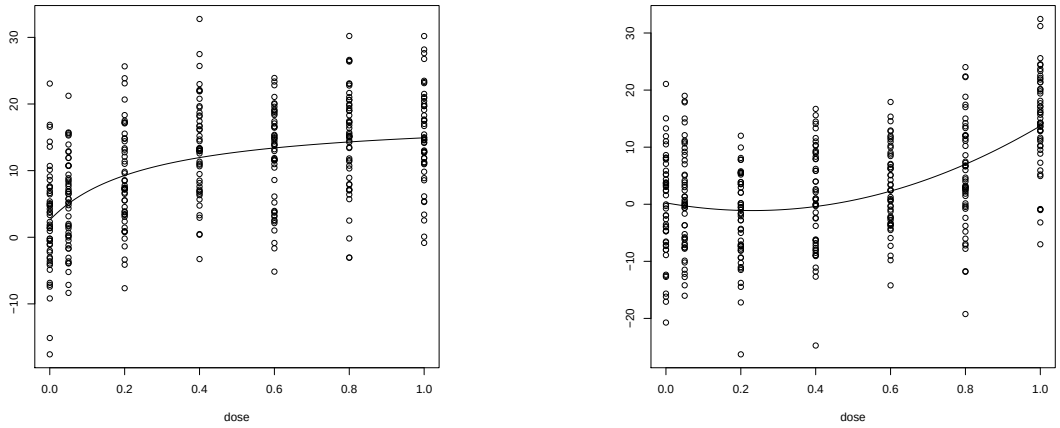

Figure 1: *Fit of an Emax (efficacy) and a quadratic model to the data generated by a model discussed in Tao et al. (2015).*

(PSO), which is a prominent member of the class of nature-inspired metaheuristic algorithms. PSO has been widely used to solve hard and large dimensional optimization problems in engineering and computer science, and it has only been used recently to find optimal designs [see Kim and Li (2011), Chen et al. (2015) or Phoa et al. (2015)]. For space consideration, we omit details on PSO and refer the interested reader to Qiu et al. (2014) and Wong et al. (2015) for details and illustrations.

Example 5.1 Tao et al. (2015) used an Emax-model with parameters $\theta_1^e = (2.5, 14.5, 0.2)^T$ for the mean efficacy outcome and an exponential model $\eta_1^t(d, \theta_1^t) = 0.163 + 0.037e^{(3.3\log(6)d)}$ to model the toxicity effects [see Table 1 in this reference]. As described in Section 3.3.1 of Tao et al. (2015), they used a uniform design to allocate patients to the dose levels 0, 0.05, 0.2, 0.4, 0.6, 0.8, and 1, respectively. For the error distribution in model (2.1) they assumed a two dimensional centered normal distribution with parameters $\rho = 0.4$, $\sigma_e = 7$ and $\sigma_t = 8$. We simulated data according to model (2.1) with sample sizes $n_1 = 350$ for the new drug and fitted an Emax and the quadratic model for efficacy and toxicity, respectively. The quadratic model was used, because it yields a similar shape as the exponential model and minimally supported designs are explicitly available for the combinations of the Emax and a quadratic model. The fits of both regression models to the simulated data are shown in Figure 1. The estimates for the parameters are given by $\hat{\theta}_1^e = (2.588, 15.64, 0.26)$ and $\hat{\theta}_1^t = (0.24, -11.632, 25.11)$ for the Emax and quadratic model, while the estimates for the covariance are obtained as $\hat{\rho} = 0.387$, $\hat{\sigma}_e = 7.272$ and $\hat{\sigma}_t = 8.311$. We used this information to determine a locally D-optimal design for the active controlled trial. Note that we do not require information from the model for the active control for this purpose as we are calculating D-optimal designs [see Proposition 4.1].
By Theorem 3.2(c) and Theorem 4.1, we only need to consider designs with at most six

D-optimal design					minimally supported D-optimal design			
$(0,0)$	$(0.18,0)$	$(0.49,0)$	$(1,0)$	$(C,1)$	$(0,0)$	$(0.31,0)$	$(1,0)$	$(C,1)$
0.09	0.16	0.16	0.09	0.5	0.25	0.25	0.25	0.25

Table 1: *Locally D-optimal design and minimally supported D-optimal design for a situation discussed in Tao et al. (2015). The efficacy is modeled by an Emax model and the toxicity by a quadratic model.*

support points. We first used the PSO algorithm to generate the locally D-optimal design for model (2.1) and in the second step, applied Proposition 4.1 to determine the locally optimal design for the model with an active control. The results are shown in Table 1. The locally D-optimal design has five support points and is therefore not minimally supported. The minimally supported D-optimal design can be obtained from Theorem 3.6 (4a) and is shown in the right part of Table 1. The optimality of the design for the new drug was checked by Theorem 2.1. Figure 2 displays the sensitivity function of the locally D-optimal and the minimally supported D-optimal design. The results confirm its optimality and its non-optimality, respectively. The D-efficiency of the minimally supported designs is given by 0.9886. We note that the lower bound (2.6) for the D-efficiency of the minimally supported optimal design does not need the knowledge of the locally D-optimal design and is given by 0.9532.

The good performance of the minimally supported design is also confirmed by calculating the D-efficiency of the uniform design used in Tao et al. (2015) relative to our locally D-optimal design and the minimally supported D-optimal design. These relative efficiencies are 0.575 and 0.581, respectively, showing that the performance of the design implemented by Tao et al. (2015) could be substantially improved by using locally D-optimal designs.

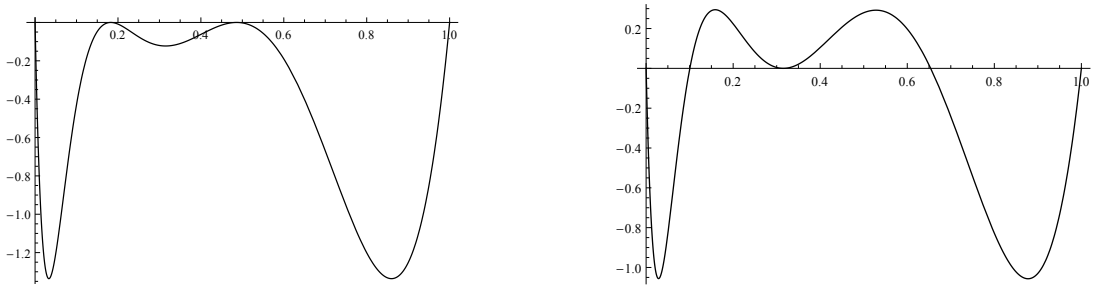

Figure 2: *The sensitivity functions of the locally D-optimal design and the minimally supported D-optimal design for an active-controlled trial. The designs are given in Table 1. The efficacy is modeled by an Emax model and the toxicity by a quadratic model.*

Example 5.2 Consider a situation where the efficacy outcome is described by an Emax model and a Michaelis-Menten model is used for the toxicity outcome. The nominal parameter values are $\theta_1 = (0, 0.466, 25, 300, 50)^T$, and the dose interval is $\mathcal{D} = [0, 150]$. We chose

ρ	D-optimal design				E-optimal design			
0.1	$(0,0)$	$(23.84,0)$	$(150,0)$	$(C,1)$	$(0,0)$	$(19.08,0)$	$(150,0)$	$(C,1)$
	0.16	0.31	0.31	0.22	0.22	0.47	0.25	0.06
0.5	$(0,0)$	$(23.84,0)$	$(150,0)$	$(C,1)$	$(0,0)$	$(19.37,0)$	$(150,0)$	$(C,1)$
	0.16	0.31	0.31	0.22	0.15	0.49	0.31	0.05
0.8	$(0,0)$	$(23.84,0)$	$(150,0)$	$(C,1)$	$(0,0)$	$(18.65,0)$	$(150,0)$	$(C,1)$
	0.16	0.31	0.31	0.22	0.11	0.51	0.33	0.05

Table 2: *Locally D- and E-optimal designs for an active-controlled trial, where the efficacy is modeled by an Emax model and the toxicity by a Michaelis-Menten model. The parameters in the two models are* $\theta_1^e = (0, 0.466, 25)^T, \theta_1^t = (300, 50)^T, \sigma_e = 0.2, \sigma_t = 20, \sigma_e^{AC} = 0.2, \sigma_t^{AC} = 29.8$ *and* $\rho \in \{0.1, 0.5, 0.8\}$.

$\sigma_e = 0.2$, $\sigma_t = 20$ and various values for the correlation in the covariance matrix are considered. By Theorem 3.3(b) and Theorem 4.1, we only need to consider designs with at most six support points. We applied the PSO algorithm to generate the locally D- and E-optimal designs for model (2.1) and Proposition 4.1 to determine the locally optimal designs for the dose finding trial with an active control. The results are shown in Table 2 for various values of the correlation ρ. By definition, an E-optimal design minimizes the maximum eigenvalue of the inverse of the information matrix, whereas a D-optimal design minimizes the volume of the confidence ellipsoid for the parameter. The locally D- and E-optimal designs for the active controlled trial have four support points and are therefore minimally supported. Consequently, the support points of the D-optimal designs do not depend on the elements of the covariance matrix Σ_1, as predicted by Theorem 3.5. On the other hand, the interior support points of the E-optimal design are slightly changing with the correlation ρ. The optimality of both designs was checked by Theorem 2.1 and Figure 3 displays the sensitivity functions of the designs that confirm their optimality for $\rho = 0.1$.

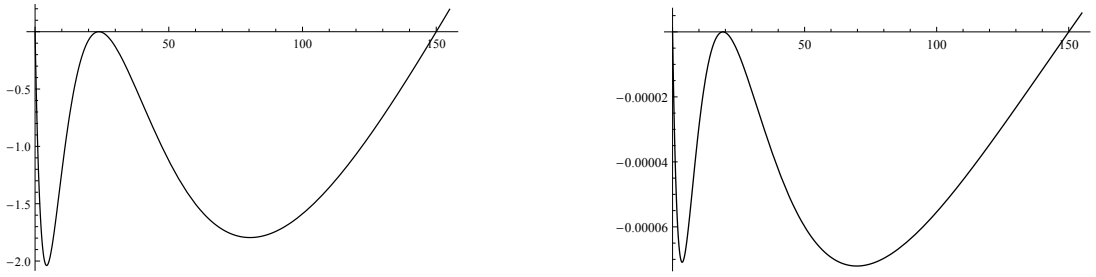

Figure 3: *The sensitivity function of the locally D-optimal design (left) and the locally E-optimal design (right) confirm the optimality of the PSO-generated designs. The efficacy is modeled by an Emax and the toxicity is modeled by a Michaelis-Menten model, where the correlation between efficacy and toxicity is given by* $\rho = 0.1$.

ρ	optimal					minimally supported D-optimal			
0.1	$(0,0)$	$(0.86,0)$	$(3.58,0)$	$(7,0)$	$(C,1)$	$(0,0)$	$(1.94,0)$	$(7,0)$	$(C,1)$
	0.225	0.15	0.15	0.225	0.25	0.25	0.25	0.25	0.25
0.5	$(0,0)$	$(0.8,0)$	$(3.73,0)$	$(7,0)$	$(C,1)$	$(0,0)$	$(1.94,0)$	$(7,0)$	$(C,1)$
	0.2175	0.1575	0.1575	0.2175	0.25	0.25	0.25	0.25	0.25
0.9	$(0,0)$	$(0.7,0)$	$(3.99,0)$	$(7,0)$	$(C,1)$	$(0,0)$	$(1.94,0)$	$(7,0)$	$(C,1)$
	0.21	0.165	0.165	0.21	0.25	0.25	0.25	0.25	0.25

Table 3: *Locally D-optimal design (left) and the minimally supported D-optimal designs (right). The efficacy and toxicity are modeled by a quadratic model with parameter* $\theta_1^e = (0.5, 0.01, 0.1)^T$ *and Emax-model with parameter* $\theta_1^t = (0.1, 2.4, 1.2)^T$*, respectively. The elements in the covariance matrix are* $\sigma_e = 0.1, \sigma_t = 0.4$ *and various correlation values.*

Example 5.3 Assume that the efficacy outcome is described by a quadratic model and the toxicity outcome by an Emax-model, where the nominal values of the model parameters are given by $\theta_1 = (0.5, 0.01, 0.1, 0.1, 2.4, 1.2)^T$. The dose interval is $\mathcal{D} = [0, 7]$ and we chose $\sigma_e = 0.1$ and $\sigma_t = 0.4$. It follows from Theorem 3.2(d) and Theorem 4.1 that only designs with at most six support points have to be considered. The locally D-optimal designs are determined in the same way as described in Example 5.1 and 5.2 and the results are listed in the left part of Table 3 for different values of the correlation. Note that the D-optimal designs are not minimally supported and the support points and weights depend on the correlation. The minimally supported D-optimal designs can be found by an application of Theorem 3.6 and do not depend on ρ [see the right part of Table 3]. The optimality of the numerically calculated D-optimal designs was checked by Theorem 2.1 and the corresponding sensitivity functions are displayed in Figure 4 for different values of the correlation, that is $\rho = 0.1, 0.5$ and 0.9. We observe that all designs calculated by the metaheuristic PSO-algorithm are in fact D-optimal. Moreover, the efficiencies of the minimally supported designs are given by $0.96, 0.81$ and 0.34 for the case $\rho = 0.1, 0.5$, and 0.9, respectively. From the efficiencies we see that the minimally supported designs are only efficient if the efficacy and toxicity outcomes are nearly uncorrelated. For a strong correlation between efficacy and toxicity minimally supported designs cannot be recommended. Finally, we note that the values of the lower bounds in (2.6) for these 3 minimally supported optimal designs are $0.87, 0.67$ and 0.18.

6. Conclusions and further research

In this paper we investigated the optimal design problem for active controlled trials with bivariate outcomes. Upper bounds on the number of support points of locally optimal have been derived, which are used to reduce the dimensionality of the corresponding optimization problems. We also determined minimally supported D-optimal designs explicitly for specific combinations of models for the efficacy and toxicity and note that in general the op-

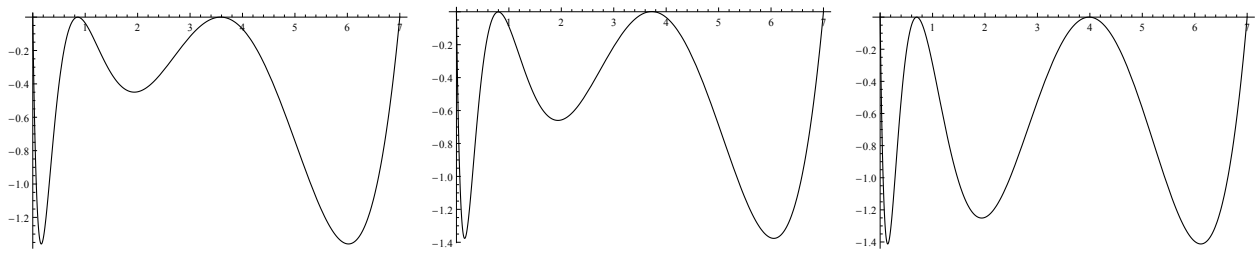

Figure 4: *The sensitivity functions of the locally D-optimal design in Table 3 confirm their optimality.*

timal designs for active controlled clinical trials with bivariate outcomes are not minimally supported. Nevertheless, it is demonstrated that for the models under consideration the minimally supported D-optimal designs are rather efficient, provided that the correlation between efficacy and toxicity is weak. Our results demonstrate that statistical inference in clinical trials with bivariate outcomes can be improved substantially by the appropriate use of efficient designs.

This paper discusses locally optimal designs, which require a-priori information about the unknown model parameters if they appear in the model in a nonlinear way [see Chernoff (1953)]. When preliminary knowledge regarding the unknown parameters of a nonlinear model is available, and the application of locally optimal designs is well justified [see for example Dette et al. (2008)]. Locally optimal designs are typically used as benchmarks for commonly used designs [see the discussion in Example 5.1]. Additionally, locally optimal designs serve as basis for constructing optimal designs with respect to more sophisticated optimality criteria, which are robust against a misspecification of the unknown parameters; see Pronzato and Walter (1985) or Chaloner and Verdinelli (1995), Dette (1997) among others. An interesting direction for future research is to further develop the methodology introduced in the present paper to address uncertainty in the preliminary information for the unknown parameters.

Acknowledgements The authors would like to thank Martina Stein, who typed parts of this manuscript with considerable technical expertise. This work has been supported in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823) of the German Research Foundation (DFG). Kettelhake, Schorning and Wong were partially supported by a grant from the National Institute Of General Medical Sciences of the National Institutes of Health under Award Number R01GM107639. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

7. Appendix

7.1 Proof of Theorems 3.1, 3.2, 3.3 and 3.4

We present the proof of Theorem 3.4 only for the case, where the effect of the drug on efficacy and toxicity is modeled by an Emax model. In this case the gradient of the outcome with respect to the parameter is given by

$$\frac{\partial}{\partial\theta_1}\eta_1(d,\theta_1) \;=\; \begin{pmatrix} 1 & \frac{d}{\vartheta_2^e+d} & -\frac{\vartheta_1^e d}{(\vartheta_2^e+d)^2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \frac{d}{\vartheta_2^t+d} & -\frac{\vartheta_1^t d}{(\vartheta_2^t+d)^2} \end{pmatrix}.$$

It is easy to see that there exists a full column rank matrix $L \in \mathbb{R}^{6\times 10}$ which does not depend on the variable d such that

$$\frac{\partial}{\partial\theta_1}\eta_1(d,\theta_1) = \begin{pmatrix} \nu^T(d) & 0 \\ 0 & \nu^T(d) \end{pmatrix} L^T \;,$$

where the vector $\nu(d)$ is defined by the linearly independent functions in the gradient, i.e.

$$\nu(d) = (1, \tfrac{1}{\vartheta_2^e+d}, \tfrac{1}{(\vartheta_2^e+d)^2}, \tfrac{1}{\vartheta_2^t+d}, \tfrac{1}{(\vartheta_2^t+d)^2})^T \in \mathbb{R}^5.$$

Consequently, we obtain for the information matrix in (2.3) the representation

$$M_1(\xi,\theta_1) = L\int_{\mathcal{D}} \begin{pmatrix} \nu(d)\nu^T(d) & \mathbf{0} \\ \mathbf{0} & \nu(d)\nu^T(d) \end{pmatrix} d\xi(d) L^T, \tag{7.1}$$

where the matrix $\mathbf{0}$ denotes a 5×5 square matrix with all entries 0 and

$$\nu(d)\nu^T(d) = \begin{pmatrix} 1 & \frac{1}{\vartheta_2^e+d} & \frac{1}{(\vartheta_2^e+d)^2} & \frac{1}{\vartheta_2^t+d} & \frac{1}{(\vartheta_2^t+d)^2} \\ \frac{1}{\vartheta_2^e+d} & \frac{1}{(\vartheta_2^e+d)^2} & \frac{1}{(\vartheta_2^e+d)^3} & \frac{1}{(\vartheta_2^e+d)(\vartheta_2^t+d)} & \frac{1}{(\vartheta_2^e+d)(\vartheta_2^t+d)^2} \\ \frac{1}{(\vartheta_2^e+d)^2} & \frac{1}{(\vartheta_2^e+d)^3} & \frac{1}{(\vartheta_2^e+d)^4} & \frac{1}{(\vartheta_2^e+d)^2(\vartheta_2^t+d)} & \frac{1}{(\vartheta_2^e+d)^2(\vartheta_2^t+d)^2} \\ \frac{1}{\vartheta_2^t+d} & \frac{1}{(\vartheta_2^e+d)(\vartheta_2^t+d)} & \frac{1}{(\vartheta_2^e+d)^2(\vartheta_2^t+d)} & \frac{1}{(\vartheta_2^t+d)^2} & \frac{1}{(\vartheta_2^t+d)^3} \\ \frac{1}{(\vartheta_2^t+d)^2} & \frac{1}{(\vartheta_2^e+d)(\vartheta_2^t+d)^2} & \frac{1}{(\vartheta_2^e+d)^2(\vartheta_2^t+d)^2} & \frac{1}{(\vartheta_2^t+d)^3} & \frac{1}{(\vartheta_2^t+d)^4} \end{pmatrix}. \tag{7.2}$$

Now Theorem 14.2.9 in Harville (1997) shows that an improvement with respect to the Loewner ordering can be obtained by improving the common block

$$\int_{\mathcal{D}} \nu(d)\nu^T(d) d\xi(d)$$

in the matrix (7.1). For this purpose we now use Theorem 2.2. The functions $\psi_0(d) = 1$ and

$$\begin{aligned}\psi_1(d) &= \tfrac{1}{\vartheta_2^e+d},\ \psi_2(d) = \tfrac{1}{(\vartheta_2^e+d)^2},\ \psi_3(d) = \tfrac{1}{(\vartheta_2^e+d)^3},\ \psi_4(d) = \tfrac{1}{(\vartheta_2^e+d)^4},\\ \psi_5(d) &= \tfrac{1}{\vartheta_2^t+d},\ \psi_6(d) = \tfrac{1}{(\vartheta_2^t+d)^2},\ \psi_7(d) = \tfrac{1}{(\vartheta_2^t+d)^3}, \psi_8(d) = \tfrac{1}{(\vartheta_2^t+d)^4}.\end{aligned}$$

fulfill the conditions specified in the paragraph before Theorem 2.2. It follows by an application of Theorem 1.1 in Chapter IX of Karlin and Studden (1966) that the sets $\{\psi_0, \ldots, \psi_7\}$ and $\{\psi_0, \ldots, \psi_8\}$ are Tchebycheff systems and Theorem 2.2 is applicable with $k = 8$. Part (A2) of this result yields that there exists a design ξ^* with at most five support points including L and R such that

$$\int_{\mathcal{D}} \nu(d)\nu^T(d)d\xi(d) \leq \int_{\mathcal{D}} \nu(d)\nu^T(d)d\xi^*(d),$$

and the assertion follows. We note that an application of Theorem 3.1 in Dette and Melas (2011) is not possible because the different functions from matrix (7.2) do not form a Tchebycheff system.

□

7.2 Proof of Theorem 3.5

Let ξ be a minimally supported design of the form (4.2). As $s_1^e = s_1^t$ we have $k = s_1^t + 1$. Considering the Cholesky decomposition $\Sigma_1^{-1} = \tilde{\Sigma}\tilde{\Sigma}^T$ of the inverse of the covariance matrix Σ_1 we obtain for the information matrix $M_1(\xi, \theta_1)$ the representation

$$\begin{aligned}M_1(\xi, \theta_1) &= \sum_{i=1}^{k} \omega_i (\tfrac{\partial}{\partial\theta_1}\eta_1(d_i, \theta_1))^T \tilde{\Sigma}\tilde{\Sigma}^T (\tfrac{\partial}{\partial\theta_1}\eta_1(d_i, \theta_1))\\ &= G^T \mathrm{Diag}(\omega_1, \omega_1, \ldots \omega_k, \omega_k) G, \qquad (7.3)\end{aligned}$$

where the matrix G is defined by

$$G = \begin{pmatrix} \tilde{\Sigma}^T(\frac{\partial}{\partial\theta_1}\eta_1(d_1, \theta_1)) \\ \vdots \\ \tilde{\Sigma}^T(\frac{\partial}{\partial\theta_1}\eta_1(d_k, \theta_1)) \end{pmatrix} = (I_k \otimes \tilde{\Sigma}^T) \begin{pmatrix} (\frac{\partial}{\partial\theta_1}\eta_1(d_1, \theta_1)) \\ \vdots \\ (\frac{\partial}{\partial\theta_1}\eta_1(d_k, \theta_1)) \end{pmatrix} \in \mathbb{R}^{2k\times 2k}. \qquad (7.4)$$

and $A \otimes B$ denotes the Kroecker product of the matrices A and B. Now

$$\det(M_1(\xi, \theta_1)) = (\det G)^2 \prod_{i=1}^{k} w_i^2$$

and consequently, the minimally supported D-optimal design must have equal weights. Moreover, the representation

$$\det(G) = \big(\det(\tilde{\Sigma})\big)^k \det\Big(\big(\tfrac{\partial}{\partial\theta_1}\eta_1(d_j,\theta_1)\big)_{j=1,\ldots,k}\Big)$$

shows that the support points of the minimally supported D-optimal design do not depend on the elements of the matrix Σ_1. This completes the proof of Theorem 3.5. □

7.3 Proof of Theorem 3.6

We show only the proof of part 1(b) as the proofs for other cases are similar. If a linear and a Michaelis-Menten model are used to describe the effect of the drug on efficacy and toxicity, at least two support points, say d_1, d_2, are necessary to guarantee invertibility of the information matrix. From Theorem 3.5 it follows $\omega_1^* = \omega_2^* = \frac{1}{2}$. Consider now the determinant of the information matrix of a design ξ with equal weights at the points d_1 and d_2, then it follows by a straightforward calculation that

$$\det(M_1(\xi,\theta_1)) = \frac{{\vartheta_1^t}^2 d_1^2 d_2^2 (d_1-d_2)^4}{16\left(\rho^2-1\right)^2 \sigma_e^4 \sigma_t^4 (\vartheta_2^t + d_1)^4 (\vartheta_2^t + d_2)^4}. \tag{7.5}$$

If we assume w.l.o.g. that $d_1 < d_2$, then the right hand side of (7.5) is a monotone function of d_2. Consequently the right boundary point R is one of the optimal support points, that is $d_2 = R$. Maximizing the remaining expression with respect to the point d_1 in the interval $[L, R]$ gives

$$d_1 = L \vee \frac{1}{2}(\sqrt{R^2 + 10R\vartheta_2^t + 9(\vartheta_2^t)^2} - R - 3\vartheta_2^t),$$

which proves the result. □

7.4 Proof of Theorem 4.1

Assume that $\tilde{\xi}$ is not admissible, that is there exists a design

$$\tilde{\eta} = \begin{pmatrix} (\overline{d}_1, 0) & \ldots & (\overline{d}_l, 0) & (C, 1) \\ \overline{\omega}_1 & \ldots & \overline{\omega}_l & \overline{\omega}_{l+1} \end{pmatrix}$$

such that $M(\tilde{\eta},\theta_1) \neq M(\tilde{\xi},\theta_1)$ and $M(\tilde{\eta},\theta_1) \geq M(\tilde{\xi},\theta_1)$. This yields immediately $\overline{\omega}_{l+1} \geq \tilde{\omega}_{k+1}$ and

$$(1-\overline{\omega}_{l+1}) M_1(\eta,\theta_1) \geq (1-\tilde{\omega}_{k+1}) M_1(\xi,\theta_1),$$

where η denotes the design with masses $\frac{\overline{\omega}_1}{1-\overline{\omega}_{l+1}}, \ldots, \frac{\overline{\omega}_l}{1-\overline{\omega}_{l+1}}$ at the points $\overline{d}_1, \ldots, \overline{d}_l$, respectively. Therefore we obtain

$$(1-\overline{\omega}_{l+1})M_1(\eta,\theta_1) \geq (1-\tilde{\omega}_{k+1})M_1(\tilde{\xi},\theta_1) \geq (1-\overline{\omega}_{l+1})M_1(\xi,\theta_1).$$

Because the design ξ is admissible we have $M_1(\eta,\theta_1) = M_1(\xi,\theta_1)$. Using the block structure of the information matrix and the assumption that the design $\tilde{\xi}$ is not admissible it follows that

$$(\overline{\omega}_{l+1} - \tilde{\omega}_{k+1})M_1(\xi,\theta_1) \leq 0 \quad \text{and} \quad (\tilde{\omega}_{k+1} - \overline{\omega}_{l+1})\mathcal{I}(\theta_2) \leq 0.$$

This yields $\overline{\omega}_{l+1} = \tilde{\omega}_{k+1}$ and $M(\tilde{\eta},\theta_1) = M(\tilde{\xi},\theta_1)$, which is a contradiction to the assumption that the design $\tilde{\xi}$ is not admissible. The desired result follows. □

References

Atashgah, A. B. and Seifi, A. (2007). Application of semi-definite programming to the design of multi-response experiments. *IEE Transactions*, 39(7):763–769.

Atkinson, A. C. (1996). The usefulness of optimum experimental designs. *Journal of the Royal Statistical Society, Ser. B*, 58(1):59–76.

Berger, M. P. F. and Wong, W. K. (2005). *Applied Optimal Designs*. John Wiley & Sons, England.

Boroujerdi, M. (2002). *Pharmacokinetics: Principles and Applications*. McGraw Hill, New York.

Carrothers, T. J. (2011). Decision-making in drug development: application of a clinical utility index. *AAPS*, 1:85–107.

Chaloner, K. and Verdinelli, I. (1995). Bayesian experimental design: A review. *Statistical Science*, 10(3):273–304.

Chang, S. (1997). An algorithm to generate near D-optimal designs for multiple response surface models. *IEE Transactions*, 29(12):1073–1081.

Chen, R. B., Wang, W., Chang, S. P., and Wong, W. K. (2015). Minimax optimal designs via particle swarm optimization methods. *Statistics and Computing*, 25(5):975–988.

Chernoff, H. (1953). Locally optimal designs for estimating parameters. *Annals of Mathematical Statistics*, 24:586–602.

Dette, H. (1996). Lower bounds for efficiencies with applications. In Brunner, E. and Denker, M., editors, *Research Developments in Probability and Statistics: Festschrift zum 65sten Geburtstag von M.L. Puri*, pages 111–124. VSP Utrecht.

Dette, H. (1997). Designing experiments with respect to "standardized" optimality criteria. *Journal of the Royal Statistical Society, Ser. B*, 59:97–110.

Dette, H., Bornkamp, B., and Bretz, F. (2013). On the efficiency of two-stage response-adaptive designs. *Statistics in Medicine*, 32(10):1646–1660.

Dette, H., Bretz, F., Pepelyshev, A., and Pinheiro, J. C. (2008). Optimal designs for dose finding

studies. *Journal of the American Statistical Association*, 103(483):1225–1237.

Dette, H., Kettelhake, K., and Bretz, F. (2015). Designing dose finding studies with an active control for exponential families. *Biometrika; in press.*

Dette, H., Kiss, C., Benda, N., and Bretz, F. (2014). Optimal designs for dose finding studies with an active control. *Journal of the Royal Statistical Society, Ser. B*, 76(1):265–295.

Dette, H. and Melas, V. B. (2011). A note on the de la Garza phenomenon for locally optimal designs. *Annals of Statistics*, 39(2):1266–1281.

Dette, H. and Schorning, K. (2013). Complete classes of designs for nonlinear regression models and principal representations of moment spaces. *Annals of Statistics*, 41(3):1260–1267.

Dragalin, V. and Fedorov, V. (2006). Adaptive designs for dose-finding based on efficacy-toxicity response. *Journal of Statistical Planing and Inference*, 136(6):1800–1823.

Dragalin, V., Fedorov, V. V., and Wu, Y. (2008). Two-stage design for dose-finding that accounts for both efficacy and safety. *Statistics in Medicine*, 27(25):5156–5176.

Fahrmeir, L. and Tutz, G. (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models*. Springer Series in Statistics.

Fan, S. K. and Chaloner, K. (2004). Optimal designs and limiting optimal designs for a trinomial response. *Journal of Statistical Planning and Inference*, 126:347–360.

Harville, D. A. (1997). *Matrix Algebra from a statistician's perspective*. Springer.

Hasselblad, V. and Kong, F. (2001). Statistical methods for compaison to placebo in active-control trials. *Drug Information Journal*, (4):435–449.

Heise, M. A. and Myers, R. H. (1996). Optimal designs for bivariate logistic regression. *Biometrics*, 52(2):613–624.

Helms, H., Benda, N., and Friede, T. (2015a). Point and interval estimators of the target dose in clinical dose-finding studies with active control. *Journal of Biopharmaceutical Statistics (Early View)*, 25(5):939–957.

Helms, H., Benda, N., Zinserling, P., Kneib, T., and Friede, T. (2015b). Spline-based procedures for dose-finding studies with active control. *Statistics in Medicine*, 34(2):232–248.

Karlin, S. and Studden, W. J. (1966). *Tchebycheff Systems: with Applications in Analysis and Statistics*. Wiley and Sons, New York.

Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *Annals of Statistics*, 2:849–879.

Kim, S. and Li, G. (2011). A novel global search algorithm for nonlinear mixed-effects models using particle swarm optimization. *J. Pharmacokinetics and Pharmacodynamics*, 38(4):471–495.

Magnusdottir, B. T. (2013). c-optimal designs for the bivariate emax model. *m0Da 10: Advances in MOdel-Oriented Design and Analysis*, pages 153–161.

Phoa, K. H. F., Chen, R. B., Wang, W. C., and Wong, W. K. (2015). Optimizing two-level supersaturated designs using swarm intelligence techniques. *Technometrics; in press.*

Pronzato, L. and Walter, E. (1985). Robust experimental design via stochastic approximation.

Mathematical Biosciences, 75:103–120.

Pukelsheim, F. (2006). *Optimal Design of Experiments*. SIAM, Philadelphia.

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

Qiu, J. H., Chen, R. B., Wang, W. C., and Wong, W. K. (2014). Using animal instincts to design efficient biomedical studies. *Swarm and Evolutionary Computation Journal*, 18:1–10.

Sagnol, G. (2011). Computing optimal designs of multiresponse experiments reduces to second-order cone programming. *Journal of Statistical Planning and Inference*, 141(5):1684–1708.

Sidorenkov, G. and Navis, G. (2014). Safety of ace inhibitor therapies in patients with chronic kidney disease. *Expert Opinion on Drug Safety*, 13(10):1383–1395.

Splawinski, J. and Kuzniar, J. (2004). Clinical trials: active control vs placebo-what is ethical? *Science and Engineering Ethics*, 10(1):73–79.

Tao, A., Lin, Y., Pinheiro, J., and Shih, W. (2015). Dose finding method in joint modeling of efficacy and safety endpoints in phase ii studies. *International Journal of Statistics and Probability*, 4(1).

Tao, Y., Liu, J., Li, Z., Lin, J., Lu, T., and Yan, F. (2013). Dose-finding based on bivariate efficacy-toxicity outcome using archimedean copula. *PLoS ONE*, 8(11).

Temple, R. and Ellenberg, S. S. (2000). Placebo-controlled trials and active-control trials in the evaluation of new treatments. *Annals of Internal Medicine*, 133(6):455–463.

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

Wong, W. K., Chen, R. B., Huang, C. C., and Wang, W. C. (2015). A modified particle swarm optimization technique for finding optimal designs for mixture models. *PLosOne*, page DOI: 10.1371/journal.pone.0124720.

Yang, M. (2010). On the de la Garza Phenomenon. *Annals of Statistics*, 38(4):2499–2524.

Yang, M. and Stufken, J. (2009). Support points of locally optimal designs for nonlinear models with two parameters. *Annals of Statistics*, 37:518–541.

Yang, M. and Stufken, J. (2012). Identifying locally optimal designs for nonlinear models: A simple extension with profound consequences. *Annals of Statistics*, 40:1665–1685.