

# Optimal designs for dose finding studies with an active control

Holger Dette

Ruhr-Universität Bochum

Fakultät für Mathematik

44780 Bochum, Germany

e-mail: holger.dette@ruhr-uni-bochum.de

Christine Kiss

Ruhr-Universität Bochum

Fakultät für Mathematik

44780 Bochum, Germany

e-mail: christine.kiss@ruhr-uni-bochum.de

Norbert Benda

Federal Institute for

Drugs and Medical Devices

53175 Bonn

e-mail: norbert.benda@bfarm.de

Frank Bretz

Statistical Methodology

Novartis Pharma AG

4002 Basel, Switzerland

e-mail: frank.bretz@novartis.com

September 18, 2011

## Abstract

Dose finding studies often compare several doses of a new compound with a marketed standard treatment as an active control. In the past, however, research has focused mostly on experimental designs for placebo-controlled dose finding studies. To the best of our knowledge, optimal designs for dose finding studies with an active control have not been considered so far. As the statistical analysis for an active controlled dose finding study can be formulated in terms of a mixture of two regression models, the related design problem is different to what has been investigated before in the literature. We present a rigorous approach to the problem of determining optimal designs for estimating the smallest dose achieving the same treatment effect as the active control. We determine explicitly the locally optimal designs for a broad class of models employed in such studies. We also discuss robust design strategies and determine related Bayesian and standardized minimax optimal designs. We illustrate the results by investigating alternative designs for a clinical trial which has recently appeared in a consulting project of one of the authors.

Keywords and Phrases. minimax design, Bayesian optimal designs, dose response, dose estimation, active control

# 1 Introduction

One of the critical steps in developing a medicinal drug is a proper understanding and characterization of its dose response relationship. Failing to characterize well the dose response relationship may have severe consequences once the drug is available to patients: selecting too high doses may lead to unacceptable safety problems, while selecting too low doses may lead to insufficient efficacy. Further applications, where dose response modeling is of particular importance, include the investigation of a new herbicide or fertilizer, a molecular entity, an environmental toxin, or an industrial chemical.

Much literature is available on dose response studies including a placebo group (see Ruberg (1995), Ting (2006), Bretz et al. (2008) among many others). However, in some drug development programs the dose dependent efficacy relative to a standard treatment is of major interest, especially in preparation for an active-controlled confirmatory non-inferiority trial. In addition to regulatory requirements related to drug approval, health technology assessments for national reimbursement decisions may be improved by dose finding studies that evaluate the incremental dose effect as compared to the standard treatment. Furthermore, in many situations the use of placebo could be considered unethical or unfeasible, even in a Phase II dose finding study. If no placebo is used the extrapolation of the dose response from the lowest dose to the zero dose (i.e. placebo) becomes problematic and the use of an active control could facilitate the assessment of the overall efficacy level of the dose response curve.

The considerable interest by regulatory agencies in active-controlled studies becomes evident from several related guidelines. For example, the tripartite ICH E4 guideline on dose finding encourages the inclusion of an active comparator in a dose finding study to improve assay sensitivity of the trial as well as to generate better data on comparative effectiveness and safety (ICH, 1994). In addition, several international disease-specific regulatory guidelines recommend the use of an active comparator in pivotal Phase III studies (EMA, 2006, 2011). In a more general context, the EMA guideline on the choice of a non-inferiority margin states that a placebo-controlled trial is usually not sufficient and that the comparison between test and reference will often be of importance in its own right (EMA, 2005). It thus becomes evident that due to the regulatory requirements on active-controlled Phase III trials, dose finding studies with an active control contribute significantly to the proper choice of a dose to be used in Phase III and lead to a better risk-benefit profile in comparison with a marketed drug.

The research in the present paper is motivated by an active-controlled dose-finding Phase II study to determine the optimal dose of the new compound for the management of acute flare in gout adult patients who are refractory or contraindicated to standard therapies. The primary objective is to determine the target dose of the new compound, which is the dose that leads to the same efficacy as the active control. It will be identified by assessing the dose response relationship of various doses of the new compound with regard to pain intensity in the target joint at 72 hours (Day 4) post-dose measured on a 0 – 100mm Visual Analog

Scale (VAS). Approximately 200 patients will be included in the study. Patients who meet the entry criteria will be randomized to receive either the active control or one dose of the new compound. An important problem consists in specifying the dose levels for the new compound as well as the allocation ratio of patients across all treatments arms in this study. Once the optimal dose of the new compound is selected on the basis of this Phase II trial, Phase III studies will be conducted to evaluate further the efficacy and safety of the new compound in the respective patient population (either acute or chronic gout patients).

It is well known that optimal designs can substantially improve the efficiency of statistical analyzes and numerous authors have worked on the problem of constructing optimal designs for placebo-controlled dose response studies (see Miller et al. (2007), Dragalin et al. (2007), Dette et al. (2008), Dette et al. (2010) among others). However, to our best knowledge, optimal design problems for active controlled dose finding studies have not been considered in the literature so far. In this paper we propose a strategy to obtain efficient designs for such situations. In Section 2 we introduce the statistical model that includes an active control together with several dose levels of the compound under investigation. Locally optimal designs for estimating the target dose for the new compound are constructed explicitly. These designs require a-priori information about the unknown model parameters [see Chernoff (1953), Ford et al. (1992), Fang and Hedayat (2008)] and usually serve as benchmarks for commonly used designs. In addition, 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), Imhof (2001) among others]. In Section 3 we consider standardized minimax and Bayesian optimal designs, which minimize the maximal efficiency and average efficiency over a given range of the unknown parameters, respectively. For several models, including the widely used EMAX model, it is demonstrated that the robust designs are saturated (i.e. the number of different conditions coincides with the number of parameters for the underlying model) and optimal designs with respect to these criteria are determined explicitly. Several examples illustrating the results are presented in Section 4, where we also study the efficiency of commonly used designs for the case study described above. Finally, some conclusions and directions for further research are presented in Section 5, while Section 6 contains the proofs of our main results.

## **2 A statistical model for active-controlled dose finding studies**

We assume that patients are treated either with an active control (a standard treatment administered at a fixed dose level) or with the new drug for which the dose response relationship is unknown. For a given total sample size, say  $n$ , the goal of an experimental design is to allocate  $n_1$  and  $n_2 = n - n_1$  patients to the new drug and the active control, respectively, and to determine the dose levels under which the  $n_1$  patients are treated with

the new drug.

For the statistical analysis we assume that the observations are realizations of independent random variables  $Y_1, \dots, Y_{n_1}, Z_1, \dots, Z_{n_2}$  according to the model

$$\begin{aligned} Y_i &= f(d_i, \theta) + \varepsilon_i & \text{for } 1 \leq i \leq n_1, \\ Z_i &= \mu + \varepsilon_{n_1+i} & \text{for } 1 \leq i \leq n_2, \end{aligned}$$

where  $\varepsilon_1, \dots, \varepsilon_n$  are independent and normally distributed with expectation 0 and variance  $\sigma^2$ . The random variable  $Y_i$  corresponds to a patient receiving the new drug at dose level  $d_i$  ( $i = 1, \dots, n_1$ ) and  $Z_i$  corresponds to a patient receiving the active control (at a fixed single dose level). Furthermore,  $\mu$  is the expected effect for the active control,  $\theta = (\vartheta_0, \dots, \vartheta_p)^T \in \Theta \subset \mathbb{R}^{p+1}$  denotes a vector of unknown parameters and  $f$  is a given function which describes the average response dose of the new drug at a given dose.

Let  $\kappa$  be an indicator, whether a patient receives the new drug ( $\kappa = 0$ ) or the active control ( $\kappa = 1$ ). The design space is therefore given by the set

$$\mathcal{X} = (\mathcal{X}_D \times \{0\}) \cup \{(C, 1)\},$$

where  $\mathcal{X}_D$  denotes the dose range for the new drug,  $C$  is the dose level of the active control and the second component of an experimental condition  $(d, \kappa)$  determines the treatment ( $\kappa = 0, 1$ ). Straightforward calculation shows that the Fisher information at  $(d, \kappa)$  is given by the matrix  $\gamma(d, \kappa)\gamma^T(d, \kappa)$  where the function  $\gamma : \mathcal{X} \rightarrow \mathbb{R}^{p+2}$  is defined by

$$\gamma((d, \kappa), \theta) = \begin{cases} (g^T(d, \theta), 0)^T & \text{if } \kappa = 0 \\ (0^T, 1)^T & \text{if } \kappa = 1 \end{cases} \quad (2.1)$$

and

$$g(d, \theta) = \left( \frac{\partial f(d, \theta)}{\partial \vartheta_0}, \dots, \frac{\partial f(d, \theta)}{\partial \vartheta_p} \right)^T \quad (2.2)$$

denotes the gradient of the regression function  $f(d, \theta)$  with respect to the vector  $\theta$ . 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  $\mathcal{X}$ . Therefore, an experimental design is given by

$$\xi = \begin{pmatrix} (d_1, 0) & \dots & (d_k, 0) & (C, 1) \\ w_1 & \dots & w_k & w_{k+1} \end{pmatrix}$$

where  $w_1, \dots, w_{k+1}$  are positive weights, such that  $\sum_{j=1}^{k+1} w_j = 1$ . The weight  $w_{k+1}$  denotes the relative proportion of total observations treated with the active control,  $d_1, \dots, d_k$  the different dose levels used for the new drug and  $w_j$  the relative proportion of patients treated at dose level  $d_j$  ( $j = 1, \dots, k$ ). The information matrix of an approximate design is given by

the  $(p + 2) \times (p + 2)$  matrix

$$\begin{aligned} M(\xi, \theta) &= \int_{\mathcal{X}} \gamma((d, \kappa), \theta) \gamma^T((d, \kappa), \theta) d\xi(d, \kappa) \\ &= \sum_{j=1}^k w_j \begin{pmatrix} g(d_j, \theta) \\ 0 \end{pmatrix} (g^T(d_j, \theta), 0) + w_{k+1} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} (1 - w_{k+1})\tilde{M}(\tilde{\xi}, \theta) & 0 \\ 0 & w_{k+1} \end{pmatrix} \end{aligned}$$

where the  $(p + 1) \times (p + 1)$  matrix  $\tilde{M}(\tilde{\xi}, \theta)$  is defined by

$$\tilde{M}(\tilde{\xi}, \theta) = \int_{\mathcal{X}_D} g(d, \theta) g^T(d, \theta) d\tilde{\xi}(d) \quad (2.3)$$

and

$$\tilde{\xi} = \begin{pmatrix} d_1 & \dots & d_k \\ \tilde{w}_1 & \dots & \tilde{w}_k \end{pmatrix} \quad (2.4)$$

denotes a design (on the design space  $\mathcal{X}_D$ ) for the new drug with weights

$$\tilde{w}_i = \frac{w_i}{1 - w_{k+1}}, \quad i = 1, \dots, k.$$

In the following we consider models of the form

$$f(d, \theta) = \vartheta_0 + \vartheta_1 f_{\theta_2}(d) \quad (2.5)$$

where  $\theta_2 = (\vartheta_2, \dots, \vartheta_p)^T$ , the function  $f_{\theta_2}$  is assumed to be strictly increasing, see Bretz et al. (2005). The aim is to estimate the target dose  $d^* = d^*(\theta, \mu) = f^{-1}(\mu, \theta) = f_{\theta_2}^{-1}(\frac{\mu - \vartheta_0}{\vartheta_1})$  for a fixed value of  $\mu$ , i.e. the smallest dose of the new compound achieving the same treatment effect as the active control. A natural estimate of  $d^*$  is given by  $\hat{d}^* = f^{-1}(\hat{\mu}, \hat{\theta})$ , where  $(\hat{\theta}, \hat{\mu})$  is the maximum likelihood estimate of the parameter  $\bar{\theta} = (\theta, \mu)$ . Standard calculation shows that the variance of this estimator is approximately given by

$$\text{Var}(\hat{d}^*) \approx \frac{\sigma^2}{n} \psi(\xi, \bar{\theta}),$$

where the function  $\psi$  is defined by

$$\begin{aligned} \psi(\xi, \bar{\theta}) = \psi(\xi, \theta, \mu) &= \nabla d^*(\theta, \mu) M^{-}(\xi, \theta) \nabla d^{*T}(\theta, \mu) \\ &= \frac{1}{1 - w_{k+1}} \nabla_{\theta} d^*(\theta, \mu) \tilde{M}^{-}(\tilde{\xi}, \theta) \nabla_{\theta} d^{*T}(\theta, \mu) + \left( \frac{\partial d^*(\theta, \mu)}{\partial \mu} \right)^2 \frac{1}{w_{k+1}}. \end{aligned} \quad (2.6)$$

Here,  $\nabla$  denotes the gradient of the function  $d^*$  with respect to the parameter  $\bar{\theta} = (\theta, \mu)$ ,  $\nabla_{\theta}$  the gradient with respect to  $\theta$ , and  $M^{-}(\xi, \theta)$  and  $\tilde{M}^{-}(\xi, \theta)$  are generalized inverses of the information matrices  $M(\xi, \theta)$  and  $\tilde{M}(\xi, \theta)$ , respectively. A design  $\xi_{\theta}^*$  is called locally AC-optimal design (for Active Control) if  $\nabla d^*(\theta, \mu) \in \text{Range}(M(\xi, \theta))$  and if  $\xi_{\theta}^*$  minimizes

the function  $\psi(\xi, \bar{\theta})$  among all designs satisfying this condition. The identity in (2.6) holds, because for a design with  $\nabla d^*(\theta, \mu) \in \text{Range}(M(\xi, \theta))$  the variance is independent of the choice of the general inverse [see Pukelsheim (2006)]. Therefore we choose a generalized inverse with the same block structure as the information matrix. Note that the optimality criterion (2.6) is a special case of the  $c$ -optimality criterion, which corresponds to minimizing the expression

$$c^T M^-(\xi, \theta) c, \quad (2.7)$$

for a given vector  $c \in \mathbb{R}^{p+1}$  in the class of all design  $\xi$ , such that  $c$  is estimable by the given design  $\xi$ , i.e.  $\text{Range}(c) \subset \text{Range}(M(\xi, \theta))$ .

In the next result we determine locally AC-optimal designs for a broad class of nonlinear regression models accounting for an active control by minimizing the criterion (2.6). These designs serve as benchmarks for commonly used designs and are the basis for the construction of optimal designs with respect to more sophisticated optimality criteria, in particular standardized minimax and Bayesian optimal designs discussed in the following section. Since the seminal paper of Chernoff (1953) numerous authors have worked on the problem of constructing locally optimal designs for many regression problems [see for example Ford et al. (1992), He et al. (1996), Fang and Hedayat (2008) or Yang (2010) among many others] but - to the best knowledge of the authors - optimal design problems for active controlled dose finding studies have not been considered in the literature. In the following result we show that for all models of the form (2.5) there exist locally AC-optimal designs with two support points independently of the dimension of the parameter vector  $\theta$ . The proof is based on the implicit function theorem and given in the Appendix.

**Theorem 2.1** *For a model of the form (2.5) the optimality criterion defined in (2.6) has the representation*

$$\psi(\xi, \theta, \mu) = \left( \frac{\partial d^*(\theta, \mu)}{\partial \mu} \right)^2 \left( \frac{1}{1 - w_{k+1}} g^T(d^*, \theta) \tilde{M}^-(\tilde{\xi}, \theta) g(d^*, \theta) + \frac{1}{w_{k+1}} \right),$$

where  $d^* = d^*(\theta, \mu) = f_{\theta_2}^{-1}(\frac{\mu - \vartheta_0}{\vartheta_1})$  and the matrix  $\tilde{M}(\tilde{\xi}, \theta)$  and the design  $\tilde{\xi}$  are defined by (2.3) and (2.4), respectively. Moreover, for the model (2.5) with  $d^* \in \mathcal{X}_D$  the design

$$\xi_{\theta}^* = \begin{pmatrix} (d^*, 0) & (C, 1) \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \quad (2.8)$$

is a locally AC-optimal design. In particular, the minimum value of the criterion  $\psi$  defined in (2.6) is given by

$$\min_{\xi} \psi(\xi, \theta, \mu) = \psi(\xi_{\theta}^*, \theta, \mu) = 4 \left( \frac{\partial d^*(\theta, \mu)}{\partial \mu} \right)^2.$$

It is of interest to note that the locally optimal designs determined in Theorem 2.1 are not necessarily unique. To see this we consider the case  $p = 1$ , where we use the notation  $f_{\theta_2} = f_0$

because the function  $f_{\theta_2}$  does not depend on the parameter  $\theta_2$ . In this case the model (2.5) reduces to a linear model

$$f(d, \theta) = \vartheta_0 + \vartheta_1 f_0(d), \quad (2.9)$$

and there also exist locally AC-optimal designs with three or more support points.

**Theorem 2.2** *Let  $d_1, \dots, d_k \in \mathcal{X}_D$  and  $\tilde{w}_1, \dots, \tilde{w}_k \in (0, 1)$  such that*

$$\sum_{j=1}^k \tilde{w}_j = 1 \quad \text{and} \quad \sum_{j=1}^k \tilde{w}_j f_0(d_j) = f_0(d^*). \quad (2.10)$$

*Then the design with masses  $\frac{\tilde{w}_1}{2}, \dots, \frac{\tilde{w}_k}{2}$  and  $\frac{1}{2}$  at the points  $(d_1, 0), \dots, (d_k, 0)$  and  $(C, 1)$  is a locally AC-optimal design for the model (2.9).*

Note that Theorem 2.2 generalizes a result of Herzberg and Cox (1972), who considered the special case  $f_0(d) = d$  and showed that designs with masses  $\frac{1}{4}, \frac{1}{4}$  and  $\frac{1}{2}$  at the points  $(d^* - x, 0), (d^* + x, 0)$  and  $(C, 1)$  are locally AC-optimal designs on the design space  $\mathcal{X} = [L, R]$  whenever  $x \leq \min \{ d^* - L, R - d^* \}$ .

### 3 Robust optimal AC-optimal designs

Locally optimal designs are often sensitive with respect to misspecification of the initial parameters and several alternative design strategies have been developed to address this issue. The literature mainly differentiates between adaptive/sequential [see Chaudhuri and Mykland (1995) or Dragalin et al. (2010) among others] and Bayesian/minimax optimal designs [see for example Chaloner and Verdinelli (1995) and Dette (1997) among others]. In this section we will investigate two robust design strategies for active controlled dose finding studies, namely minimax and Bayesian AC-optimal designs. To be precise let  $\Theta$  and  $\mathcal{M}$  denote sets for the possible values for  $\theta$  and  $\mu$ , respectively. A design  $\xi^*$  is called standardized minimax AC-optimal design for the active control model with respect to the set  $\bar{\Theta} = \{ \bar{\theta} = (\theta, \mu) | \theta \in \Theta, \mu \in \mathcal{M} \}$  if  $\nabla d^*(\theta, \mu) \in \text{Range}(M(\xi, \theta))$  for all  $(\theta, \mu) \in \bar{\Theta}$  and if  $\xi^*$  minimizes the maximum efficiency

$$\Psi_M(\xi) = \max \{ \text{eff}(\xi, \bar{\theta}) \mid \bar{\theta} \in \bar{\Theta} \} \quad (3.1)$$

calculated over a given range  $\bar{\Theta}$  of the parameters. The efficiency here is defined by

$$\text{eff}(\xi, \bar{\theta}) = \frac{\psi(\xi, \bar{\theta})}{\psi(\xi_{\bar{\theta}}^*, \bar{\theta})} \in [1, \infty], \quad (3.2)$$

where  $\psi(\xi, \bar{\theta})$  denotes the criterion function introduced in (2.6) and  $\xi_{\bar{\theta}}^*$  is the locally AC-optimal design. The range  $\bar{\Theta}$  denotes a given set specified by the experimenter which reflects

its prior belief about the unknown vector of parameters. Therefore a design minimizing (3.1) is expected to yield reasonable efficiencies for all values in the set  $\bar{\Theta}$ .

Similarly, the Bayesian AC-optimal design minimizes an average of the efficiencies. To be precise let  $\pi$  denote a prior distribution with support given by  $\bar{\Theta}$ . Then a design  $\xi^*$  is called Bayesian AC-optimal design with respect to the prior  $\pi$  if  $\nabla d^*(\theta, \mu) \in \text{Range}(M(\xi, \theta))$  for all  $(\theta, \mu) \in \bar{\Theta}$  and if  $\xi^*$  minimizes a weighted average of the efficiencies (3.1), that is

$$\int_{\bar{\Theta}} \text{eff}(\xi, \bar{\theta}) d\pi(\bar{\theta}) \quad (3.3)$$

From Theorem 2.1 it follows that the efficiency in (3.1) and (3.3) is given by

$$\text{eff}(\xi, \bar{\theta}) = \frac{1}{4} \left( \frac{1}{1 - w_{k+1}} g^T(d^*, \theta) \tilde{M}^{-1}(\tilde{\xi}, \theta) g(d^*, \theta) + \frac{1}{w_{k+1}} \right)$$

where  $d^* = d^*(\bar{\theta}) = f_{\theta_2}^{-1}(\frac{\mu - \vartheta_0}{\vartheta_1})$ . In general, the determination of standardized minimax and Bayesian optimal designs is a very difficult problem [see for example Imhof (2001) or Braess and Dette (2007)] and in most cases these designs have to be found numerically. In the following discussion we will describe some general properties of these designs and construct robust optimal designs for some models in explicit form. For this purpose we consider again model (2.5) and assume that for each parameter  $\bar{\theta} \in \bar{\Theta}$  the quantity  $d^* = f_{\theta_2}^{-1}(\frac{\mu - \vartheta_0}{\vartheta_1})$  is well defined and an element of the design space, that is

$$\mathcal{X}_0 = \left\{ f_{\theta_2}^{-1} \left( \frac{\mu - \vartheta_0}{\vartheta_1} \right) \mid \bar{\theta} = (\vartheta_0, \vartheta_1, \theta_2^T, \mu) \in \bar{\Theta} \right\} \subset \mathcal{X}_D. \quad (3.4)$$

In other words, for any set of parameters  $\bar{\theta} \in \bar{\Theta}$  there exists a (unique) dose  $d^* = d^*(\theta, \mu) \in \mathcal{X}_D$ , such that  $f(d^*, \theta) = \mu$ . In the following discussion we consider the criterion (3.1) and (3.3), where the parameter space is of the form

$$\bar{\Theta}_\beta = \{ \bar{\theta} = (\vartheta_0, \vartheta_1, \theta_2^T, \mu)^T \in \bar{\Theta} \mid \theta_2^T = \beta \}$$

for some fixed vector  $\beta \in \mathbb{R}^{p-1}$ . Throughout this section we assume that  $\bar{\Theta}$  and  $\bar{\Theta}_\beta$  contain at least two points (otherwise we have a locally optimal criterion and the results of the previous section are applicable). Similarly, if the prior  $\pi$  is supported on the set  $\bar{\Theta}_\beta$  we reflect this in our notation, i.e.  $\pi = \pi_\beta$ , and assume that  $\pi_\beta$  has at least two support points.

### 3.1 Standardized minimax optimal designs

For a fixed  $\theta_2 = \beta$  we introduce for a design of the form (2.4) an induced design on the design space  $\mathcal{Z} = \{ f_\beta(d) \mid d \in \mathcal{X}_D \}$  by

$$\tilde{\eta} = \begin{pmatrix} z_1 & \dots & z_k \\ \tilde{w}_1 & \dots & \tilde{w}_k \end{pmatrix}$$

with  $z_i = f_\beta(d_i) \in \mathcal{Z}$ ,  $i = 1, \dots, k$ . It is easy to see that condition (3.4) is equivalent to

$$\mathcal{Z}_0 = \{ f_\beta(d) \mid d \in \mathcal{X}_0 \} = \left\{ \frac{\mu - \vartheta_0}{\vartheta_1} \mid \bar{\theta} = (\vartheta_0, \vartheta_1, \theta_2, \mu)^T \in \bar{\Theta}_\beta \right\} \subset \mathcal{Z}. \quad (3.5)$$

Observing that for the model (2.5) we have

$$d^*(\bar{\theta}) = f_\beta^{-1} \left( \frac{\mu - \vartheta_0}{\vartheta_1} \right) = f_\beta^{-1}(z^*)$$

with an obvious definition of  $z^* = z^*(\bar{\theta}) = f_\beta(d^*(\bar{\theta}))$  it follows that the efficiency in (3.2) is given by

$$\text{eff}(\xi, \bar{\theta}) = \frac{1}{4} \left( \frac{1}{1 - w_{k+1}} h(z^*, \theta)^T \tilde{M}^-(\tilde{\eta}, \theta) h(z^*, \theta) + \frac{1}{w_{k+1}} \right),$$

where the vector  $h$  is defined by

$$h(z, \theta) = g(f_\beta^{-1}(z), \theta) = \left( 1, z, \frac{\partial f(d, \theta)}{\partial \vartheta_2} \Big|_{d=f_\beta^{-1}(z)}, \dots, \frac{\partial f(d, \theta)}{\partial \vartheta_p} \Big|_{d=f_\beta^{-1}(z)} \right)^T \quad (3.6)$$

and  $\tilde{M}(\tilde{\eta}, \theta) = \sum_{j=1}^k \tilde{w}_j h(z_j, \theta) h^T(z_j, \theta)$ . In model (2.5) we have  $\frac{\partial f}{\partial \vartheta_0} = 1$ ,  $\frac{\partial f}{\partial \vartheta_1} = f_\beta$ ,  $\frac{\partial f}{\partial \vartheta_j} = \vartheta_1 \frac{\partial f_\beta}{\partial \vartheta_j}$  for  $j \geq 2$  and consequently the expression  $h^T(z, \theta) \tilde{M}^-(\tilde{\eta}, \theta) h(z, \theta)$  does not depend on the parameters  $\vartheta_0, \vartheta_1$  and  $\mu$ . Because  $\theta_2 = \beta$  is assumed to be fixed, we reflect this property by the notation

$$h^T(z, \theta) \tilde{M}^-(\tilde{\eta}, \theta) h(z, \theta) = h^T(z) \tilde{M}^-(\tilde{\eta}) h(z) \quad (3.7)$$

where the definition of  $h(z)$  and  $\tilde{M}(\eta)$  depend on the specific context. Therefore the optimization problem for the standardized minimax AC-optimal design with respect to the set  $\Theta_\beta$  reduces to

$$\min_{\xi} \sup_{\bar{\theta} \in \bar{\Theta}_\beta} \frac{\psi(\xi, \bar{\theta})}{\psi(\xi_{\bar{\theta}}^*, \bar{\theta})} = \min_{w_{k+1} \in (0,1)} \min_{\tilde{\eta}} \sup_{z^* \in \mathcal{Z}_0} \frac{1}{4} \left( \frac{1}{1 - w_{k+1}} h^T(z^*) \tilde{M}^-(\tilde{\eta}) h(z^*) + \frac{1}{w_{k+1}} \right). \quad (3.8)$$

As a consequence, the solution of the standardized minimax optimal design can be found in two steps. First, one determines a minimax optimal interpolation design  $\tilde{\eta}^*$  in a regression model with Fisher information  $h(z)h^T(z)$  and design space  $\mathcal{Z}$ , where the range of interpolation is given by  $\mathcal{Z}_0$ , that is

$$\tilde{\eta}^* = \underset{\tilde{\eta}}{\text{argmin}} \sup_{z^* \in \mathcal{Z}_0} h^T(z^*) \tilde{M}^-(\tilde{\eta}) h(z^*). \quad (3.9)$$

Secondly, one determines the optimal weight  $w_{k+1}$ , which specifies the proportion of patients treated with the control. Problems of the type (3.9) have been discussed by Kiefer and Wolfowitz (1964a,b) for polynomial regression models and these papers indicated that explicit solutions of (3.9) are extremely difficult to obtain. Additionally, Dette and O'Brien (1999)

presented further examples. In particular these authors proved the following equivalence Theorem, which can be used to check the optimality of a given design.

**Theorem 3.1** *For a design  $\tilde{\eta}$  on the induced design space  $\mathcal{Z}$  define*

$$d(z, \tilde{\eta}) = h^T(z)M^-(\tilde{\eta})h(z)$$

and

$$\Pi(\tilde{\eta}) = \left\{ z \in \mathcal{Z}_0 \mid d(z, \tilde{\eta}) = \sup_{x \in \mathcal{Z}_0} d(x, \tilde{\eta}) \right\}.$$

Then the following two properties are equivalent:

1.  $\tilde{\eta}^*$  minimizes  $\sup_{z \in \mathcal{Z}_0} d(z, \tilde{\eta})$ .
2. There exists a probability measure  $\pi$  on the set  $\Pi(\tilde{\eta}^*)$ , such that for all  $x \in \mathcal{Z}$  the inequality

$$\int_{\mathcal{Z}_0} d^2(x, z, \tilde{\eta}^*) d\pi(z) \leq \sup_{z \in \mathcal{Z}_0} d(z, \tilde{\eta}^*) \quad (3.10)$$

holds with  $d(x, z, \tilde{\eta}) = h^T(x)\tilde{M}^-(\tilde{\eta})h(z)$ . Moreover, in this case there is equality in (3.10) for each support point of the design  $\tilde{\eta}^*$ .

In Subsection 3.3 and 3.4 we will use this result to identify standardized minimax AC-optimal designs for the model (2.9) and the EMAX model.

## 3.2 Bayesian optimal designs

An analogue of (3.8) is obtained by similar arguments as in the previous section for the Bayesian optimality criterion with respect to the prior  $\pi_\beta$ , that is

$$\begin{aligned} \Psi_B(\xi) &= \int_{\bar{\Theta}} \frac{\psi(\xi, \bar{\theta})}{\psi(\xi_\beta^*, \bar{\theta})} d\pi_\beta(\bar{\theta}) \\ &= \int_{\bar{\Theta}_\beta} \frac{1}{4} \left( \frac{1}{1 - w_{k+1}} g^T(d^*(\bar{\theta}), \theta) \tilde{M}^-(\bar{\xi}, \theta) g(d^*(\bar{\theta}), \theta) + \frac{1}{w_{k+1}} \right) d\pi_\beta(\bar{\theta}) \\ &= \frac{1}{4} \left( \frac{1}{1 - w_{k+1}} \int_{\mathcal{Z}_0} h^T(z) \tilde{M}^-(\tilde{\eta}) h(z) d\pi_\beta^{\mathcal{Z}_0}(z) + \frac{1}{w_{k+1}} \right), \end{aligned} \quad (3.11)$$

where  $\pi_\beta^{\mathcal{Z}_0}$  is the prior induced on the set  $\mathcal{Z}_0$  by the transformation  $z = f_\beta \circ d^*$ . Consequently, the Bayesian AC-optimal design problem can be solved, by first determining the design  $\tilde{\eta}$  minimizing

$$\int_{\mathcal{Z}_0} h^T(z) \tilde{M}^-(\tilde{\eta}) h(z) d\pi_\beta^{\mathcal{Z}_0}(z) = \text{tr}(\tilde{M}^-(\tilde{\eta})A) \quad (3.12)$$

and then determining the optimal weight  $w_{k+1}^*$ , where the matrix  $A$  is given by  $A = \int_{\mathcal{Z}_0} h(z)h^T(z) d\pi_\beta^{\mathcal{Z}_0}(z)$ . Note that the optimality criterion (3.12) is a classical  $A$ -optimality

criterion which has been well studied in the literature. For example, the equivalence theorem states that a design  $\tilde{\eta}$  is A-optimal if and only if the inequality

$$h^T(z)\tilde{M}^-(\tilde{\eta})A\tilde{M}^-(\tilde{\eta})h(z) \leq \text{tr}(\tilde{M}^-(\tilde{\eta})A) \quad (3.13)$$

holds for all  $z \in \mathcal{Z}$ . Moreover, if  $\tilde{\eta}$  is A-optimal, there is equality in (3.13) for all support points of the design  $\tilde{\eta}$ .

### 3.3 Linear model

It follows from the definition of model (2.9) that the vector  $h(z, \theta)$  in (3.7) is given by  $h(z, \theta) = (1, z)^T$ . The following results show that standardized minimax and Bayesian AC-optimal designs for model (2.9) are saturated. The proofs are given in the Appendix.

**Theorem 3.2** *If assumption (3.4) is satisfied, then the standardized minimax AC-optimal design for model (2.9) has at most three support points. Moreover, if  $\mathcal{X}_D = [L, R]$  is an interval, then the standardized minimax AC-optimal design with respect to the set  $\bar{\Theta}$  is given by*

$$\xi^* = \begin{pmatrix} (L, 0) & (R, 0) & (C, 0) \\ \frac{\sqrt{\rho}(2f_0(R)-f_0(d_L)-f_0(d_R))}{2(1+\sqrt{\rho})(f_0(R)-f_0(L))} & \frac{\sqrt{\rho}(f_0(d_L)+f_0(d_R)-2f_0(L))}{2(1+\sqrt{\rho})(f_0(R)-f_0(L))} & \frac{1}{1+\sqrt{\rho}} \end{pmatrix} \quad (3.14)$$

where  $d_L = \min_{d \in \mathcal{X}_0} d$ ,  $d_R = \max_{d \in \mathcal{X}_0} d$ , and the parameter  $\rho$  is given by

$$\rho = \frac{2(2f_0(d_L)f_0(d_R) - f_0(R)(f_0(d_L) + f_0(d_R)) - f_0(L)(f_0(d_L) + f_0(d_R)))}{(f_0(d_L) + f_0(d_R) - 2f_0(L))(f_0(d_L) + f_0(d_R) - 2f_0(R))}.$$

**Theorem 3.3** *If assumption (3.4) is satisfied, then the standardized Bayesian AC-optimal design with respect to the prior distribution  $\pi$  has at most three support points.*

*Moreover, if  $\mathcal{X}_D = [L, R]$  is an interval, then the standardized Bayesian AC-optimal design with respect to the prior  $\pi$  on the set  $\bar{\Theta}$  is given by*

$$\xi^* = \begin{pmatrix} (L, 0) & (R, 0) & (C, 1) \\ p(1-w_3) & (1-p)(1-w_3) & w_3 \end{pmatrix}$$

where

$$p = \frac{\sqrt{1+2\mu_1+\mu_2}}{\sqrt{1+2\mu_1+\mu_2} + \sqrt{1-2\mu_1+\mu_2}}, \quad (3.15)$$

$$w_3 = \frac{1}{1 + \frac{1}{2}(\sqrt{1+2\mu_1+\mu_2} + \sqrt{1-2\mu_1+\mu_2})}$$

and  $\mu_1$  and  $\mu_2$  denote the first and second moment of the induced prior distribution  $\pi^{\mathcal{Z}_0}$  on the set  $\mathcal{Z}_0$ , respectively.

### 3.4 EMAX model

In our second example we consider standardized minimax and Bayesian AC-optimal designs for the EMAX model

$$f(d, \theta) = \vartheta_0 + \frac{\vartheta_1 d}{1 + \vartheta_2 d}, \quad (3.16)$$

for which an explicit determination of standardized minimax AC-optimal designs is substantially harder. A straightforward calculation shows that for model (3.16) the gradient of the regression function can be written in the form

$$g(d, \theta) = h(f_{\vartheta_2}(d), \theta) = (1, f_{\vartheta_2}(d), -\vartheta_1 f_{\vartheta_2}^2(d))^T,$$

where  $f_{\vartheta_2}(x) = x/(1 + \vartheta_2 x)$ . Observing the definition of the induced design space  $\mathcal{Z}$  it follows that the vector  $h$  in (3.6) is given by  $h(z, \theta) = (1, z, -\vartheta_1 z^2)^T$  and (3.7) holds with  $h(z) = (1, z, z^2)^T$ . The following results describe the standardized minimax and Bayesian AC-optimal designs with respect to the set  $\bar{\Theta}_\beta$ .

**Theorem 3.4** *If assumption (3.4) is satisfied, then the standardized minimax AC-optimal design with respect to the set  $\bar{\Theta}_\beta$  for the EMAX model (3.16) has at most four support points. Moreover, if additionally  $\mathcal{X}_D = [L, R]$  and  $\mathcal{X}_0 = [d_L, d_R]$  are intervals and the set*

$$\mathcal{Z}_0 = \left\{ \frac{d}{1 + \beta d} \mid d \in \mathcal{X}_0 \right\}$$

*is symmetric with respect to the center of the interval  $\mathcal{Z} = f_\beta([L, R])$  the following statements hold. If*

$$a = \frac{(d_R - d_L)(1 + \beta L)(1 + \beta R)}{(R - L)(1 + \beta d_L)(1 + \beta d_R)} \geq a_0 := \sqrt{\frac{1}{6}(5 - \sqrt{13})},$$

*then the standardized minimax AC-optimal design is given by*

$$\xi^* = \begin{pmatrix} (L, 0) & \frac{L+R+2\beta LR}{2+\beta(L+R)} & (R, 0) & (C, 1) \\ p(1 - w_4) & (1 - 2p)(1 - w_4) & p(1 - w_4) & w_4 \end{pmatrix} \quad (3.17)$$

*where  $p = \frac{1}{6}(1 + a^2)$  and  $w_4 = \left(1 + \sqrt{\frac{3}{2-a^2}}\right)^{-1}$ . Otherwise, if  $a \leq a_0$ , the standardized minimax AC-optimal design is given by (3.17) where*

$$p = \frac{a^2(1 + a^2) - a(1 - a^2)\sqrt{1 + a^2}}{2(3a^2 - 1)}$$

*and*

$$w_4 = \left(1 + \sqrt{1 - a^2 + 2a^4 + 2a(1 - a^2)\sqrt{1 + a^2}}\right)^{-1}.$$

**Theorem 3.5** *If assumption (3.4) is satisfied, then the standardized Bayesian AC-optimal design with respect to the prior distribution  $\pi_\beta$  has at most four support points.*

Moreover, if  $\mathcal{X}_D = [L, R]$  is an interval and the induced prior distribution  $\pi_\beta^{\mathcal{Z}_0}$  in (3.11) is symmetric with respect to the center of the interval  $\mathcal{Z} = f_\beta([L, R])$ , then the standardized Bayesian AC-optimal design with respect to the prior  $\pi_\beta$  is given by

$$\xi^* = \begin{pmatrix} (L, 0) & \frac{L+R+2\beta LR}{2+\beta(L+R)} & (R, 0) & (C, 1) \\ p(1-w_4) & (1-2p)(1-w_4) & p(1-w_4) & w_4 \end{pmatrix}, \quad (3.18)$$

where

$$p = \frac{\sqrt{\mu_2 + \mu_4}}{2(\sqrt{1 + 2\mu_2 + \mu_4} + \sqrt{\mu_2 + \mu_4})},$$

$$w_4 = \frac{1}{1 + \sqrt{1 + 2\mu_2 + \mu_4} + \sqrt{\mu_2 + \mu_4}},$$

and  $\mu_j = \int_{\mathcal{Z}_0} z^j d\pi_\beta^{\mathcal{Z}_0}(z)$  denotes the  $j$ th moment of the induced prior distribution  $\pi_\beta^{\mathcal{Z}_0}$  ( $j = 2, 4$ ).

## 4 Examples

In this section we illustrate the methodology by calculating several AC-optimal designs and investigate the efficiency of two designs considered by the clinical team for the active controlled dose-finding study described in Section 1.

### 4.1 AC-optimal designs

We determine various AC-optimal designs for the EMAX model (3.16), where the dose range  $\mathcal{X}_D$  is given by the interval  $[10, 150]$ . Information available at the design stage of the dose finding study led to best guesses of the model parameters, namely  $\vartheta_2 = 0.025$ ,  $\vartheta_1 = 1.125$ ,  $\vartheta_0 = 2.5$  and  $\mu = 22.5$ . By Theorem 2.1 the corresponding locally AC-optimal design is given by  $(32, 0)$  and  $(C, 1)$ . That is, 50% of the patients are treated with the new drug (at dose level 32) and the control, respectively. Table 1 displays the locally AC-optimal design for other parameter constellations, as obtained from Theorem 2.1. For each configuration, the corresponding locally AC-optimal design advises the experimenter to treat 50% of the patients with the control and 50% with the new drug at the dose level listed in the table.

$\vartheta_2$	$\vartheta_1$	$\vartheta_0$	$\mu$	<b>local</b>	$\vartheta_2$	$\vartheta_1$	$\vartheta_0$	$\mu$	<b>local</b>
0.025	1.145	2.5	22.5	31.0	0.0283	1.145	2.5	22.5	34.5
0.025	1.12	2.5	23.25	34.2	0.0283	1.12	2.5	23.25	38.6
0.025	1.085	2.5	23.5	37.5	0.0283	1.085	2.5	23.5	42.8

Table 1: *Locally AC-optimal designs for various parameter specifications.*

The locally AC-optimal designs considered in Table 1 were calculated under the assumption

that the elicited parameter values would be the true ones. In order to account for the uncertainty about the parameter values, we next determine several standardized minimax and Bayesian AC-optimal designs. We initially keep the parameter  $\vartheta_2 = 0.025$  fixed and allow some uncertainty for the other parameters. Table 2 displays the results for  $\vartheta_0 \in [1, 4]$  and different example intervals for  $\vartheta_1$  and  $\mu$ . We only show the dose levels and corresponding weights for the new drug, because the proportion of patients treated with the active control is easily calculated from these quantities. For example, if

$$\vartheta_2 = 0.025, \quad \vartheta_1 \in [0.91, 1.33], \quad \mu \in [21.5, 25], \quad (4.1)$$

and the standardized minimax optimal design is used, we obtain from the corresponding row in Table 2 that 12.5%, 32% and 12.5% of the patients should be treated with the new drug at dose levels 10, 39.2 and 150, while the remaining 43% of the patients are treated with the standard treatment. The AC-robust designs reported in Table 2 correspond to a set  $\mathcal{Z}_0$  which is symmetric with respect to the center of the induced design space  $\mathcal{Z}$ . The Bayesian AC-optimal designs have been calculated with respect to the uniform distribution on  $\mathcal{X}_0$  and were determined numerically, even when assuming a fixed parameter  $\vartheta_2 = 0.025$ , because the induced prior distribution is not symmetric. The corresponding standardized minimax AC-optimal designs in Table 2, however, were calculated using Theorem 3.4. For example, in the scenario (4.1) the set  $\mathcal{X}_0$  defined in (3.4) is given by  $\mathcal{X}_0 = [19.61, 77.83]$  which yields  $L = 10, R = 150$  and  $d_L = 19.61, d_R = 77.83$ . The induced design space is given by  $\mathcal{Z} = [8, 31.58]$  and we obtain  $\mathcal{Z}_0 = f_{\vartheta_2}(\mathcal{X}_0) = [13.16, 26.42]$ . Therefore, the standardized AC-optimal design can be directly obtained by an application of Theorem 3.4.

In order to obtain efficient designs which are robust against misspecification of  $\vartheta_2$ , Table 2 also displays the results under the assumption that an interval for the parameter  $\vartheta_2$  is specified, more specifically that  $\vartheta_2 \in [0.023, 0.027]$  and  $\vartheta_2 \in [0.016, 0.025]$ . In this case, the standardized minimax AC-optimal designs have to be determined numerically as well. A comparison of the standardized minimax optimal designs shows only minor differences between the cases  $\vartheta_2 \in [0.023, 0.027]$  and  $\vartheta_2 = 0.025$ . However, if  $\vartheta_2 \in [0.016, 0.025]$  the smallest dose level 10 receives more weight while the highest dose level 150 received less weight. Moreover, the dose level in the interior design space  $\mathcal{X}_D$  is larger. It is interesting to note that the proportion of patients treated with the active control is essentially not changing if there is more uncertainty about the parameter  $\vartheta_2$ . Similar observations can also be made for the Bayesian AC-optimal designs.

If the set  $\mathcal{Z}_0$  is not symmetric all designs have to be calculated numerically, even in the case where the parameter  $\vartheta_2$  is fixed. Table 3 displays results for some examples in this case. There are no major differences between the Bayesian AC-optimal designs for  $\vartheta_2 = 0.025$  and  $\vartheta_2 \in [0.023, 0.027]$ . The differences between the standardized minimax AC-optimal designs are more pronounced. The lowest dose level receives approximately three times more weight if uncertainty about the parameter  $\vartheta_2$  is taken into account in the optimality criterion. In addition, the weight at the dose level in the interior design space is decreased by 25%.

$\vartheta_2$	$\vartheta_1$	$\mu$	Bayes			minimax		
{0.025}	[0.86, 1.43]	[19, 26]	10	35.2	150	10	39.2	150
			0.09	0.41	0.05	0.16	0.28	0.16
	[0.91, 1.33]	[21.5, 25]	10	36.5	150	10	39.2	150
			0.06	0.44	0.04	0.125	0.32	0.125
	[0.97, 1.20]	[23, 24]	10	38.0	150	10	39.2	150
			0.03	0.47	0.03	0.08	0.39	0.08
[0.023, 0.027]	[0.86, 1.43]	[19, 26]	10	35.2	150	10	37.1	150
			0.09	0.41	0.06	0.15	0.26	0.19
	[0.91, 1.33]	[21.5, 25]	10	36.6	150	10	39.7	150
			0.06	0.44	0.04	0.14	0.32	0.11
	[0.97, 1.20]	[23, 24]	10	38.1	150	10	39.2	150
			0.03	0.47	0.03	0.08	0.38	0.10
[0.016, 0.025]	[0.86, 1.43]	[19, 26]	10	33.9	150	10	42.9	150
			0.11	0.40	0.04	0.19	0.27	0.14
	[0.91, 1.33]	[21.5, 25]	10	34.1	150	10	43.1	150
			0.08	0.43	0.03	0.17	0.32	0.09
	[0.97, 1.20]	[23, 24]	10	34.6	150	10	41.1	150
			0.05	0.46	0.02	0.13	0.37	0.07

Table 2: Bayesian and standardized minimax AC-optimal designs for various specifications for the parameter  $\vartheta_1$ ,  $\vartheta_2$  and  $\mu$  where the intercept satisfies  $\vartheta_0 \in [1, 4]$ . The set  $\mathcal{Z}_0$  defined in (3.5) is symmetric with respect to the center of the induced space  $\mathcal{Z}$ . First row: dose levels of the new compound. Second row: proportion of patients treated at these dose levels.

Again the proportion of patients treated with the control is essentially the same in all three scenarios, regardless of whether a Bayesian or standardized minimax approach is employed. Note that the intermediate dose level for the Bayesian AC-optimal is smaller than for the standardized minimax AC-optimal design. Moreover, the weights at the dose levels 10 and 150 are mostly smaller for the Bayesian designs compared with the minimax designs.

## 4.2 Relative efficiencies

In this section we investigate the relative efficiencies of the robust designs determined in Section 4.1 in comparison with two designs considered by the clinical team at the planning stage of the dose finding study. These two standard designs  $S_1$  and  $S_2$  reflect the current practice of allocating patients equally across several doses, which are often chosen to be equally distant in the original or in a logarithmic scale, that is

$$\xi_{S_1} = \begin{pmatrix} (10, 0) & (45, 0) & (80, 0) & (115, 0) & (150, 0) & (C, 1) \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}, \quad (4.2)$$

$$\xi_{S_2} = \begin{pmatrix} (10, 0) & (20, 0) & (39, 0) & (76, 0) & (150, 0) & (C, 1) \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}. \quad (4.3)$$

$\vartheta_2$	Bayes				minimax			
{0.025}	$\xi_{B_1}$	10	34.0	150	$\xi_{M_1}$	10	39.2	150
		0.07	0.44	0.03		0.05	0.44	0.05
[0.023, 0.027]	$\xi_{B_2}$	10	33.9	150	$\xi_{M_2}$	10	39.9	150
		0.07	0.44	0.03		0.14	0.33	0.10
[0.016, , 0.025]	$\xi_{B_3}$	10	31.9	150	$\xi_{M_3}$	10	43.4	150
		0.09	0.42	0.02		0.18	0.34	0.05

Table 3: *Bayesian and standardized minimax AC-optimal designs under the assumption that  $\vartheta_0 \in [1, 2]$ ,  $\vartheta_1 \in [0.92, 1.38]$  and  $\mu \in [20, 23]$ . The range for the parameter  $\vartheta_2$  is shown in the first column. The induced space  $\mathcal{Z}_0$  defined in (3.5) is not symmetric with respect to the center of the induced design space  $\mathcal{Z}$ . First row: dose levels of the new compound. Second row: proportion of patients treated at these dose levels.*

It follows from the discussion in Section 3 that the relative efficiencies depend on the unknown parameter  $\bar{\theta}$  only through the target dose  $d^* = f_{\vartheta_2}^{-1}(\frac{\mu - \vartheta_0}{\vartheta_1})$ . Table 4 exemplarily displays the relative efficiencies of various designs for estimating the target dose  $d^* = 32$ . A design is better if its relative efficiency (which is always larger or equal than 1) is closer to 1. As seen from Table 4 the robust designs achieve considerably better relative efficiencies than the two standard designs. The Bayesian AC-optimal designs usually yield better relative efficiencies for estimating the target dose than the minimax AC-optimal designs. However, this observation depends on the specific target dose (in our case  $d^* = 32$ ). This can be seen in Figure 1, where we exemplarily show the relative efficiency of the standardized minimax AC-optimal design  $\xi_{M_3}$  and Bayesian AC-optimal design  $\xi_{B_3}$  for various values of the target dose  $d^*$  and the parameter  $\vartheta_2$  (these are determined by the specification of the set  $\bar{\Theta}$ ). We observe that the relative efficiencies of the standardized minimax AC-optimal designs vary between 1.6 and 2.0, while the range of efficiencies obtained from the Bayesian AC-optimal design is considerably larger, namely [1.4, 3.9]. If, for example, the target does is  $d^* = 45$ , the standardized minimax AC-optimal design performs better than the Bayesian AC-optimal design.

Finally, we compare these surface with the corresponding relative efficiencies obtained from the two standard designs, as displayed in Figure 2. Both standard designs perform uniformly worse than the standardized minimax AC-optimal design. The efficiency of the equidistant design  $\xi_{S_1}$  varies between 2.5 and 2.9, while the efficiency of the design  $\xi_{S_2}$  is smaller and varies between 2.1 and 2.5. On the other hand the Bayesian AC-optimal design outperforms the two standard designs whenever the target does is less than 45 and larger than 20. If

	$\xi_{S_1}$	$\xi_{S_2}$	$\xi_{B_1}$	$\xi_{B_2}$	$\xi_{B_3}$	$\xi_{M_1}$	$\xi_{M_2}$	$\xi_{M_3}$
$\text{eff}(\xi, \theta)$	2.94	2.40	1.58	1.58	1.66	1.68	1.89	1.95

Table 4: *Relative efficiencies of reference and robust designs for estimating the target dose  $d^* = 32$ .*

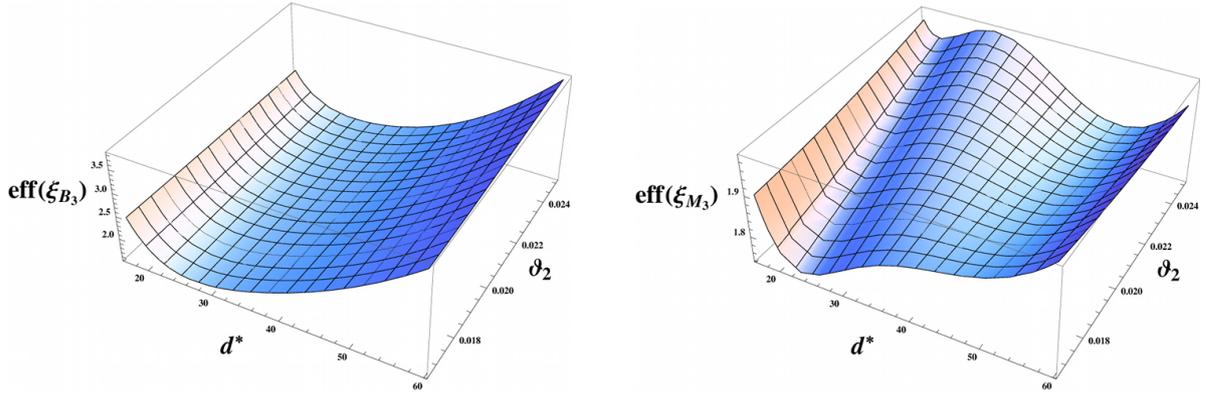


Figure 1: *Relative efficiencies of the designs  $\xi_{B_3}$  and  $\xi_{M_3}$  with varying parameter  $\vartheta_2 \in [0.016, , 0.025]$  and target dose  $d^* \in [16.7, 59.5]$ .*

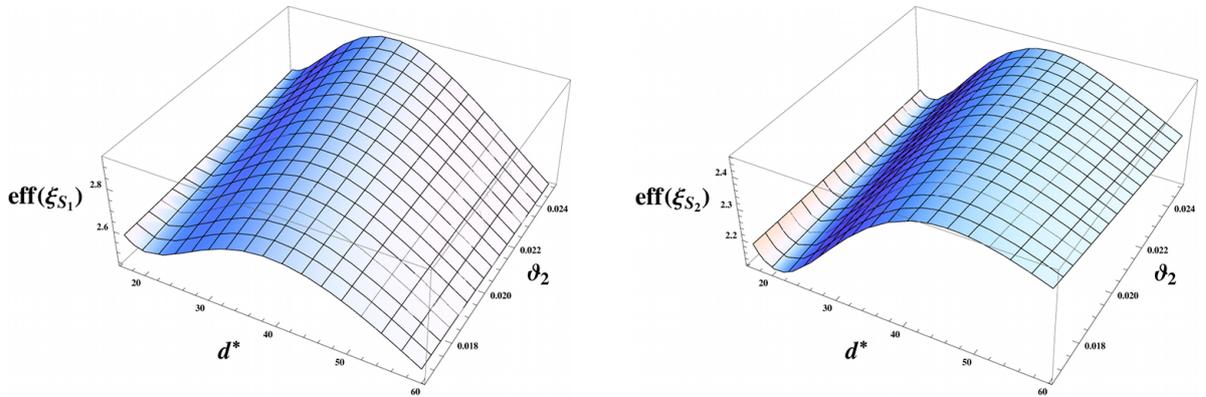


Figure 2: *Relative efficiencies of the references designs  $\xi_{S_1}$  and  $\xi_{S_2}$  with varying parameter  $\vartheta_2 \in [0.016, , 0.025]$  and target dose  $d^* \in [16.7, 59.5]$ .*

$d^* < 20$  there are no substantial differences between the Bayesian AC-optimal and the two standard designs, while the latter have a better performance if  $d^* > 45$ . On the basis of these calculations (and similar results for other parameter specifications, which are omitted here for the sake of brevity), we recommend using a standardized minimax AC-optimal design.

## 5 Conclusion

In this paper we present a rigorous approach for the construction of optimal designs for dose finding studies with an active control. Despite of their practical importance, optimal design problems of this type have not been studied in the literature so far to our best knowledge. Locally optimal designs for estimating the target dose are derived explicitly. These designs are used for the construction of robust designs which require much less prior knowledge about the parameters of the model used for describing the dose response relationship. It is demonstrated that the new designs outperform several standard designs which are currently used in clinical practice.

A challenging future research project in this context is the important problem of model uncertainty which typically appears in this type of investigations, because in many applications it is very difficult to specify an adequate nonlinear model for the description of the dose response relationship. In most cases there exist several competing models (e.g. EMAX, log-linear and logistic) for this purpose. A typical strategy to obtain a good model for the description of the dose response relationship is to test the null hypothesis of a constant dose-response curve at a significance level of, say 5% (adjusted for multiplicity), against the alternative hypothesis of a non-constant dose response curve for each of the four candidate models. Among those models where the null hypothesis is rejected, the model with the highest value of the AIC (Akaike information criterion) will be selected for the estimation of the target dose, see e.g. Bretz et al. (2005). In a recent paper Dette et al. (2008) proposed model robust designs for MED estimation for a class of “classical” dose finding models by maximizing multiple objective criteria [see for example Cook and Wong (1994)]. In the future we plan to adapt this methodology to the problem of constructing AC-optimal designs for several competing active control models [see also Bornkamp et al. (2011) for a Bayesian approach to address model uncertainty in a “classical” dose finding study].

In this paper we focused on one possibility of estimating a target dose that takes the treatment effect of an active control into account. However, alternative target dose definitions might be used in special situations. For example, if the safety profile of the new compound is better than that of the active control, one might be interested in estimating the smallest dose that is not relevantly inferior, i.e.  $d^* = d^*(\theta, \mu, \delta) = f^{-1}(\mu, \theta, \delta) = f_{\theta_2}^{-1}(\frac{\mu - \delta - \vartheta_0}{\vartheta_1})$  for a fixed non-inferiority margin  $\delta > 0$ . Yet a different alternative target dose arises in dose finding studies which include both an active control and placebo. In such situations, one might be interested in a combined objective by estimating the smallest dose that is not worse than the active control and is still better than placebo by a certain clinically relevant amount. A careful inspection of the proofs in Section 6 shows that the methodology can directly be applied to this problem by replacing  $\mu$  with  $\mu - \delta$ ; details are left to the reader.

A different line of research is to derive optimal designs for endpoints that do not follow a normal distribution. For example, in chronic gout studies (which is a different situation than the acute gout study considered in Section 1) the primary endpoint is often defined as the number of flares occurring per subject within 16 weeks of randomization. These flares can be modeled using a negative binomial distribution and a common overdispersion (variance divided by expectation minus 1) for all treatment arms. The logarithm of the expectation of the number of flares during 16 weeks is then described by a regression model for the dose-response relationship between the (single) dose groups of the new compound and a constant parameter for the comparator. Again, we leave the determination of optimal designs in such situations for future research.

## 6 Appendix. Technical details

**Proof of Theorem 2.1:** We will show at the end of this proof that for the models under consideration we have

$$\frac{\partial d^*(\theta, \mu)}{\partial \mu} g(d^*, \theta) = -\nabla_{\theta} d^*(\theta, \mu), \quad (6.1)$$

where  $\nabla_{\theta}$  denotes the gradient with respect to the parameter  $\theta$ . Therefore we obtain for the gradient with respect to  $\bar{\theta} = (\theta, \mu)$

$$\nabla d^*(\theta, \mu) = \left( \nabla_{\bar{\theta}}^T d^*(\theta, \mu), \frac{\partial d^*(\theta, \mu)}{\partial \mu} \right)^T = \frac{\partial d^*(\theta, \mu)}{\partial \mu} (-g^T(d^*, \theta), 1)^T,$$

and the optimality criterion defined in (2.6) is in fact a special case of the  $c$ -optimality criterion (2.7) with the vector

$$c = (-g^T(d^*, \theta), 1)^T. \quad (6.2)$$

It can be shown [see Pukelsheim (2006)] that a design  $\xi$  minimizes  $c^T M^{-}(\xi, \theta) c$  if and only if there exists a generalized inverse  $G$  of  $M^{-}(\xi, \theta)$  such that for all  $(d, \kappa) \in \mathcal{X}$  the inequality

$$c^T M^{-}(\xi, \theta) c \geq (\gamma^T((d, \kappa), \theta) G c)^2 \quad (6.3)$$

is satisfied. Moreover, equality holds in (6.3) for all support points of any locally  $c$ -optimal design. Therefore a design  $\xi_{\bar{\theta}}^*$  is a locally optimal AC-design if and only if there exists a generalized inverse of the matrix  $M(\xi_{\bar{\theta}}^*, \theta)$ , such that the inequalities

$$c^T M^{-}(\xi_{\bar{\theta}}^*, \theta) c \geq ((g^T(d, \theta), 0) G c)^2 \quad (6.4)$$

$$c^T M^{-}(\xi_{\bar{\theta}}^*, \theta) c \geq ((0, 1) G c)^2 \quad (6.5)$$

hold for all  $d \in \mathcal{X}_D$ , where the vector  $c$  is given by (6.2). Note that the inequalities (6.4) and (6.5) correspond to the case  $\kappa = 0$  and  $\kappa = 1$  in (2.7), respectively. The information matrix of the candidate design  $\xi_{\bar{\theta}}^*$  defined in (2.8) is obtained as

$$M(\xi_{\bar{\theta}}^*, \theta) = \frac{1}{2} \begin{pmatrix} g(d^*, \theta) g^T(d^*, \theta) & 0 \\ 0 & 1 \end{pmatrix}.$$

Observing that in model (2.5) the first coordinate of the vector  $g(d, \theta)$  is given by 1 it now follows by a straightforward calculation that the matrix  $G = 2(e_1 e_1^T + e_{p+2} e_{p+2}^T)$  is a generalized inverse of the information matrix  $M(\xi_{\bar{\theta}}^*, \theta)$ , where  $e_i \in \mathbb{R}^{p+2}$  denotes the  $i$ -th unit vector. Using this generalized inverse in the equivalence theorem with the vector  $c = (-g(d^*, \theta)^T, 1)^T$  it follows that the design  $\xi_{\bar{\theta}}^*$  fulfills the inequalities (6.4) and (6.5) because both sides are equal to 4.

We conclude the proof showing the relation (6.1) which is essential for the argument above.

For this purpose we recall the definition of the vector  $\bar{\theta} = (\theta, \mu)^T \in \mathbb{R}^{p+2}$  and consider the function

$$F(\bar{\theta}, d) = \mu - f(d, \theta).$$

By assumption there exists a solution, say  $d^*(\bar{\theta}) = d^*(\theta, \mu)$  of the equation  $F(\bar{\theta}, d) = 0$  with respect to  $d$  and  $\frac{\partial f(d, \theta)}{\partial d} \neq 0$ . Therefore the implicit function theorem implies that the function  $\bar{\theta} \rightarrow d^*(\bar{\theta})$  is differentiable with derivative

$$\begin{aligned} \frac{\partial d^*(\bar{\theta})}{\partial \bar{\theta}} &= - \left( \frac{\partial F(\bar{\theta}, d)}{\partial d} \Big|_{d=d^*(\bar{\theta})} \right)^{-1} \frac{\partial F(\bar{\theta}, d)}{\partial \bar{\theta}} \Big|_{d=d^*(\bar{\theta})} \\ &= - \left( \frac{\partial F(\bar{\theta}, d)}{\partial d} \Big|_{d=d^*(\bar{\theta})} \right)^{-1} (-g^T(d^*, \theta), 1)^T. \end{aligned}$$

This yields

$$\nabla_{\theta} d^*(\bar{\theta}) = \left( \frac{\partial F(\bar{\theta}, d)}{\partial d} \Big|_{d=d^*(\bar{\theta})} \right)^{-1} g(d^*, \theta), \quad \frac{\partial d^*(\bar{\theta})}{\partial \mu} = - \left( \frac{\partial F(\bar{\theta}, d)}{\partial d} \Big|_{d=d^*(\bar{\theta})} \right)^{-1},$$

and proves (6.1), which completes the proof of Theorem 2.1.  $\square$

**Proof of Theorem 2.2:** In model (2.9) the gradient in (2.2) is given by  $g(d, \theta) = (1, f_0(d))^T$  and we obtain from (2.10) and Elfving's theorem [see Elfving (1952)] that the design  $\tilde{\xi}$  defined in (2.4) minimizes  $g^T(d^*, \theta) \tilde{M}^{-1}(\tilde{\xi}, \theta) g(d^*, \theta)$  where the minimum value is given by 1. Therefore it follows from Theorem 2.1 that the design  $\xi_{\theta}^*$  defined in Theorem 2.2 is locally AC-optimal.  $\square$

**Proof of Theorem 3.2:** We first show that the optimal design  $\tilde{\eta}^*$  defined by (3.9) is supported at at most 2 points, which implies the statement regarding the number of support points. For this purpose we apply Theorem 3.1 and recall that for model (2.9) the vector  $h$  in this criterion is given by  $h(z, \theta) = (1, z)^T$ . Consequently, it is easy to see that the left hand side of inequality (3.10) is a polynomial of degree 2. Because equality holds in (3.10) for all support points of  $\tilde{\eta}^*$  this implies that the optimal design  $\tilde{\eta}^*$  has at most 2 support points. Consequently, the standardized minimax AC-optimal design has at most 3 support points.

We will also use Theorem 3.1 to show that the design  $\tilde{\eta}^*$  minimizing (3.9) is of the form

$$\tilde{\eta}^* = \begin{pmatrix} z_L & z_R \\ \frac{z_l + z_r - 2z_R}{2(z_L - z_R)} & 1 - \frac{z_l + z_r - 2z_R}{2(z_L - z_R)} \end{pmatrix} \quad (6.6)$$

where the points  $z_L, z_R, z_l$  and  $z_r$  are defined by  $z_L = f_0(L) = \min_{d \in \mathcal{X}_D} f_0(d)$ ,  $z_R = f_0(R) = \max_{d \in \mathcal{X}_D} f_0(d)$ ,  $z_l = f_0(d_L) = \min_{d \in \mathcal{X}_0} f_0(d)$  and  $z_r = f_0(d_R) = \max_{d \in \mathcal{X}_0} f_0(d)$ , respectively. For this purpose we note that a straightforward but tedious calculation gives  $\Pi(\tilde{\eta}^*) = \{z_l, z_r\}$  in Theorem 3.1. Equivalently,  $h^T(z_l) \tilde{M}^{-1}(\tilde{\eta}^*) h(z_l) = h^T(z_r) \tilde{M}^{-1}(\tilde{\eta}^*) h(z_r)$ , where  $h(z) =$

$(1, z)^T$  and  $\tilde{M}(\tilde{\eta}^*) = \int h(z)h^T(z)d\tilde{\eta}^*(z)$ . This yields

$$\sup_{z \in \mathcal{Z}_0} d(z, \tilde{\eta}^*) = d(z_l, \tilde{\eta}^*) = 1 - \frac{(z_l - z_r)^2}{(z_l + z_r - 2z_L)(z_l + z_r - 2z_R)}$$

and

$$d^2(x, z, \tilde{\eta}^*) = \frac{4(x(2z - z_l - z_r) - z(z_l + z_r) + z_L(z_l + z_r) + (z_l + z_r - 2z_L)z_R)^2}{(-2z_L + z_l + z_r)^2(-2z_R + z_l + z_r)^2}.$$

If  $\pi$  is a prior distribution on the set  $\Pi(\tilde{\eta}^*)$  with weights  $p$  and  $1 - p$  at the points  $z_l$  and  $z_r$ , respectively, we have

$$\int_{z \in \mathcal{Z}_0} d^2(x, z, \tilde{\eta}^*)d\pi(z) = pd^2(x, z_l, \tilde{\eta}^*) + (1 - p)d^2(x, z_r, \tilde{\eta}^*) = P_2(x).$$

If the design  $\tilde{\eta}^*$  in (6.6) was optimal, then it follows from Theorem 3.1 that for  $x = z_L$  there must be equality in (3.10) which determines the weight  $p$ , that is

$$p = \frac{1}{2} \left( 2 + \frac{z_L - z_l}{z_r + z_l - 2z_L} + \frac{z_R - z_l}{z_r + z_l - 2z_R} \right).$$

The function  $P_2(x)$  is a polynomial of degree two, with positive leading coefficient

$$\frac{4(z_l - z_r)^2}{(-2z_L + z_l + z_r)^2(-2z_R + z_l + z_r)^2}$$

and minimum at the point  $\frac{z_L + z_R}{2}$ . Therefore it attains maxima in the set  $\mathcal{Z}$  at the points  $z_L$  and  $z_R$  with value  $d(z_l, \tilde{\eta}^*)$  and by Theorem 3.1 the design  $\tilde{\eta}^*$  is optimal i.e. it minimizes (3.9). For the determination of the standardized minimax AC-optimal design it remains to determine the weight  $w_3$ . Inserting  $\tilde{\eta}^*$  into (3.8) leads to

$$k(w_3) = \frac{1}{1 - w_3} h(z^*)\tilde{M}^{-1}(\tilde{\eta}^*)h(z^*) + \frac{1}{w_3} = \frac{\rho}{1 - w_3} + \frac{1}{w_3},$$

where

$$\rho = \frac{2(2z_l z_r - z_R(z_l + z_r) - z_L(z_l + z_r - 2z_R))}{(z_l + z_r - 2z_L)(z_l + z_r - 2z_R)}.$$

The function  $k$  is minimal for  $w_3 = \frac{1}{1 + \sqrt{\rho}}$  and with definition (2.4) it follows that

$$w_1 = \frac{\sqrt{\rho}(2z_R - z_l - z_r)}{2(1 + \sqrt{\rho})(z_R - z_L)} \quad \text{and} \quad w_2 = \frac{\sqrt{\rho}(z_l + z_r - 2z_L)}{2(1 + \sqrt{\rho})(z_R - z_L)}.$$

Resubstitution of  $z_l, z_r, z_L$  and  $z_R$  shows that the design given by (3.14) is standardized minimax AC-optimal.  $\square$

**Proof of Theorem 3.3:** It follows from the discussion in Section 3.2 that the Bayesian AC-optimal design can be found by solving the  $A$ -optimal design problem (3.12), where  $A = (\mu_{i+j})_{i,j=0,1} = (\int_{\mathcal{Z}_0} z^{i+j} d\pi_{\beta}^{\mathcal{Z}_0}(z))_{i,j=0,1}$  and  $h(z) = (1, z)^T$ . The statement regarding the support points is obtained in the same way as given in the proof of Theorem 3.2 using the inequality (3.13). A further application of (3.13) shows that the design  $\tilde{\eta}^*$  with masses  $p$  and  $1 - p$  at the points  $-1$  and  $1$  minimizes (3.12), where the weight  $p$  is given by (3.15). The corresponding minimal value is given by

$$\text{tr}(M^{-1}(\tilde{\eta}^*)A) = \frac{1}{4} \left( \sqrt{1 - 2\mu_1 + \mu_2} + \sqrt{1 + 2\mu_1 + \mu_2} \right)^2$$

Therefore the assertion follows by minimizing the expression (3.11) with respect to the remaining weight  $w_3$  and transforming the design  $\tilde{\eta}^*$  onto the design space  $\mathcal{X}_D$ .  $\square$

**Proof of Theorem 3.4:** The statement regarding the number of support points follows along the lines given in the proof of Theorem 3.2, where in this case the vector  $h$  in (3.7) is given by  $h(z) = (1, z, -z^2)$ . In order to prove the second part of the result we note that the set  $\mathcal{Z}_0$  is an interval and consequently the design  $\eta^*$  minimizing (3.9) has a nonsingular information matrix  $M(\tilde{\eta}^*)$  and has therefore exactly 3 support points. Note also that the criterion  $d(z, \tilde{\eta}^*) = h^T(z)M^{-1}(\tilde{\eta}^*)h(z)$  is invariant with respect to linear transformations of the form  $z \rightarrow \alpha z + \beta$  of the set  $\mathcal{Z}_0$  and  $\mathcal{Z}$  and therefore we assume without loss of generality that  $\mathcal{Z} = [-1, 1]$  and  $\mathcal{Z}_0 = [-a, a]$ . From a standard convexity argument it follows that the design  $\tilde{\eta}^*$  minimizing

$$\max_{z \in [-a, a]} h^T(z) \tilde{M}^{-1}(\tilde{\eta}) h(z)$$

is symmetric with masses  $p, 1 - 2p, p$  at the points  $-1, 0, 1$  and inverse information matrix

$$\tilde{M}^{-1}(\tilde{\eta}^*) = \frac{1}{2p(1 - 2p)} \begin{pmatrix} 2p & 0 & -2p \\ 0 & 1 - 2p & 0 \\ -2p & 0 & 1 \end{pmatrix}.$$

Recalling the definition of the optimization problem (3.8) we obtain that the function  $z \rightarrow h^T(z) \tilde{M}^{-1}(\tilde{\eta}^*) h(z)$  is a symmetric polynomial of degree 4 with positive leading coefficient. Therefore the maximum in (3.9) is attained at most at the two boundary points of the set  $\mathcal{Z}_0$  and one interior point of  $\mathcal{Z}_0$ , and it follows by symmetry that we have to distinguish three cases for the set  $\Pi(\tilde{\eta}^*)$  in Theorem 3.1, that is

$$(1) \quad \Pi(\tilde{\eta}^*) = \{-a, 0, a\}, \quad (2) \quad \Pi(\tilde{\eta}^*) = \{-a, a\}, \quad (3) \quad \Pi(\tilde{\eta}^*) = \{0\}.$$

In the first case we note that the identity  $d(a, \tilde{\eta}^*) = d(0, \eta^*)$  determines the weight  $p$ , i.e.  $p = \frac{1}{6}(1 + a^2)$  with value

$$\max_{z \in [-a, a]} d(z, \tilde{\eta}^*) = d(0, \eta^*) = \frac{3}{2 - a^2}. \quad (6.7)$$

In the second case the condition

$$d(-a, \tilde{\eta}^*) = d(a, \tilde{\eta}^*) = \frac{2p + (1 - 6p)a^2 + a^4}{2p(1 - 2p)} \quad (6.8)$$

is satisfied for any symmetric design and cannot be used to determine  $p$  directly. However, the optimal design would minimize  $\max_{z \in [-a, a]} d(z, \tilde{\eta}) = d(a, \tilde{\eta})$ , and this yields

$$p = \frac{a^2 + a^4 - \sqrt{a^2(1 - a^2)(1 - a^4)}}{2(-1 + 3a^2)} \quad (6.9)$$

with value

$$\max_{z \in [-a, a]} d(z, \tilde{\eta}^*) = d(\pm a, \tilde{\eta}^*) = 1 - a^2 + 2a^4 + 2a(1 - a^2)\sqrt{1 + a^2}. \quad (6.10)$$

Finally, for the third case note that by (6.8) the inequality  $(1 - 2p)^{-1} = d(0, \tilde{\eta}^*) > d(\pm a, \tilde{\eta}^*)$  is equivalent to the inequality  $p \geq (1 + a^2)/6$ . Therefore minimizing  $(1 - 2p)^{-1}$  with respect to this condition gives  $p = (1 + a^2)/6$  and a design satisfying (1), that is  $\Pi(\tilde{\eta}^*) = \{-a, 0, a\}$ . Consequently, the third case can not occur and we only have to compare the results corresponding to cases (1) and (2), for which the optimal values are given by (6.7) and (6.10), respectively. A simple calculation shows that the inequality

$$\frac{3}{2 - a^2} \leq 1 - 2a^2 + 2a^4 + 2a(1 - a^2)\sqrt{1 + a^2}$$

holds in the interval  $[0, 1]$  if and only if the inequality  $a \geq a_0 = \sqrt{(5 - \sqrt{13})/6} \approx 0.482087$  is satisfied. It follows that the criterion (3.9) is minimized for the design  $\tilde{\eta}^*$  with masses  $p, 1 - 2p, p$  at the points  $-1, 0, 1$ , respectively, where the weight  $p$  is given by  $(1 + a^2)/6$  if  $a \geq a_0$  and (6.9) if  $a < a_0$ . The assertion of Theorem 3.4 finally follows by transforming these results to the original design space (note that the transformation is nonlinear).  $\square$

**Proof of Theorem 3.5:** We may assume without loss of generality that  $\mathcal{Z} = [-1, 1]$ . Now by assumption the induced prior distribution is symmetric and as a consequence the elements in the matrix  $A = (\int_{\mathcal{Z}_0} z^{i+j} d\pi_{\beta}^{\mathcal{Z}_0}(z))_{i,j=0}^2$  vanish, whenever  $i + j$  is odd. Therefore a standard argument shows that there exists a symmetric design  $\tilde{\eta}^*$  minimizing (3.12) and in the following we will investigate if a symmetric design supported at the points  $-1, 0$ , and  $1$  is optimal. For such a design the optimal weights can be computed using a result from chapter 8 in Pukelsheim (2006), which gives for the corresponding weights

$$\tilde{w}_1 = \tilde{w}_3 = \frac{\sqrt{\mu_2 + \mu_4}}{2(\sqrt{1 + 2\mu_2 + \mu_4} + \sqrt{\mu_2 + \mu_4})}, \quad \tilde{w}_2 = \frac{\sqrt{1 + 2\mu_2 + \mu_4}}{\sqrt{1 + 2\mu_2 + \mu_4} + \sqrt{\mu_2 + \mu_4}},$$

respectively, where  $\mu_j$  denotes the  $j$ th moment of the distribution  $\pi_{\beta}^{\mathcal{Z}_0}$ . Finally, a straightforward application of the equivalence theorem (3.13) shows that the design with weights  $\tilde{w}_1, \tilde{w}_2$

and  $\tilde{w}_3$  at the points  $-1, 0$  and  $1$  is in fact minimizing (3.12). The value of the corresponding criterion in (3.12) is given by  $(\sqrt{1 + 2\mu_2 + \mu_4} + \sqrt{\mu_2 + \mu_4})^2$ . Minimizing the criterion (3.11) with respect to the remaining weight  $w_4$  and transforming the support points from the induced design space  $\mathcal{Z}$  to the given design space  $\mathcal{X}_D$  shows that the Bayesian AC-optimal design with respect to the prior  $\pi_\beta$  is given by (3.18), which completes the proof of Theorem 3.5.  $\square$

**Disclaimer for Norbert Benda:** The views expressed in this article are views of the author and do not necessarily reflect the views of the BfArM.

**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).

## References

- Bornkamp, B., Bretz, F., Dette, H., and Pinheiro, J. (2011). Response-adaptive dose-finding under model uncertainty. *Annals of Applied Statistics, to appear*.
- Braess, D. and Dette, H. (2007). On the number of support points of maximin and Bayesian  $D$ -optimal designs in nonlinear regression models. *Annals of Statistics*, 35:772–792.
- Bretz, F., Pinheiro, J., and Branson, M. (2005). Combining multiple comparisons and modeling techniques in dose-response studies. *Biometrics*, 61:738–748.
- Bretz, F. Hsu, J., Pinheiro, J., and Liu, Y. (2008). Dose finding - a challenge in statistics. *Biometrical Journal*, 50:480–504.
- Chaloner, K. and Verdinelli, I. (1995). Bayesian experimental design: a review. *Statistical Science*, 10(3):273–304.
- Chaudhuri, P. and Mykland, P. A. (1995). On efficient designing of nonlinear experiments. *Statistica Sinica*, 5(2):421–440.
- Chernoff, H. (1953). Locally optimal designs for estimating parameters. *Annals of Mathematical Statistics*, 24:586–602.
- Cook, R. D. and Wong, W. K. (1994). On the equivalence of constrained and compound optimal designs. *Journal of the American Statistical Association*, 89:687–692.
- Dette, H. (1997). Designing experiments with respect to "standardized" optimality criteria. *Journal of the Royal Statistical Society, Ser. B*, 59:97–110.

- 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., Kiss, C., Bevanda, M., and Bretz, F. (2010). Optimal designs for the EMAX, log-linear and exponential models. *Biometrika*, 97(2):513–518.
- Dette, H. and O’Brien, T. (1999). Optimality criteria for regression models based on predicted variance. *Biometrika*, 86:93–106.
- Dragalin, V., Bornkamp, B., Bretz, F., Miller, F., Padmanabhan, S. K., Patel, N., Perevozskaya, I., Pinheiro, J., and Smith, J. R. (2010). A simulation study to compare new adaptive dose-ranging designs. *Statistics in Biopharmaceutical Research*, in press.
- Dragalin, V., Hsuan, F., and Padmanabhan, S. K. (2007). Adaptive designs for dose-finding studies based on sigmoid  $e_{\max}$  model. *Journal of Biopharmaceutical Statistics*, 17(6):1051–1070.
- Elfving, G. (1952). Optimal allocation in linear regression theory. *Annals of Statistics*, 23:255–262.
- EMA (2005). Cpmp guideline on the choice of the non-inferiority margin. doc. ref. [emea/cpmp/ewp/2158/99](http://www.ema.europa.eu/ema). Technical report, Available at <http://www.ema.europa.eu/ema>.
- EMA (2006). Cpmp guideline on clinical investigation of medicinal products for the treatment of multiple sclerosis. doc. ref. [cpmp/ewp/561/98](http://www.ema.europa.eu/ema). Technical report, Available at <http://www.ema.europa.eu/ema>.
- EMA (2011). Chmp guideline on medicinal products for the treatment of insomnia. doc. ref. [ema/chmp/16274/2009](http://www.ema.europa.eu/ema). Technical report, Available at <http://www.ema.europa.eu/ema>.
- Fang, X. and Hedayat, A. S. (2008). Locally D-optimal designs based on a class of composed models resulted from blending Emax and one-compartment models. *Annals of Statistics*, 36:428–444.
- Ford, I., Torsney, B., and Wu, C. F. J. (1992). The use of canonical form in the construction of locally optimum designs for nonlinear problems. *Journal of the Royal Statistical Society, Ser. B*, 54:569–583.
- He, Z., Studden, W. J., and Sun, D. (1996). Optimal designs for rational models. *Annals of Statistics*, 24:2128–2142.
- Herzberg, A. M. and Cox, D. R. (1972). Some optimal designs for interpolation and extrapolation. *Biometrika*, 59(3):551–561.

- ICH (1994). Ich topic e 4. dose response information to support drug registration. Technical report, Available at <http://www.ema.europa.eu/ema>.
- Imhof, L. A. (2001). Maximin designs for exponential growth models and heteroscedastic polynomial models. *Annals of Statistics*, 29:561–576.
- Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *Annals of Statistics*, 2:849–879.
- Kiefer, J. and Wolfowitz, J. (1964a). Optimum extrapolation and interpolation designs I. *Ann. Inst. Stat. Math.*, 16:79–108.
- Kiefer, J. and Wolfowitz, J. (1964b). Optimum extrapolation and interpolation designs II. *Ann. Inst. Stat. Math.*, 16:295–303.
- Miller, F., Guilbaud, O., and Dette, H. (2007). Optimal designs for estimating the interesting part of a dose-effect curve. *Journal of Biopharmaceutical Statistics*, 17:1097–1115.
- Pronzato, L. and Walter, E. (1985). Robust experimental design via stochastic approximation. *Math. Biosciences*, 75:103–120.
- Pukelsheim, F. (2006). *Optimal Design of Experiments*. SIAM, Philadelphia.
- Ruberg, S. J. (1995). Dose response studies I. Some design considerations. *J. Biopharm. Statist.*, 5:1–14.
- Ting, N. (2006). *Dose Finding in Drug Development*. Springer, New York.
- Yang, M. (2010). On the de la Garza Phenomenon. *Annals of Statistics*, to appear.