

# Bayesian optimal designs for dose-response curves with common parameters

Kirsten Schorning, Maria Konstantinou  
Fakultät für Mathematik  
Ruhr-Universität Bochum  
44780 Bochum  
Germany

## Abstract

The issue of determining not only an adequate dose but also a dosing frequency of a drug arises frequently in Phase II clinical trials. This results in the comparison of models which have some parameters in common. Planning such studies based on Bayesian optimal designs offers robustness to our conclusions since these designs, unlike locally optimal designs, are efficient even if the parameters are misspecified. In this paper we develop approximate design theory for Bayesian  $D$ -optimality for nonlinear regression models with common parameters and investigate the cases of common location or common location and scale parameters separately. Analytical characterisations of saturated Bayesian  $D$ -optimal designs are derived for frequently used dose-response models and the advantages of our results are illustrated via a numerical investigation.

## 1 Introduction

One of the key objectives in pharmaceutical drug development is to properly characterise the dose-response relationship as this has a direct impact on the determination of an efficacious and safe dose in Phase II clinical trials. As pointed out by Dette et al. (2008) and Bretz et al. (2010), too high of a dose is likely to result in toxicity and safety problems whereas choosing too low of a dose reduces the chance of establishing adequate efficacy and thus the chance of approval of the drug. In fact, in a recent review Sacks et al. (2014) identify the failure to select an optimal dose due to statistical uncertainty as one of the most frequent reasons for the delay or denial of approval of drugs by the Food and Drug Administration (FDA).

The efficient planning of dose-finding studies can significantly improve the statistical accuracy for the selection of an efficacious and safe dose. For this reason numerous authors

investigate the construction of optimal experimental designs for such studies and consider regression models which are frequently used to describe the dose-response relationship (see, for example, Dette et al. (2008), Dragalin et al. (2008) and Bretz et al. (2010)).

In this paper we consider the closely related issue of determining not only a safe and efficacious dose but also the dosing frequency. This is often of interest in Phase II dose-finding studies since the frequency of administration affects the pharmacological effect of the drug under investigation. That is, if the same drug is administered at different dosing intervals, for example, once or twice a day, the drug exposure inside the body during the day will be higher right after administration and lower right before administration in the case of once a day administration whereas if the drug is administered twice a day the drug exposure during the day will be more uniform.

Under this scenario, estimating the dose-response curves corresponding to the different administration frequencies separately is inefficient. This is due to the fact that the regression models will often have common parameters thus suggesting a borrowing of strength. One of the parameters that it is reasonable to assume is the same for all models is the parameter corresponding to the placebo effect. Another assumption that is made in some situations is that there exists a biological maximum attainable drug effect. In this case, the parameter describing the efficacy for high doses can also be considered to be common for all models. In what follows we will investigate separately the cases where only the parameter corresponding to the placebo effect or both of the aforementioned parameters are common in all the regression models.

The aim of the present paper is the efficient planning of such dose-finding studies where the regression models, used to model the dose-response relationship for the different administration frequencies, share some of the parameters. More specifically, we are interested in finding designs providing us with the doses to utilise in the different administration groups and also the way to split the total sample size between the groups, such that the design is optimal in terms of estimating all of the model parameters with high precision. Despite of the practical interest, the construction of optimal designs for dose-finding studies involving different administration frequencies has only been investigated by Feller et al. (2017) who develop the optimal design theory corresponding to this scenario and derive explicit expressions for locally  $D$ -optimal designs. However, locally  $D$ -optimal designs require a best guess of the parameter values in order to be used in practice which can result in inefficient designs if the parameters are misspecified. Therefore, a design strategy that is robust to parameter misspecification is of great importance in applications.

Throughout this paper we adopt the robust design strategy of Bayesian optimal designs introduced by Pronzato and Walter (1985), Chaloner (1989) and Chaloner and Larntz (1989). Given an uncertainty space for the parameter values, Bayesian optimality incorporates this uncertainty in the formulation of the optimality criteria via a prior distribution on the given parameter space and the resulting Bayesian optimality criteria are based on classical optimality criteria (see, for example, Chaloner (1993) and Chaloner and Verdinelli (1995) for more details). The construction of Bayesian optimal designs for several regression models has been studied, for example, by Dette et al. (2007), Dette

et al. (2008) and Chaloner and Larntz (1992). However, to the best of the authors knowledge, Bayesian optimal designs for dose-finding studies involving several administration frequency groups have not been considered in the literature.

In the present paper we consider the Bayesian  $D$ -optimal design problem for comparing regression models with common parameters. Since these designs are based on the results for the individual models, in Section 2 we derive analytical characterisations of Bayesian  $D$ -optimal designs for the individual regression models which, to the best of our knowledge, do not exist in the literature so far. In Section 3 we introduce the regression models with common parameters and in Section 3.1 we derive some general results concerning the Bayesian  $D$ -optimal designs. The scenarios where the regression models have only the location parameter or both the location and scale parameter in common are considered in Sections 3.2 and 3.3, respectively. In both cases, we provide explicit characterisations of the saturated Bayesian  $D$ -optimal designs for regression models that are frequently used in dose-response studies. Finally, in Section 4 we assess the performance of the saturated designs constructed in Section 3 in a situation where they are not globally optimal.

## 2 Bayesian $D$ -optimal designs for individual regression models

It turns out that the locally optimal designs for models with common parameters are based on the locally optimal designs for the individual models. Therefore, we begin our investigation by considering the construction of Bayesian optimal designs for individual regression models. Besides some results for specific regression models, the construction of such designs in a general framework has not been considered in the literature so far.

### 2.1 Models and optimality criterion

We consider regression models of the form

$$Y_{jl} = f(d_j, \tilde{\theta}) + \varepsilon_{jl}, \quad j = 1, \dots, \tilde{k}; l = 1, \dots, \tilde{n}_j, \quad (2.1)$$

where  $\varepsilon_{jl}$  are independent centred normally distributed, that is,  $\varepsilon_{jl} \stackrel{iid}{\sim} N(0, \sigma^2)$ ,  $\tilde{\theta} = (\theta_1, \dots, \theta_{\tilde{m}})^T \in \tilde{\Theta} \subset \mathbb{R}^{\tilde{m}}$  is the vector of model parameters and  $\tilde{n}_j$  is the total number of observations taken at each dose level  $d_j$  ( $j = 1, \dots, \tilde{k}$ ) taking values in the design space  $\mathcal{X} = [0, d_{max}]$ . We collect the information provided by the different dose levels in a probability measure, say  $\tilde{\xi}$ , on the design space  $\mathcal{X}$  defined as

$$\tilde{\xi} = \left\{ \begin{array}{ccc} d_1 & \dots & d_{\tilde{k}} \\ \omega_1 & \dots & \omega_{\tilde{k}} \end{array} \right\}, \quad 0 < \omega_j < 1, \quad \sum_{j=1}^{\tilde{k}} \omega_j = 1,$$

(see Kiefer (1974)). The measure  $\tilde{\xi}$  is called design, the  $d_j$ 's are called the support points of the design and the corresponding  $\omega_j$ 's are the weights of the design and represent the proportion of observations to be taken at each experimental condition  $d_j$ . We denote the support of the design  $\tilde{\xi}$  by  $\text{supp}(\tilde{\xi})$ . Let  $\tilde{n} = \sum_{j=1}^{\tilde{k}} \tilde{n}_j$  be the total sample size. Under the assumption

$$\lim_{n \rightarrow \infty} \frac{\tilde{n}_j}{\tilde{n}} = \omega_j > 0, \quad j = 1, \dots, \tilde{k},$$

the maximum likelihood estimator  $\hat{\boldsymbol{\theta}}$  of the parameter vector  $\boldsymbol{\theta}$  satisfies

$$\sqrt{\tilde{n}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{\mathcal{D}} N_{\tilde{m}}(\mathbf{0}, M_{\text{ind}}^{-1}(\tilde{\xi}, \boldsymbol{\theta})),$$

where

$$M_{\text{ind}}^{-1}(\tilde{\xi}, \boldsymbol{\theta}) = \int_{\mathcal{X}} \tilde{h}(d, \boldsymbol{\theta}) \tilde{h}^T(d, \boldsymbol{\theta}) d\tilde{\xi}(d), \quad (2.2)$$

and

$$\tilde{h}(d, \boldsymbol{\theta}) = \frac{1}{\sigma} \left( \frac{\partial}{\partial \boldsymbol{\theta}} f(d, \boldsymbol{\theta}) \right)^T. \quad (2.3)$$

Our goal is to find optimal designs of this form such that the parameter vector  $\boldsymbol{\theta}$  is estimated with high precision even when there is imperfect knowledge of the true parameter values.

As discussed in the introduction, throughout this paper we adopt the Bayesian approach as a robust design strategy and thus we address the issue of parameter dependence of locally optimal designs. Since interest is in estimating all of the model parameters, the Bayesian criterion is based on the classical  $D$ -optimality criterion (see Chaloner and Verdinelli (1995) for more details). In particular, if  $\tilde{\pi}$  is a prior distribution on the parameter space  $\Theta$ , a design  $\tilde{\xi}_{\tilde{\pi}}^*$  is called Bayesian  $D$ -optimal with respect to the prior  $\tilde{\pi}$  for the individual model (2.1) if it maximises the function

$$\Phi_{\tilde{\pi}}(\tilde{\xi}) = \int_{\Theta} \log \det \{ M_{\text{ind}}(\tilde{\xi}, \boldsymbol{\theta}) \} \tilde{\pi}(d\boldsymbol{\theta}). \quad (2.4)$$

## 2.2 Bayesian $D$ -optimal designs

A very important tool for characterising and checking the optimality of a candidate design is the general equivalence theorem. Using Theorem 3.3 in Dette et al. (2007) the general equivalence theorem for Bayesian  $D$ -optimality of models of the form (2.1) is given below.

**Theorem 2.1.** *A design  $\tilde{\xi}_{\tilde{\pi}}^*$  is Bayesian  $D$ -optimal with respect to the prior  $\tilde{\pi}$  for maximum likelihood estimation in model (2.1) if and only if the inequality*

$$\int_{\Theta} g(d, \tilde{\xi}_{\tilde{\pi}}^*, \boldsymbol{\theta}) \tilde{\pi}(d\boldsymbol{\theta}) := \int_{\Theta} h^T(d, \boldsymbol{\theta}) M_{\text{ind}}^{-1}(\tilde{\xi}_{\tilde{\pi}}^*, \boldsymbol{\theta}) h(d, \boldsymbol{\theta}) \tilde{\pi}(d\boldsymbol{\theta}) \leq \tilde{m},$$

holds for all  $d \in \mathcal{X} = [0, d_{\text{max}}]$ . Equality holds for any point  $d \in \text{supp}(\tilde{\xi}_{\tilde{\pi}}^*)$ .

An approximate design for maximum likelihood estimation in model (2.1) must have at least  $\tilde{m}$  support points in order to be estimable for all the model parameters. A well known result in the optimal design theory is that of equally weighted  $D$ -optimal saturated designs, that is, of designs with as many support points as model parameters. Lemma 2.1 shows that this result holds also in the case of Bayesian  $D$ -optimal saturated designs for maximum likelihood estimation in model (2.1) and it is proven in the appendix.

**Lemma 2.1.** *If  $\tilde{\xi}_{\tilde{\pi}}^*$  is a Bayesian  $D$ -optimal saturated design with respect to the prior  $\tilde{\pi}$  for maximum likelihood estimation in model (2.1), then it has equal weights  $\frac{1}{\tilde{m}}$  at its support points.*

For the rest of this section we will focus on three well-known nonlinear models which are widely used for the modelling of the dose-response relationship. Namely, we consider the Emax, the exponential and the linear-in-log models specified by the regression functions

$$f_1(d, \tilde{\boldsymbol{\theta}}) = \theta_0 + \frac{\theta_1 d}{\theta_2 + d}, \quad d \in [0, d_{\max}], \quad (2.5)$$

$$f_2(d, \tilde{\boldsymbol{\theta}}) = \theta_0 + \theta_1 \left( \exp\left(\frac{d}{\theta_2}\right) - 1 \right), \quad d \in [0, d_{\max}], \quad (2.6)$$

$$f_3(d, \tilde{\boldsymbol{\theta}}) = \theta_0 + \theta_1 \log\left(\frac{d}{\theta_2} + 1\right), \quad d \in [0, d_{\max}], \quad (2.7)$$

respectively. The locally  $D$ -optimal designs for reparameterisations of these models are found in Dette et al. (2010) and are proven to be saturated, that is, they have exactly three support points, with the end-points of the design space always being two of the support points. Therefore, for each of the nonlinear models given above, we consider the construction of Bayesian  $D$ -optimal saturated designs for which we provide analytical characterisations presented in the following theorem. Its proof can be found in the appendix.

**Theorem 2.2.** *A Bayesian  $D$ -optimal saturated design with respect to the prior  $\tilde{\pi} = \tilde{\pi}_0 \times \tilde{\pi}_1 \times \tilde{\pi}_2$  for maximum likelihood estimation in the Emax, exponential or the linear-in-log model is equally supported at points  $0, \tilde{d}^*, d_{\max}$  where the non-trivial support point  $\tilde{d}^*$  is a solution in the interval  $(0, d_{\max})$  of the equation*

$$\int_{\tilde{\Theta}_2} \left( \frac{1}{d} - \frac{1}{d_{\max} - d} - \frac{2}{\theta_2 + d} \right) \tilde{\pi}_2(d\theta_2) = 0, \quad (2.8)$$

for the Emax model, of the equation

$$\int_{\tilde{\Theta}_2} \frac{\exp(d/\theta_2) (d_{\max} \exp(d_{\max}/\theta_2) - (d + \theta_2)(\exp(d_{\max}/\theta_2) - 1))}{d_{\max} \exp(d_{\max}/\theta_2)(\exp(d/\theta_2) - 1) - d \exp(d/\theta_2)(\exp(d_{\max}/\theta_2) - 1)} \tilde{\pi}_2(d\theta_2) = 0 \quad (2.9)$$

for the exponential model and of the equation

$$\int_{\tilde{\Theta}_2} \frac{1}{d + \theta_2} \frac{\theta_2 \log(d_{\max}/\theta_2 + 1)(d_{\max} + \theta_2) - d_{\max}(d + \theta_2)}{d(d_{\max} + \theta_2) \log(d_{\max}/\theta_2 + 1) - d_{\max}(d + \theta_2) \log(d/\theta_2 + 1)} \tilde{\pi}_2(d\theta_2) = 0, \quad (2.10)$$

for the linear-in-log model.

Note that these analytical characterisations reduce the design problem to solving an equation in one variable and therefore, the numerical effort for design search is reduced substantially. Furthermore, the Bayesian  $D$ -optimal designs require only the marginal prior distribution  $\tilde{\pi}_2$  on  $\tilde{\Theta}_2$ , that is, they require prior information only for the parameter  $\theta_2$  and not for the full parameter vector, because the considered models are linear in the first two parameters  $\theta_0$  and  $\theta_1$ .

### 3 Bayesian $D$ -optimal designs for regression models with common parameters

In this section we focus on our core investigation regarding the case of  $M$  different administration frequency groups modelled using the same parametric form and as it is discussed in the introduction, these models could share some parameters while others are different. More precisely, we consider regression models of the form

$$Y_{ijl} = f(d_j^{(i)}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2^{(i)}) + \varepsilon_{ijl}, \quad i = 1, \dots, M; j = 1, \dots, k_i; l = 1, \dots, n_{ij}, \quad (3.1)$$

where  $\varepsilon_{ijl} \stackrel{iid}{\sim} N(0, \sigma_i^2)$  and in each administration frequency group  $i = 1, \dots, M$  a total of  $n_{ij}$  observations are taken at each dose level  $d_1^{(i)}, \dots, d_{k_i}^{(i)}$ . We further assume that the dose levels  $d_j^{(i)}$ ,  $i = 1, \dots, M; j = 1, \dots, k_i$ , vary in possibly different design spaces for each group, that is,  $\mathcal{X}_i = [0, d_{max}^{(i)}]$ ,  $i = 1, \dots, M$ . The vector  $\boldsymbol{\theta}_1 = (\theta_1, \dots, \theta_p)^T \in \Theta_1 \subset \mathbb{R}^p$  corresponds to the parameters that are the same in all groups whereas  $\boldsymbol{\theta}_2^{(i)} = (\theta_1^{(i)}, \dots, \theta_q^{(i)})^T \in \Theta_2^{(i)} \subset \mathbb{R}^q$ ,  $i = 1, \dots, M$ , is the vector of parameters that are different for different groups.

#### 3.1 General results

Planning a dose-finding study with different administration frequency groups is concerned with identifying the doses to be utilised in each group and also identifying how to split the total sample size between the groups. We formulate this using approximate designs in the sense of Kiefer (1974). In particular, for each group  $i = 1, \dots, M$ , we define the approximate design

$$\xi_i = \left\{ \begin{array}{ccc} d_1^{(i)} & \dots & d_{k_i}^{(i)} \\ \omega_{i1} & \dots & \omega_{ik_i} \end{array} \right\}, \quad 0 < \omega_{ij} < 1, \quad \sum_{j=1}^{k_i} \omega_{ij} = 1,$$

as a probability measure on the design space  $\mathcal{X}_i$  assigning a proportion  $\omega_{ij}$  of the observations in the  $i$ th group at each of the dose levels  $d_j^{(i)}$  ( $i = 1, \dots, M; j = 1, \dots, k_i$ ). Furthermore, we define the probability measure

$$\mu = \left\{ \begin{array}{ccc} 1 & \dots & M \\ \lambda_1 & \dots & \lambda_M \end{array} \right\}, \quad 0 < \lambda_i < 1, \quad \sum_{i=1}^M \lambda_i = 1,$$

on the set  $\{1, \dots, M\}$  assigning a proportion  $\lambda_i$  of the total sample size to the  $i$ th group. The collection of these designs in the vector  $\xi = (\xi_1, \dots, \xi_M, \mu)$  is a probability measure on the design space  $\mathcal{X}_1 \times \dots \times \mathcal{X}_M \times \{1, \dots, M\}$  and is also called a design throughout this paper.

As discussed in the introduction, estimating the dose-response curves corresponding to each group separately, using the results of section 2, is wasteful because of the borrowing of strength due to the common parameters in the models. We therefore consider the simultaneous maximum likelihood estimation of the full parameter vector  $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2^{(1)}, \dots, \boldsymbol{\theta}_2^{(M)})^T \in \Theta_1 \times \Theta_2^{(1)} \times \dots \times \Theta_2^{(M)} = \Theta \subset \mathbb{R}^m$  where  $m = p + qM$ .

Let  $n_i = \sum_{j=1}^{k_i} n_{ij}$  be the sample size in the  $i$ th group and denote by  $n = \sum_{i=1}^M n_i$  the total sample size. From Feller et al. (2017) we have that under common regularity assumptions and the assumptions

$$\lim_{n_i \rightarrow \infty} \frac{n_{ij}}{n_i} = \omega_{ij} \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{n_i}{n} = \lambda_i \quad (i = 1, \dots, M; j = 1, \dots, k_i),$$

the maximum likelihood estimator  $\hat{\boldsymbol{\theta}}$  of the full parameter vector  $\boldsymbol{\theta}$  satisfies

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{\mathcal{D}} N_m(\mathbf{0}, M^{-1}(\xi, \boldsymbol{\theta})),$$

where

$$M(\xi, \boldsymbol{\theta}) = \int \int_{\mathcal{X}_z} h_z(d, \boldsymbol{\theta}) h_z^T(d, \boldsymbol{\theta}) d\xi_z(d) d\mu(z) = \sum_{i=1}^M \lambda_i M^{(i)}(\xi_i, \boldsymbol{\theta}), \quad (3.2)$$

is the information matrix of the design  $\xi = (\xi_1, \dots, \xi_M, \mu)$ , the matrices  $M^{(i)}$  are defined as

$$M^{(i)}(\xi_i, \boldsymbol{\theta}) = \int_{\mathcal{X}_i} h_i(d, \boldsymbol{\theta}) h_i^T(d, \boldsymbol{\theta}) d\xi_i(d), \quad i = 1, \dots, M, \quad (3.3)$$

and

$$h_i(d, \boldsymbol{\theta}) = \frac{1}{\sigma_i} \left( \left( \frac{\partial}{\partial \boldsymbol{\theta}_1} f(d, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2^{(i)}) \right)^T, \mathbf{0}_{q(i-1)}^T, \left( \frac{\partial}{\partial \boldsymbol{\theta}_2^{(i)}} f(d, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2^{(i)}) \right)^T, \mathbf{0}_{q(M-i)}^T \right)^T \in \mathbb{R}^m, \quad (3.4)$$

is the gradient of the regression function  $f(d, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2^{(i)})$  with respect to the parameter vector  $\boldsymbol{\theta} \in \mathbb{R}^m$ , with  $\mathbf{0}_{q(i-1)} \in \mathbb{R}^{q(i-1)}$  and  $\mathbf{0}_{q(M-i)} \in \mathbb{R}^{q(M-i)}$  denoting vectors with all entries equal to zero.

Let  $\pi$  be a prior distribution on the parameter space  $\Theta = \Theta_1 \times \Theta_2^{(1)} \times \dots \times \Theta_2^{(M)}$ . A design  $\xi_\pi^*$  is Bayesian  $D$ -optimal with respect to the prior  $\pi$  for maximum likelihood estimation in models (3.1) if it maximises

$$\Phi_\pi(\xi) = \int_{\Theta} \log \det\{M(\xi, \boldsymbol{\theta})\} \pi(d\boldsymbol{\theta}). \quad (3.5)$$

The following theorem provides the general equivalence theorem for characterising and checking the Bayesian  $D$ -optimality of a candidate design in the multiple administration frequency groups scenario we consider. Its proof is given in the appendix.

**Theorem 3.1.** A design  $\xi_\pi^* = (\xi_1^*, \dots, \xi_M^*, \mu^*)$  is Bayesian  $D$ -optimal with respect to the prior  $\pi$  for maximum likelihood estimation in models (3.1), if and only if the  $M$  inequalities

$$\tau_i(d) = \int_{\Theta} g_i(d, \xi_\pi^*, \boldsymbol{\theta}) \pi(d\boldsymbol{\theta}) - m := \int_{\Theta} h_i^T(d, \boldsymbol{\theta}) M^{-1}(\xi_\pi^*, \boldsymbol{\theta}) h_i(d, \boldsymbol{\theta}) \pi(d\boldsymbol{\theta}) - (p + qM) \leq 0, \quad (3.6)$$

are satisfied for all  $d \in \mathcal{X}_i$ ,  $i = 1, \dots, M$ . Equality holds for any points  $(d_1, \dots, d_M, z) \in \text{supp}(\xi_1^*) \times \dots \times \text{supp}(\xi_M^*) \times \text{supp}(\mu^*)$ .

Now let

$$\Xi_m^M = \left\{ \xi_\pi = (\xi_1, \dots, \xi_M, \mu) \mid \sum_{i=1}^M |\text{supp}(\xi_i)| = m \right\}, \quad (3.7)$$

be the set of all designs on  $\mathcal{X}_1 \times \dots \times \mathcal{X}_M \times \{1, \dots, M\}$ , with a total of exactly  $m$  different dose levels in the  $M$  groups. Therefore, this is the class of saturated designs on  $\mathcal{X}_1 \times \dots \times \mathcal{X}_M \times \{1, \dots, M\}$ . Lemma 3.1 provides us with the weights of Bayesian  $D$ -optimal designs in the class  $\Xi_m^M$  and is proven in the appendix.

**Lemma 3.1.** Let  $\xi_\pi = (\xi_1, \dots, \xi_M, \mu) \in \Xi_m^M$  denote a design on  $\mathcal{X}_1 \times \dots \times \mathcal{X}_M \times \{1, \dots, M\}$  and  $m_i$  denote the number of support points of  $\xi_i$  ( $i = 1, \dots, M$ ). Also assume that the  $m = \sum_{i=1}^M m_i$  vectors  $h_1(d_1^{(1)}, \boldsymbol{\theta}), \dots, h_1(d_{m_1}^{(1)}, \boldsymbol{\theta}), \dots, h_M(d_1^{(M)}, \boldsymbol{\theta}), \dots, h_M(d_{m_M}^{(M)}, \boldsymbol{\theta})$  are linearly independent for all parameters  $\boldsymbol{\theta} \in \Theta$ , where  $d_j^{(i)} \in \text{supp}(\xi_i)$ ,  $j = 1, \dots, m_i$ ;  $i = 1, \dots, M$ .

If  $\xi_\pi^* = (\xi_1^*, \dots, \xi_M^*, \mu^*)$  is Bayesian  $D$ -optimal with respect to the prior  $\pi$  in the class  $\Xi_m^M$ , then each  $\xi_i^*$  has equal weights at its support points and the weights of  $\mu^*$  at its support points  $1, \dots, M$  are given by  $\frac{m_1}{m}, \dots, \frac{m_M}{m}$  respectively.

In what follows we focus on the construction of designs which are Bayesian  $D$ -optimal in the class  $\Xi_m^M$ , that is, of designs which are saturated. This is because these designs have the smallest possible number of support points and thus, the number of experimental conditions where observations should be taken is minimum. Therefore, saturated designs result in lower costs for the study and are thus often preferred to designs which are optimal amongst all possible designs if the latter have more support points (see Section 4 for more details).

Moreover, in the following two sections we investigate the cases of the models (3.1) having the same location parameter and having the same location and scale parameters separately. Table 3.1 presents these two cases with respect to the forms of the regression functions for the Emax, exponential and linear-in-log models.

## 3.2 Bayesian $D$ -optimal designs for models with the same location parameter

We first consider the case where the regression models corresponding to the  $M$  administration frequency groups share only the location parameter and thus, the regression



Table 1: *E*max, exponential and linear-in-log models for  $i = 1, \dots, M$  groups. Left column: Common location parameter. Right column: Common location and scale parameters.

	location	location and scale
Emax	$\theta_1 + \theta_1^{(i)} \frac{d}{\theta_2^{(i)} + d}$	$\theta_1 + \theta_2 \frac{d}{\theta_2^{(i)} + d}$
Exponential	$\theta_1 + \theta_1^{(i)} \left( \exp\left(\frac{d}{\theta_2^{(i)}}\right) - 1 \right)$	$\theta_1 + \theta_2 \left( \exp\left(\frac{d}{\theta_2^{(i)}}\right) - 1 \right)$
Linear-in-log	$\theta_1 + \theta_1^{(i)} \log\left(\frac{d}{\theta_2^{(i)}} + 1\right)$	$\theta_1 + \theta_2 \log\left(\frac{d}{\theta_2^{(i)}} + 1\right)$

function in (3.1) is of the form

$$f(d, \theta_1, \boldsymbol{\theta}_2^{(i)}) = \theta_1 + f_0(d, \boldsymbol{\theta}_2^{(i)}), \quad i = 1, \dots, M. \quad (3.8)$$

This corresponds, for example, to the case of a common placebo effect for all groups. Here the full parameter vector is  $\boldsymbol{\theta} = (\theta_1, \boldsymbol{\theta}_2^{(1)}, \dots, \boldsymbol{\theta}_2^{(M)})^T \in \Theta \subset \mathbb{R}^m$  where  $m = 1 + qM$ .

Denote by  $\tilde{\pi}_2^{(i)}$  the marginal prior distribution on  $\Theta_2^{(i)}$  ( $i = 1, \dots, M$ ). The following theorem provides a solution of the Bayesian  $D$ -optimal design problem in the class  $\Xi_m^M$  for models (3.1) with regression functions (3.8) if the Bayesian  $D$ -optimal saturated designs for the individual models are known. Its proof can be found in the appendix.

**Theorem 3.2.** *Let  $\sigma_1^2 = \min_{i=1, \dots, M} \sigma_i^2$  and consider models (3.1) with regression functions of the form (3.8) also satisfying*

$$f_0(0, \boldsymbol{\theta}_2^{(i)}) = 0, \quad \eta_0(0, \boldsymbol{\theta}_2^{(i)}) = \frac{\partial}{\partial \boldsymbol{\theta}_2^{(i)}} f_0(d, \boldsymbol{\theta}_2^{(i)})|_{d=0} = \mathbf{0}_q, \quad (i = 1, \dots, M). \quad (3.9)$$

If the saturated design

$$\tilde{\xi}_{\tilde{\pi}_2^{(i)}}^* = \left\{ \begin{array}{cccc} 0 & d_1^{(i)} & \dots & d_q^{(i)} \\ \frac{1}{q+1} & \frac{1}{q+1} & \dots & \frac{1}{q+1} \end{array} \right\}, \quad (3.10)$$

is the Bayesian  $D$ -optimal saturated design with respect to the prior  $\tilde{\pi}_2^{(i)}$  on  $\Theta_2^{(i)}$  for maximum likelihood estimation in the individual model with variance  $\sigma_i^2$  ( $i = 1, \dots, M$ ), then the Bayesian  $D$ -optimal design in the class  $\Xi_5^2$  with respect to the prior  $\pi = \tilde{\pi}_1 \times \tilde{\pi}_2^{(1)} \times \dots \times \tilde{\pi}_2^{(M)}$  for maximum likelihood estimation in models (3.1) with regression function (3.8) is given by  $\xi_\pi^* = (\xi_1^*, \dots, \xi_M^*, \mu^*)$ , where

$$\xi_1^* = \tilde{\xi}_{\tilde{\pi}_2^{(1)}}^*; \quad \xi_i^* = \left\{ \begin{array}{ccc} d_1^{(i)} & \dots & d_q^{(i)} \\ \frac{1}{q} & \dots & \frac{1}{q} \end{array} \right\}, \quad i = 2, \dots, M; \quad \mu^* = \left\{ \begin{array}{cccc} 1 & 2 & \dots & M \\ \frac{q+1}{m} & \frac{q}{m} & \dots & \frac{q}{m} \end{array} \right\}. \quad (3.11)$$

Therefore, the common location parameter  $\theta_1$  is estimated in the group with minimal variance. Intuitively this result is reasonable since the estimator of the common location

parameter  $\theta_1$  is roughly the mean of the observations at dose level  $d = 0$  and the variance of this estimator is proportional to the variance in the considered group.

The corresponding Bayesian  $D$ -optimal saturated designs for the cases of Emax, exponential and linear-in-log models are found explicitly and are presented in Corollary 3.1. This result follows by a direct application of Theorem 3.2 using the corresponding characterisations presented in Theorem 2.2 for Bayesian  $D$ -optimal saturated designs in the individual Emax, exponential and linear-in-log models and thus, its proof is omitted.

**Corollary 3.1.** *Let  $\sigma_1^2 = \min_{i=1,\dots,M} \sigma_i^2$ . The Bayesian  $D$ -optimal design in the class (3.7) with respect to the prior  $\pi = \tilde{\pi}_1 \times \tilde{\pi}_2^{(1)} \times \dots \times \tilde{\pi}_2^{(M)}$  for maximum likelihood estimation in Emax, exponential and linear-in-log models (3.1) with regression functions (3.8) is  $\xi_\pi^* = (\xi_1^*, \dots, \xi_M^*, \mu^*)$  where*

$$\xi_1^* = \left\{ \begin{array}{ccc} 0 & d^{*(1)} & d_{max}^{(1)} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{array} \right\}; \quad \xi_i^* = \left\{ \begin{array}{cc} d^{*(i)} & d_{max}^{(i)} \\ \frac{1}{2} & \frac{1}{2} \end{array} \right\}, \quad i = 2, \dots, M; \quad \mu^* = \left\{ \begin{array}{cccc} 1 & 2 & \dots & M \\ \frac{3}{m} & \frac{2}{m} & \dots & \frac{2}{m} \end{array} \right\},$$

and the point  $\tilde{d}^{*(i)}$  ( $i = 1, \dots, M$ ) is a solution of the equation (2.8), (2.9) and (2.10) for the cases of Emax, exponential and linear-in-log individual models respectively using the corresponding marginal prior  $\tilde{\pi}_2^{(i)}$ .

### 3.3 Bayesian $D$ -optimal designs for models with the same location and scale parameters

In this section we consider models (3.1) with regression functions of the form

$$f(d, \theta_1, \theta_2, \boldsymbol{\theta}_2^{(i)}) = \theta_1 + \theta_2 f_0(d, \boldsymbol{\theta}_2^{(i)}), \quad i = 1, \dots, M, \quad (3.12)$$

that is, the models have common location and scale parameters. One such situation is, for example, when both the placebo effect and the efficacy for high doses coincide across the different models. As noted by Feller et al. (2017), the optimal design problem in this case is significantly harder. Therefore, in what follows we will present results for the case  $M = 2$  administration frequency groups and thus the full parameter vector is  $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_2^{(1)}, \theta_2^{(2)})^T \in \Theta \subset \mathbb{R}^4$ . Similar results can be obtained for  $M > 2$  with an additional amount of notation.

For the rest of this section we will focus on Emax, exponential and linear-in-log models for which we derive analytical characterisations of their corresponding Bayesian  $D$ -optimal designs. We also note that we construct Bayesian  $D$ -optimal designs in the class  $\Xi_4^2$ , defined in (3.7), of saturated designs on  $\mathcal{X}_1 \times \mathcal{X}_2 \times \{1, 2\}$ . We begin with the explicit result for two Emax models with the same location and scale parameters. The proof can be found in the appendix.

**Theorem 3.3.** Let  $r = \frac{\sigma_1^2}{\sigma_2^2}$  and  $\tilde{d}^{*,(i)}, d^{*,(i)}$  be solutions of the equation (2.8) in the interval  $(0, d_{max}^{(i)})$  and of the equation

$$\int_{\Theta_2^{(i)}} \frac{1}{d} - \frac{2}{d + \theta_2^{(i)}} \tilde{\pi}_2^{(i)}(d\theta_2^{(i)}) = 0,$$

in the interval  $[0, d_{max}^{(i)}]$  respectively using in both equations the marginal prior distribution  $\tilde{\pi}_2^{(i)}$  on  $\Theta_2^{(i)}$  for  $i = 1, 2$ . Also define the function

$$\begin{aligned} u(\tilde{d}^{*,(1)}, d^{*,(1)}, \tilde{d}^{*,(2)}, d^{*,(2)}) &:= 2 \log \left[ \frac{\tilde{d}^{*,(1)} d_{max}^{(1)} (d_{max}^{(1)} - \tilde{d}^{*,(1)}) d^{*,(2)}}{\tilde{d}^{*,(2)} d_{max}^{(2)} (d_{max}^{(2)} - \tilde{d}^{*,(2)}) d^{*,(1)}} \right] \\ &+ 4 \sum_{i=1}^2 (-1)^{i+1} \int_{\Theta_2^{(i)}} \log \left[ \frac{(d^{*,(i)} + \theta_2^{(i)})}{(\tilde{d}^{*,(i)} + \theta_2^{(i)})(d_{max}^{(i)} + \theta_2^{(i)})} \right] \tilde{\pi}_2^{(i)}(d\theta_2^{(i)}). \end{aligned}$$

The Bayesian  $D$ -optimal design in the class  $\Xi_4^2$  with respect to the prior  $\pi = \tilde{\pi}_1 \times \tilde{\pi}_2 \times \tilde{\pi}_2^{(1)} \times \tilde{\pi}_2^{(2)}$  for maximum likelihood estimation in two Emax models with regression functions of the form (3.12) is

(1)  $\xi^{a,*} = (\xi_1^{a,*}, \xi_2^{a,*}, \mu^{a,*})$  if  $\log r \leq 0$  and  $\log r \leq u(\tilde{d}^{*,(1)}, d^{*,(1)}, \tilde{d}^{*,(2)}, d^{*,(2)})$ , where

$$\xi_1^{a,*} = \left\{ \begin{array}{cc} 0 & \tilde{d}^{*,(1)} \\ \frac{1}{3} & \frac{1}{3} \end{array} \right\}, \quad \xi_2^{a,*} = \left\{ \begin{array}{c} d^{*,(2)} \\ 1 \end{array} \right\}, \quad \mu^{a,*} = \left\{ \begin{array}{cc} 1 & 2 \\ \frac{3}{4} & \frac{1}{4} \end{array} \right\},$$

(2)  $\xi^{b_1,*} = (\xi_1^{b_1,*}, \xi_2^{b_1,*}, \mu^{b_1,*})$  if  $u(\tilde{d}^{*,(1)}, d^{*,(1)}, \tilde{d}^{*,(2)}, d^{*,(2)}) < \log r \leq 0$ , where

$$\xi_1^{b_1,*} = \left\{ \begin{array}{cc} 0 & d^{*,(1)} \\ \frac{1}{2} & \frac{1}{2} \end{array} \right\}, \quad \xi_2^{b_1,*} = \left\{ \begin{array}{cc} \tilde{d}^{*,(2)} & d_{max}^{(2)} \\ \frac{1}{2} & \frac{1}{2} \end{array} \right\}, \quad \mu^{b_1,*} = \left\{ \begin{array}{cc} 1 & 2 \\ \frac{1}{2} & \frac{1}{2} \end{array} \right\},$$

(3)  $\xi^{b_2,*} = (\xi_1^{b_2,*}, \xi_2^{b_2,*}, \mu^{b_2,*})$  if  $0 \leq \log r \leq u(\tilde{d}^{*,(1)}, d^{*,(1)}, \tilde{d}^{*,(2)}, d^{*,(2)})$ , where

$$\xi_1^{b_2,*} = \left\{ \begin{array}{cc} \tilde{d}^{*,(1)} & d_{max}^{(1)} \\ \frac{1}{2} & \frac{1}{2} \end{array} \right\}, \quad \xi_2^{b_2,*} = \left\{ \begin{array}{cc} 0 & d^{*,(2)} \\ \frac{1}{2} & \frac{1}{2} \end{array} \right\}, \quad \mu^{b_2,*} = \left\{ \begin{array}{cc} 1 & 2 \\ \frac{1}{2} & \frac{1}{2} \end{array} \right\},$$

(4)  $\xi^{c,*} = (\xi_1^{c,*}, \xi_2^{c,*}, \mu^{c,*})$  if  $\log r \geq 0$  and  $\log(r) > u(\tilde{d}^{*,(1)}, d^{*,(1)}, \tilde{d}^{*,(2)}, d^{*,(2)})$ , where

$$\xi_1^{c,*} = \left\{ \begin{array}{c} d^{*,(1)} \\ 1 \end{array} \right\}, \quad \xi_2^{c,*} = \left\{ \begin{array}{ccc} 0 & \tilde{d}^{*,(2)} & d_{max}^{(2)} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{array} \right\}, \quad \mu^{c,*} = \left\{ \begin{array}{cc} 1 & 2 \\ \frac{1}{4} & \frac{3}{4} \end{array} \right\}.$$

The analogous results for the cases of two exponential and two linear-in-log models with the same location and scale parameters are presented in the following theorems. Their proofs follow along the same lines as the proof of Theorem 3.3 and are therefore omitted.

**Theorem 3.4.** Let  $r = \frac{\sigma_1^2}{\sigma_2^2}$  and  $\tilde{d}^{*,(i)}$  be a solution in the interval  $(0, d_{max}^{(i)})$  of the equation (2.9) using the marginal prior distribution  $\tilde{\pi}_2^{(i)}$  on  $\Theta_2^{(i)}$ . Also define the function

$$u(\tilde{d}^{*,(1)}, \tilde{d}^{*,(2)}) := 2 \sum_{i=1}^2 (-1)^{i+1} \int_{\Theta_2^{(i)}} 2 \log \left[ \left(1 - \frac{\tilde{d}^{*,(i)}}{d_{max}^{(i)}}\right) \exp\left(\frac{\tilde{d}^{*,(i)}}{\theta_2^{(i)}}\right) + \frac{\tilde{d}^{*,(i)}}{d_{max}^{(i)}} \exp\left(\frac{\tilde{d}^{*,(i)} - d_{max}^{(i)}}{\theta_2^{(1)}}\right) - 1 \right] \tilde{\pi}_2^{(i)}(d\theta_2^{(i)})$$

The Bayesian  $D$ -optimal design in the class  $\Xi_4^2$  with respect to the prior  $\pi = \tilde{\pi}_1 \times \tilde{\pi}_2 \times \tilde{\pi}_2^{(1)} \times \tilde{\pi}_2^{(2)}$  for maximum likelihood estimation in two exponential models with regression functions of the form (3.12) is

(1)  $\xi^{a,*} = (\xi_1^{a,*}, \xi_2^{a,*}, \mu^{a,*})$  if  $\log r \leq 0$  and  $\log r \leq u(\tilde{d}^{*,(1)}, \tilde{d}^{*,(2)})$ , where

$$\xi_1^{a,*} = \left\{ \begin{array}{cc} 0 & \tilde{d}^{*,(1)} \\ \frac{1}{3} & \frac{1}{3} \end{array} \right\}, \quad \xi_2^{a,*} = \left\{ \begin{array}{c} d_{max}^{(2)} \\ 1 \end{array} \right\}, \quad \mu^{a,*} = \left\{ \begin{array}{cc} 1 & 2 \\ \frac{3}{4} & \frac{1}{4} \end{array} \right\},$$

(2)  $\xi^{b_1,*} = (\xi_1^{b_1,*}, \xi_2^{b_1,*}, \mu^{b_1,*})$  if  $u(\tilde{d}^{*,(1)}, \tilde{d}^{*,(2)}) < \log r \leq 0$ , where

$$\xi_1^{b_1,*} = \left\{ \begin{array}{cc} 0 & d_{max}^{(1)} \\ \frac{1}{2} & \frac{1}{2} \end{array} \right\}, \quad \xi_2^{b_1,*} = \left\{ \begin{array}{cc} \tilde{d}^{*,(2)} & d_{max}^{(2)} \\ \frac{1}{2} & \frac{1}{2} \end{array} \right\}, \quad \mu^{b_1,*} = \left\{ \begin{array}{cc} 1 & 2 \\ \frac{1}{2} & \frac{1}{2} \end{array} \right\},$$

(3)  $\xi^{b_2,*} = (\xi_1^{b_2,*}, \xi_2^{b_2,*}, \mu^{b_2,*})$  if  $0 \leq \log r < u(\tilde{d}^{*,(1)}, \tilde{d}^{*,(2)})$ , where

$$\xi_1^{b_2,*} = \left\{ \begin{array}{cc} \tilde{d}^{*,(1)} & d_{max}^{(1)} \\ \frac{1}{2} & \frac{1}{2} \end{array} \right\}, \quad \xi_2^{b_2,*} = \left\{ \begin{array}{cc} 0 & d_{max}^{(2)} \\ \frac{1}{2} & \frac{1}{2} \end{array} \right\}, \quad \mu^{b_2,*} = \left\{ \begin{array}{cc} 1 & 2 \\ \frac{1}{2} & \frac{1}{2} \end{array} \right\},$$

(4)  $\xi^{c,*} = (\xi_1^{c,*}, \xi_2^{c,*}, \mu^{c,*})$  if  $\log r \geq 0$  and  $\log(r) > u(\tilde{d}^{*,(1)}, \tilde{d}^{*,(2)})$ , where

$$\xi_1^{c,*} = \left\{ \begin{array}{c} d_{max}^{(1)} \\ 1 \end{array} \right\}, \quad \xi_2^{c,*} = \left\{ \begin{array}{ccc} 0 & \tilde{d}^{*,(2)} & d_{max}^{(2)} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{array} \right\}, \quad \mu^{c,*} = \left\{ \begin{array}{cc} 1 & 2 \\ \frac{1}{4} & \frac{3}{4} \end{array} \right\}.$$

**Theorem 3.5.** Let  $r = \frac{\sigma_1^2}{\sigma_2^2}$  and  $\tilde{d}^{*,(i)}$  be a solution in the interval  $(0, d_{max}^{(i)})$  of the equation (2.10) using the marginal prior distribution  $\tilde{\pi}_2^{(i)}$  on  $\Theta_2^{(i)}$ . Also define the function

$$u(\tilde{d}^{*,(1)}, \tilde{d}^{*,(2)}) := 2 \sum_{i=1}^2 (-1)^{i+1} \int_{\Theta_2^{(i)}} \log \left[ \left( \frac{\tilde{d}^{*,(i)}}{\tilde{d}^{*,(i)} + \theta_2^{(i)}} \right) \left( g(d_{max}^{(i)}, \theta_2^{(i)}) - g(\tilde{d}^{*,(i)}, \theta_2^{(i)}) \right) \right] \tilde{\pi}_2^{(i)}(d\theta_2^{(i)})$$

where  $g(d, \theta_2) = \frac{d + \theta_2}{d} \log\left(\frac{d}{\theta_2} + 1\right)$ . The Bayesian  $D$ -optimal design in the class  $\Xi_4^2$  with respect to the prior  $\pi = \tilde{\pi}_1 \times \tilde{\pi}_2 \times \tilde{\pi}_2^{(1)} \times \tilde{\pi}_2^{(2)}$  for maximum likelihood estimation in two linear-in-log models with regression functions of the form (3.12) is

(1)  $\xi^{a,*} = (\xi_1^{a,*}, \xi_2^{a,*}, \mu^{a,*})$  if  $\log r \leq 0$  and  $\log r \leq u(\tilde{d}^{*,(1)}, \tilde{d}^{*,(2)})$ , where

$$\xi_1^{a,*} = \begin{Bmatrix} 0 & \tilde{d}^{*,(1)} & d_{max}^{(1)} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{Bmatrix}, \quad \xi_2^{a,*} = \begin{Bmatrix} d_{max}^{(2)} \\ 1 \end{Bmatrix}, \quad \mu^{a,*} = \begin{Bmatrix} 1 & 2 \\ \frac{3}{4} & \frac{1}{4} \end{Bmatrix},$$

(2)  $\xi^{b_1,*} = (\xi_1^{b_1,*}, \xi_2^{b_1,*}, \mu^{b_1,*})$  if  $u(\tilde{d}^{*,(1)}, \tilde{d}^{*,(2)}) < \log r \leq 0$ , where

$$\xi_1^{b_1,*} = \begin{Bmatrix} 0 & d_{max}^{(1)} \\ \frac{1}{2} & \frac{1}{2} \end{Bmatrix}, \quad \xi_2^{b_1,*} = \begin{Bmatrix} \tilde{d}^{*,(2)} & d_{max}^{(2)} \\ \frac{1}{2} & \frac{1}{2} \end{Bmatrix}, \quad \mu^{b_1,*} = \begin{Bmatrix} 1 & 2 \\ \frac{1}{2} & \frac{1}{2} \end{Bmatrix},$$

(3)  $\xi^{b_2,*} = (\xi_1^{b_2,*}, \xi_2^{b_2,*}, \mu^{b_2,*})$  if  $0 < \log r < u(\tilde{d}^{*,(1)}, \tilde{d}^{*,(2)})$ , where

$$\xi_1^{b_2,*} = \begin{Bmatrix} \tilde{d}^{*,(1)} & d_{max}^{(1)} \\ \frac{1}{2} & \frac{1}{2} \end{Bmatrix}, \quad \xi_2^{b_2,*} = \begin{Bmatrix} 0 & d_{max}^{(2)} \\ \frac{1}{2} & \frac{1}{2} \end{Bmatrix}, \quad \mu^{b_2,*} = \begin{Bmatrix} 1 & 2 \\ \frac{1}{2} & \frac{1}{2} \end{Bmatrix},$$

(4)  $\xi^{c,*} = (\xi_1^{c,*}, \xi_2^{c,*}, \mu^{c,*})$  if  $\log r \geq 0$  and  $\log(r) > u(\tilde{d}^{*,(1)}, \tilde{d}^{*,(2)})$ , where

$$\xi_1^{c,*} = \begin{Bmatrix} d_{max}^{(1)} \\ 1 \end{Bmatrix}, \quad \xi_2^{c,*} = \begin{Bmatrix} 0 & \tilde{d}^{*,(2)} & d_{max}^{(2)} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{Bmatrix}, \quad \mu^{c,*} = \begin{Bmatrix} 1 & 2 \\ \frac{1}{4} & \frac{3}{4} \end{Bmatrix}.$$

## 4 Example

In this section we assess the Bayesian  $D$ -optimal saturated designs characterised analytically in Section 3, via a numerical investigation of their performance in a situation where these are not Bayesian  $D$ -optimal amongst all designs.

In particular, we consider the case of  $M = 2$  different administration frequency groups where the Emax regression function is used to model the dose-response relationship for both groups. This choice of parametric model is frequent in dose-response studies see, for example, Gabrielsson and Weiner (2007) and Thomas et al. (2014) for recent references. It is also reasonable to assume that the placebo effect as well as the maximum effect that can be achieved by the drug under investigation is the same between groups. Therefore, the two Emax models are assumed to have common location and scale parameters which corresponds to the case considered in Section 3.3. Furthermore, for illustration purposes, the design spaces  $\mathcal{X}_1$  and  $\mathcal{X}_2$  are assumed to coincide with  $\mathcal{X}_1 = \mathcal{X}_2 = [0, 1]$  and also the variances in both groups are assumed to be the same with  $\sigma_1^2 = \sigma_2^2 = 1$ .

In what follows, the prior  $\pi = \tilde{\pi}_1 \times \tilde{\pi}_2 \times \tilde{\pi}_2^{(1)} \times \tilde{\pi}_2^{(2)}$  on the parameter space  $\Theta = \Theta_1 \times \Theta_2^{(1)} \times \Theta_2^{(2)}$  is used. The marginal priors  $\tilde{\pi}_1$  and  $\tilde{\pi}_2$  are set to the dirac measures  $\delta_{\theta_1}$  and  $\delta_{\theta_2}$  with  $(\theta_1, \theta_2)^T = (0, 1)^T$ , respectively, since the Emax model is linear in the location and scale parameter and varying these parameters does not affect the information matrix. For the priors  $\tilde{\pi}_2^{(1)}$  and  $\tilde{\pi}_2^{(2)}$  we use uniform distributions on the sets  $\{0.20, 0.275, 0.35, 0.425, 0.50\}$  and  $\{0.60, 0.675, 0.75, 0.825, 0.90\}$ , respectively and thus, we have a total of  $P = 5 \cdot 5 = 25$

different and equally probable parameter combinations. This corresponds to the case of no preference for specific parameter combinations and thus under the concept of uniform distributions there is no need for a prior to be specified, a step which is often difficult in practice.

The scenario described above corresponds to the first case of Theorem 3.3 according to which we calculate the Bayesian  $D$ -optimal design  $\xi^* = (\xi_1^*, \xi_2^*, \mu^*)$  in the class  $\Xi_4^2$  of saturated designs with respect to the prior  $\pi$  which is given by

$$\xi_1^* = \begin{Bmatrix} 0.00 & 0.1984207 & 1.00 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{Bmatrix}, \quad \xi_2^* = \begin{Bmatrix} 0.742427 \\ 1 \end{Bmatrix}, \quad \mu^* = \begin{Bmatrix} 1 & 2 \\ \frac{3}{4} & \frac{1}{4} \end{Bmatrix}. \quad (4.1)$$

Checking the inequalities (3.6) of the Equivalence Theorem 3.1 reveals that the design  $\xi^*$  is not Bayesian  $D$ -optimal amongst all designs since the inequality is not satisfied for the second administration frequency group (see Figure 1).

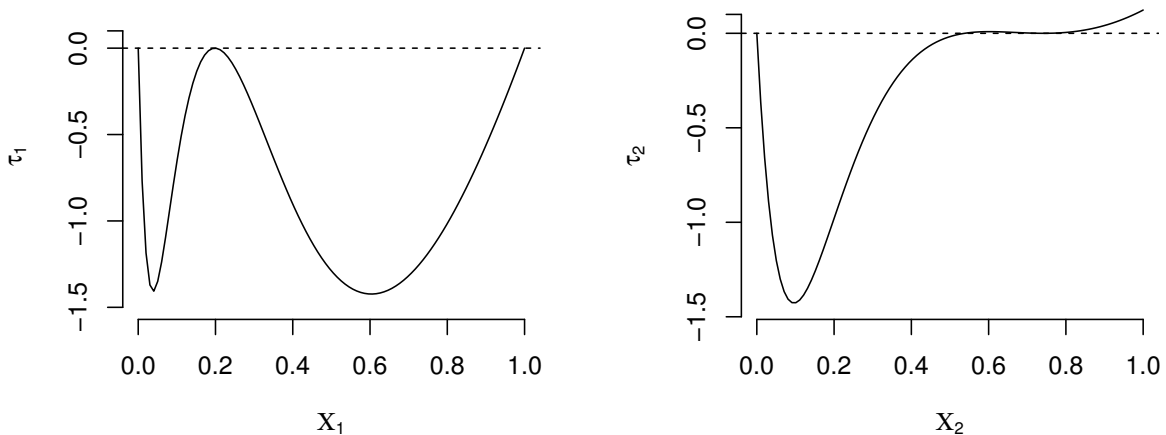


Figure 1: *Illustration of Theorem 3.1 for the Bayesian  $D$ -optimal design  $\xi^*$  in the class  $\Xi_4^2$ . The figures show the functions  $\tau_1$  and  $\tau_2$  defined by the left-hand side of inequality (3.6), respectively. Left panel:  $\tau_1$ . Right panel:  $\tau_2$ .*

Nevertheless, we further investigate the performance of the design  $\xi^*$  given in (4.1) since if it performs well then, as it is discussed in Section 3.1, it is preferred to the design that is Bayesian  $D$ -optimal amongst all designs. The reason is that  $\xi^*$  has minimum number of support points, that is, it requires fewer doses to be utilised in the two administration groups and thus it results in lower costs for the dose-response study without losing much information. We also note that finding the design that is Bayesian  $D$ -optimal amongst all possible designs is computationally intensive even for the case of only two administration

frequency groups and to the best of our knowledge no analytical characterisations of these designs, which would facilitate the evaluation, exist in the literature.

For the calculation of the design which is Bayesian  $D$ -optimal amongst all designs we used the particle swarm optimisation algorithm (see Clerc (2006) for details) and the resulting candidate design  $\xi^{*,B} = (\xi_1^{*,B}, \xi_2^{*,B}, \mu^{*,B})$  is given by

$$\begin{aligned} \xi_1^{*,B} &= \begin{Bmatrix} 0.19982 & 1.00 \\ 0.50148 & 0.49852 \end{Bmatrix}, & \xi_2^{*,B} &= \begin{Bmatrix} 0.00 & 0.56386 & 1.00 \\ 0.48649 & 0.26260 & 0.25091 \end{Bmatrix}, \\ \mu^{*,B} &= \begin{Bmatrix} 1 & 2 \\ 0.48691 & 0.51309 \end{Bmatrix}. \end{aligned} \quad (4.2)$$

As shown in Figure 2 both inequalities (3.6) of the Equivalence Theorem 3.1 are satisfied for this candidate design and therefore,  $\xi^{*,B}$  given in (4.2) is in fact Bayesian  $D$ -optimal amongst all designs.

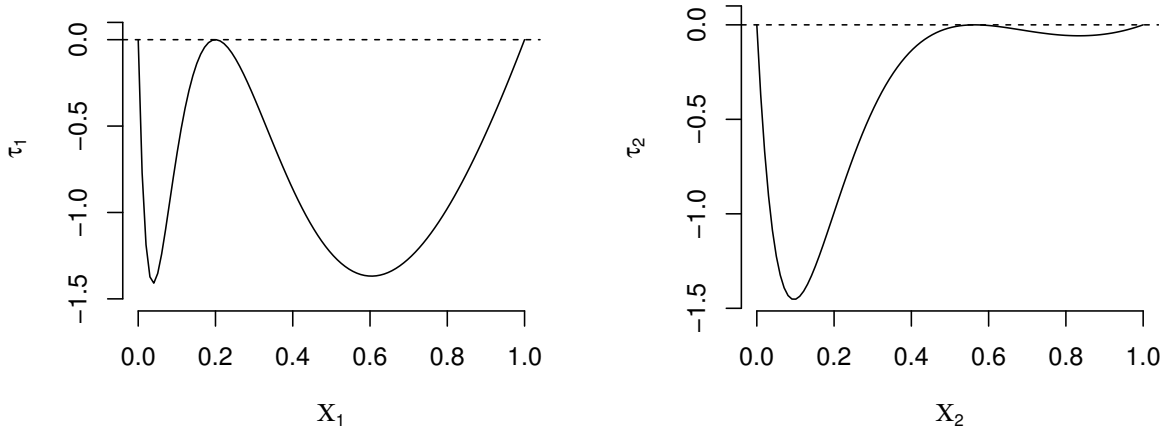


Figure 2: *Illustration of Theorem 3.1 for the Bayesian  $D$ -optimal design  $\xi^{*,B}$ . The figures show the functions  $\tau_1$  and  $\tau_2$  defined by the left-hand side of inequality (3.6), respectively. Left panel:  $\tau_1$ . Right panel:  $\tau_2$ .*

We now assess the performance of a given design  $\xi$  (for a specific parameter vector  $\boldsymbol{\theta}$ ) via the  $D$ -efficiency defined as

$$\text{eff}(\xi, \boldsymbol{\theta}) = \frac{\det(M(\xi, \boldsymbol{\theta}))^{1/4}}{\det(M(\xi^*(\boldsymbol{\theta}), \boldsymbol{\theta}))^{1/4}} \in [0, 1], \quad (4.3)$$

where  $\xi^*(\boldsymbol{\theta})$  is the locally  $D$ -optimal design with respect to the specific parameter vector  $\boldsymbol{\theta}$ . The closer the  $D$ -efficiency of  $\xi$  is to 1, the fewer replicates are required for the design  $\xi$  to be able to estimate the parameters with the same precision as  $\xi^*(\boldsymbol{\theta})$ .

Table 2 presents the  $D$ -efficiencies of  $\xi^*$  and  $\xi^{*,B}$ , given in (4.1) and (4.2) respectively, for the various parameter combinations resulting from the choice of prior  $\pi$ . For example, for the parameter vector  $\boldsymbol{\theta} = (0, 1, 0.2, 0.6)$  the  $D$ -efficiencies of  $\xi^*$  and  $\xi^{*,B}$  are 97.52% and 97.31%, respectively.

		$\theta_2^{(1)}$				
		0.20	0.275	0.35	0.425	0.50
$\text{eff}(\xi^*, (0, 1, \theta_2^{(1)}, \theta_2^{(2)}))$	$\theta_2^{(2)}$ 0.6	97.52	99.20	98.68	96.94	94.25
	0.675	97.96	99.65	99.50	98.26	96.23
	0.75	98.07	99.76	99.81	98.91	97.37
	0.825	97.94	99.62	99.77	99.10	97.90
	0.90	97.62	99.30	99.51	98.97	98.02
$\text{eff}(\xi^{*,B}, (0, 1, \theta_2^{(1)}, \theta_2^{(2)}))$	$\theta_2^{(2)}$ 0.6	97.31	99.39	99.41	98.38	96.59
	0.675	97.58	99.59	99.88	99.22	97.95
	0.75	97.58	99.53	99.94	99.52	98.60
	0.825	97.40	99.30	99.74	99.46	98.78
	0.90	97.07	98.93	99.38	99.16	98.63

Table 2: The efficiencies (in %) of the Bayesian  $D$ -optimal design  $\xi^*$  in the class  $\Xi_4^2$  and of the Bayesian  $D$ -optimal design  $\xi^{*,B}$  for the  $P = 25$  different parameter combinations.

Note that the  $D$ -efficiencies of both designs  $\xi^*$  and  $\xi^{*,B}$  depicted in Table 2 are between 94% and 99%. Consequently, using either one of these two designs results in parameter estimation with accuracy very close to the one the locally  $D$ -optimal design  $\xi^*(\boldsymbol{\theta})$  would have achieved. Moreover, for every parameter combination the  $D$ -efficiency of the design  $\xi^*$  which is Bayesian  $D$ -optimal only in the class of saturated designs is close to the corresponding  $D$ -efficiency of the design  $\xi^{*,B}$  which is Bayesian  $D$ -optimal amongst all designs. We therefore, conclude that our analytical characterisations of Bayesian  $D$ -optimal saturated designs can be used without detriment thus reducing both the numerical effort for design search as well as the cost of the dose-response study.

### Acknowledgements

This work has been supported in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823, Project C2) of the German Research Foundation (DFG). Schorning was partially supported by a grant from the National Institute Of General Medical Sciences of the National Institutes of Health under Award Number R01GM107639. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institutes of Health).

## A Appendix

### *Proof of Lemma 2.1:*



*Proof.* Let  $\tilde{\xi} = \{d_1, \dots, d_{\tilde{m}}; \omega_1, \dots, \omega_{\tilde{m}}\}$  be any saturated design on  $\mathcal{X} = [0, d_{max}]$ . Also let  $\tilde{X}$  be the  $\tilde{m} \times \tilde{m}$  matrix with  $j$ th column given by  $\tilde{h}(d_j, \boldsymbol{\theta})$ ,  $j = 1, \dots, \tilde{m}$  and  $\tilde{W} = \text{diag}(\omega_1, \dots, \omega_{\tilde{m}})$ . Using this notation the information matrix defined in (2.2) and (2.3) becomes

$$M_{\text{ind}}(\tilde{\xi}, \tilde{\boldsymbol{\theta}}) = \tilde{X}\tilde{W}\tilde{X}^T,$$

and thus its determinant is given by

$$\det \left\{ M_{\text{ind}}(\tilde{\xi}, \tilde{\boldsymbol{\theta}}) \right\} = [\det\{\tilde{X}\}]^2 \det\{\tilde{W}\}.$$

Hence the Bayesian  $D$ -optimality criterion given in (2.4) becomes

$$\Phi_{\tilde{\pi}}(\tilde{\xi}) = \int_{\tilde{\Theta}} \left( 2 \log |\det\{\tilde{X}\}| + \sum_{j=1}^{\tilde{m}} \log \omega_j \right) \tilde{\pi}(d\tilde{\boldsymbol{\theta}}).$$

Maximising the above expression with respect to the weights  $\omega_1, \dots, \omega_{\tilde{m}}$  under the condition  $\sum_{j=1}^{\tilde{m}} \omega_j = 1$  gives  $\omega_j = \frac{1}{\tilde{m}}$  for all  $j = 1, \dots, \tilde{m}$  which proves the assertion. □

### ***Proof of Theorem 2.2:***

*Proof.* From Lemma 2.1 we know that a Bayesian  $D$ -optimal saturated design has equal weights. Therefore, let  $\tilde{\xi} = \{d_1, d_2, d_3; 1/3, 1/3, 1/3\}$  be an equally weighted three-point design where  $0 \leq d_1 < d_2 < d_3 \leq d_{max}$ .

Theorem 4.1 in Dette et al. (2010) shows that for any of the models (2.5), (2.6) and (2.7) the locally  $D$ -optimal design is supported at exactly three points two of which are the end-points of the design space. Therefore, for any of these models the determinant of the information matrix (2.2) of the equally weighted three-point design  $\tilde{\xi}$  is decreasing with  $d_1$  and increasing  $d_3$ .

As a result, the Bayesian  $D$ -optimality criterion function given in (2.4) is also decreasing with  $d_1$  and increasing with  $d_3$  and thus maximised at  $d_1^* = 0$  and  $d_3^* = d_{max}$ . That is, for any of the models (2.5), (2.6) and (2.7) a Bayesian  $D$ -optimal saturated design is always supported at the end-points of the design space. The non-trivial support point can be found by solving the equation

$$\frac{\partial}{\partial d_2} \Phi_{\tilde{\pi}}(\tilde{\xi})|_{d_1=0, d_3=d_{max}} = 0 \iff \int_{\tilde{\Theta}_2} \frac{\partial}{\partial d_2} \left( \log \det\{M_{\text{ind}}(\tilde{\xi}, \tilde{\boldsymbol{\theta}})\} \right) |_{d_1=0, d_3=d_{max}} \tilde{\pi}_2(d\boldsymbol{\theta}_2) = 0,$$

for  $d_2 \in (0, d_{max})$ . Using the appropriate information matrix for each of the three models under consideration yields the equations of Theorem 2.2. □

**Proof of Theorem 3.1:**

*Proof.* For a fixed  $\boldsymbol{\theta}$  consider the set

$$\mathcal{M}_{\boldsymbol{\theta}} = \{M(\xi_{\pi}, \boldsymbol{\theta}) | \xi_{\pi} = (\xi_1, \dots, \xi_M, \mu) \in \Xi_1(\mathcal{X}_1) \times \dots \times \Xi_M(\mathcal{X}_M) \times \Xi(\{1, \dots, M\})\},$$

where  $\Xi_1(\mathcal{X}_1) \times \dots \times \Xi_M(\mathcal{X}_M) \times \Xi(\{1, \dots, M\})$  denotes the class of all designs on  $\mathcal{X}_1 \times \dots \times \mathcal{X}_M \times \{1, \dots, M\}$ , and note that  $\mathcal{M}_{\boldsymbol{\theta}}$  is the convex hull of

$$\mathcal{D}_{\boldsymbol{\theta}} = \{M^{(i)}(\delta_d, \boldsymbol{\theta}) = h_i(d, \boldsymbol{\theta})h_i^T(d, \boldsymbol{\theta}) | i \in \{1, \dots, M\}, d \in \mathcal{X}_i\}.$$

Using Theorem 3.3 in Dette et al. (2007), a design  $\xi_{\pi}^* = (\xi_1^*, \dots, \xi_M^*, \mu^*)$  is Bayesian  $D$ -optimal, with respect to the prior  $\pi$ , in the class  $\Xi_1(\mathcal{X}_1) \times \dots \times \Xi_M(\mathcal{X}_M) \times \Xi(\{1, \dots, M\})$  if and only if the inequality

$$\int_{\Theta} \text{tr}\{(M(\xi_{\pi}^*, \boldsymbol{\theta}))^{-1} E_{\boldsymbol{\theta}}\} \pi(d\boldsymbol{\theta}) \leq m,$$

holds for all matrices  $E_{\boldsymbol{\theta}} \in \mathcal{M}_{\boldsymbol{\theta}}$ .

Since  $\mathcal{M}_{\boldsymbol{\theta}}$  is the convex hull of  $\mathcal{D}_{\boldsymbol{\theta}}$ , it is sufficient to prove inequality for all  $E_{\boldsymbol{\theta}} \in \mathcal{D}_{\boldsymbol{\theta}}$ . Therefore,  $\xi_{\pi}^*$  is Bayesian  $D$ -optimal, with respect to the prior  $\pi$ , in the class  $\Xi_1(\mathcal{X}_1) \times \dots \times \Xi_M(\mathcal{X}_M) \times \Xi(\{1, \dots, M\})$  if and only if the inequality

$$\int_{\Theta} \text{tr}\{(M(\xi_{\pi}^*, \boldsymbol{\theta}))^{-1} h_i(d, \boldsymbol{\theta})h_i^T(d, \boldsymbol{\theta})\} \pi(d\boldsymbol{\theta}) \leq m,$$

holds for all  $d \in \mathcal{X}_i$  and  $i = 1, \dots, M$ . Rewriting the above gives the inequalities of Theorem 3.1.  $\square$

**Proof of Lemma 3.1:**

*Proof.* The information matrix of the design  $\xi_{\pi}$  can be rewritten as

$$M(\xi_{\pi}, \boldsymbol{\theta}) = XW X^T,$$

where

$$X = \left( h_1(d_1^{(1)}), \dots, h_1(d_{m_1}^{(1)}), \dots, h_M(d_1^{(M)}), \dots, h_M(d_{m_M}^{(M)}) \right) \in \mathbb{R}^{m \times m},$$

and

$$W = \text{diag}(\lambda_1 \omega_{11}, \dots, \lambda_1 \omega_{1m_1}, \dots, \lambda_M \omega_{M1}, \dots, \lambda_M \omega_{Mm_M}) \in \mathbb{R}^{m \times m}.$$

Then the Bayesian  $D$ -optimality criterion function given in (3.5) becomes

$$\int_{\Theta} \log \det\{M(\xi_{\pi}, \boldsymbol{\theta})\} \pi(d\boldsymbol{\theta}) = \int_{\Theta} (2 \log |\det\{X\}| + \log \det\{W\}) \pi(d\boldsymbol{\theta}).$$

Maximising

$$\log \det\{W\} = \sum_{i=1}^M \log \lambda_i \sum_{j=1}^{m_i} \log \omega_{ij},$$

under the conditions  $\sum_{i=1}^M \lambda_i = 1$ ,  $\sum_{j=1}^{m_i} \omega_{ij} = 1$  and  $\sum_{i=1}^M m_i = m$  gives

$$\lambda_i = \frac{m_i}{m}, \quad \omega_{ij} = \frac{1}{m_i}, (i = 1, \dots, M; j = 1, \dots, m_i),$$

which proves the assertion. □

### ***Proof of Theorem 3.2***

*Proof.* For the sake of transparency we restrict ourselves to the case  $M = 2$  such that  $m = 2q + 1$ .

By an application of Lemma 3.1 we obtain the following two candidates  $\xi^a = (\xi_1^a, \xi_2^a, \mu^a)$  and  $\xi^b = (\xi_1^b, \xi_2^b, \mu^b)$  for a Bayesian  $D$ -optimal design in the class  $\Xi_{2q+1}^2$  with

$$\begin{aligned} \xi_1^a &= \left\{ \begin{matrix} d_0^{(1)} & d_1^{(1)} & \dots & d_q^{(1)} \\ \frac{1}{q+1} & \frac{1}{q+1} & \dots & \frac{1}{q+1} \end{matrix} \right\}, & \xi_2^a &= \left\{ \begin{matrix} d_1^{(2)} & \dots & d_q^{(2)} \\ \frac{1}{q} & \dots & \frac{1}{q} \end{matrix} \right\}, & \mu^a &= \left\{ \begin{matrix} 1 & 2 \\ \frac{q+1}{2q+1} & \frac{q}{2q+1} \end{matrix} \right\}, \\ \xi_1^b &= \left\{ \begin{matrix} d_1^{(1)} & \dots & d_q^{(1)} \\ \frac{1}{q} & \dots & \frac{1}{q} \end{matrix} \right\}, & \xi_2^b &= \left\{ \begin{matrix} d_0^{(2)} & d_1^{(2)} & \dots & d_q^{(2)} \\ \frac{1}{q+1} & \frac{1}{q+1} & \dots & \frac{1}{q+1} \end{matrix} \right\}, & \mu^b &= \left\{ \begin{matrix} 1 & 2 \\ \frac{q}{2q+1} & \frac{q+1}{2q+1} \end{matrix} \right\}. \end{aligned}$$

We now evaluate the criterion values  $\Phi_{\tilde{\pi}}(\xi^a)$  and  $\Phi_{\tilde{\pi}}(\xi^b)$  and maximise them with respect to the support points of the design  $\xi^a$  and  $\xi^b$ , respectively.

For the design  $\xi^a$ , the criterion value  $\Phi_{\tilde{\pi}}(\xi^a)$  can be rewritten by

$$\begin{aligned} \Phi_{\tilde{\pi}}(\xi^a) &= \int_{\Theta_2^{(1)}} (q+1) \log \frac{1}{\sigma_1^2} + (q+1) \log \frac{1}{q+1} + 2 \log |\det\{X_1(d_0^{(1)}, d_1^{(1)}, \dots, d_q^{(1)}, \theta_2^{(1)})\}| \tilde{\pi}_2^{(1)}(d\theta_2^{(1)}) \\ &\quad + \int_{\Theta_2^{(2)}} q \log \frac{1}{\sigma_2^2} + q \log \frac{1}{q} + 2 \log |\det\{X_2(d_1^{(2)}, \dots, d_q^{(2)}, \theta_2^{(2)})\}| \tilde{\pi}_2^{(2)}(d\theta_2^{(2)}) + c(q), \end{aligned} \tag{A.1}$$

where the  $(q+1) \times (q+1)$ -dimensional matrix  $X_1(d_0, d_1, \dots, d_q, \theta_2)$  is given by

$$X_1(d_0, d_1, \dots, d_q, \theta_2) = \begin{pmatrix} 1 & 1 & \dots & 1 \\ \eta_0(d_0, \theta_2) & \eta_0(d_1, \theta_2) & \dots & \eta_0(d_q, \theta_2) \end{pmatrix},$$

the  $q \times q$ -dimensional matrix  $X_2(d_1, \dots, d_q, \theta_2)$  is given by

$$X_2(d_1, \dots, d_q, \theta_2) = (\eta_0(d_1, \theta_2) \quad \dots \quad \eta_0(d_q, \theta_2)),$$

the constant  $c(q)$  only depends on the number of the non-common parameters  $q$ .

The first term of the sum in (A.1) coincides with Bayesian  $D$ -criterion for the individual model of the form (3.8) with respect to the prior  $\tilde{\pi}_2^{(1)}$  and is therefore maximised by the corresponding Bayesian  $D$ -optimal design  $\tilde{\xi}_{\tilde{\pi}_2^{(1)}}^*$  given by (3.10). Because of the conditions in (3.9) the second term of the sum in (A.1) can be rewritten by

$$\begin{aligned} & \int_{\Theta_2^{(2)}} q \log \frac{1}{\sigma_2^2} + q \log \frac{1}{q} + 2 \log |\det \{X_2(d_1^{(2)}, \dots, d_q^{(2)}, \theta_2^{(2)})\}| \tilde{\pi}_2^{(2)}(d\theta_2^{(2)}) \\ &= \int_{\Theta_2^{(2)}} q \log \frac{1}{\sigma_2^2} + q \log \frac{1}{q} + 2 \log |\det \{X_1(0, d_1^{(2)}, \dots, d_q^{(2)}, \theta_2^{(2)})\}| \tilde{\pi}_2^{(2)}(d\theta_2^{(2)}), \end{aligned}$$

where the latter expression again corresponds to the Bayesian  $D$ -criterion for the individual model of the form (3.8) with respect to the prior  $\tilde{\pi}_2^{(2)}$  under the additional constraint that the smallest support point  $d_0^{(2)} = 0$  is fixed to zero. Maximising this expression with respect to the remaining  $q$  support points  $d_1^{(2)}, \dots, d_q^{(2)}$  we receive that  $\xi_2^a = \xi_2^*$  where  $\xi_2^*$  is defined by (3.11).

A similar calculation for the criterion value  $\Phi_{\tilde{\pi}}(\xi^b)$  results in the maximising design  $\xi^b$  where  $\xi_1^b = \xi_1^*$  and  $\xi_2^b = \tilde{\xi}_{\tilde{\pi}_2^{(2)}}^*$  which are defined in (3.11) and (3.10), respectively.

Taking into account that  $\sigma_1^2 \leq \sigma_2^2$  a comparison of the criterion values evaluated in  $\xi^a$  and  $\xi^b$  results in Theorem 3.2.  $\square$

### ***Proof of Theorem 3.3:***

*Proof.* Let  $\xi_{\pi} = (\xi_1, \xi_2, \mu)$  be an arbitrary design in the class  $\Xi_4^2$  of saturated designs on  $\mathcal{X}_1 \times \mathcal{X}_2 \times \{1, 2\}$ . The information matrix of the design  $\xi_{\pi}$  is given by

$$M(\xi_{\pi}, \boldsymbol{\theta}) = \lambda \int_0^{d_{max}^{(1)}} h_1(d, \boldsymbol{\theta}) h_1^T(d, \boldsymbol{\theta}) d\xi_1(\boldsymbol{\theta}) + (1 - \lambda) \int_0^{d_{max}^{(2)}} h_2(d, \boldsymbol{\theta}) h_2^T(d, \boldsymbol{\theta}) d\xi_2(\boldsymbol{\theta}),$$

where for Emax models

$$h_1(d, \boldsymbol{\theta}) = \frac{1}{\sigma_1} \left( 1, \frac{d}{d + \theta_2^{(1)}}, \frac{d}{(d + \theta_2^{(1)})^2}, 0 \right)^T, \quad h_2(d, \boldsymbol{\theta}) = \frac{1}{\sigma_2} \left( 1, \frac{d}{d + \theta_2^{(2)}}, 0, \frac{d}{(d + \theta_2^{(2)})^2} \right)^T.$$

Let  $\pi = \tilde{\pi}_1 \times \tilde{\pi}_2 \times \tilde{\pi}_2^{(1)} \times \tilde{\pi}_2^{(2)}$  be a prior distribution on the parameter space  $\Theta = \Theta_1 \times \Theta_2 \times \Theta_2^{(1)} \times \Theta_2^{(2)}$ . By an application of Lemma 3.1 we obtain the following candidate designs for the Bayesian  $D$ -optimal design in the class  $\Xi_4^2$ .

$$\begin{aligned} \xi_1^a &= \left\{ \begin{matrix} d_1^{(1)} \\ \frac{1}{3} \end{matrix}, \begin{matrix} d_2^{(1)} \\ \frac{1}{3} \end{matrix}, \begin{matrix} d_3^{(1)} \\ \frac{1}{3} \end{matrix} \right\}, & \xi_2^a &= \left\{ \begin{matrix} d_1^{(2)} \\ 1 \end{matrix} \right\}, & \mu^a &= \left\{ \begin{matrix} 1 & 2 \\ \frac{3}{4} & \frac{1}{4} \end{matrix} \right\}, \\ \xi_1^b &= \left\{ \begin{matrix} d_1^{(1)} \\ \frac{1}{2} \end{matrix}, \begin{matrix} d_2^{(1)} \\ \frac{1}{2} \end{matrix} \right\}, & \xi_2^b &= \left\{ \begin{matrix} d_1^{(2)} \\ \frac{1}{2} \end{matrix}, \begin{matrix} d_2^{(2)} \\ \frac{1}{2} \end{matrix} \right\}, & \mu^b &= \left\{ \begin{matrix} 1 & 2 \\ \frac{1}{2} & \frac{1}{2} \end{matrix} \right\}, \end{aligned}$$

$$\xi_1^c = \left\{ \begin{matrix} d_1^{(1)} \\ 1 \end{matrix} \right\}, \quad \xi_2^c = \left\{ \begin{matrix} d_1^{(2)} & d_2^{(2)} & d_3^{(2)} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{matrix} \right\}, \quad \mu^c = \left\{ \begin{matrix} 1 & 2 \\ \frac{1}{4} & \frac{3}{4} \end{matrix} \right\}.$$

We now evaluate the Bayesian  $D$ -optimality criterion function given in (3.5) for each of these candidate designs and maximise it with respect to the support points.

For the design  $\xi_\pi^a = (\xi_1^a, \xi_2^a, \mu^a)$  the criterion function becomes

$$\begin{aligned} \Phi_\pi(\xi_\pi^a) &= \int_{\Theta_2 \times \Theta_2^{(1)}} \log \left[ \det \left\{ M_{\text{ind}}(\xi_1^a, \theta_2, \theta_2^{(1)}) \right\} \right] (\tilde{\pi}_2 \times \tilde{\pi}_2^{(1)})(d\theta_2 d\theta_2^{(1)}) \\ &+ \int_{\Theta_2 \times \Theta_2^{(2)}} \log \left[ \left( \frac{1}{4} \right)^4 3^3 \frac{1}{\sigma_2^2} \theta_2^2 \left( \frac{d_1^{(2)}}{(d_1^{(2)} + \theta_2^{(2)})^2} \right)^2 \right] (\tilde{\pi}_2 \times \tilde{\pi}_2^{(2)})(d\theta_2 d\theta_2^{(2)}). \end{aligned}$$

From Theorem 2.2 we have that the first term is maximised for  $d_1^{(1)} = 0, d_3^{(1)} = d_{max}^{(1)}$  and  $d_2^{(1)} = \tilde{d}^{*,(1)}$  where the latter is a solution of equation (2.8) for the marginal prior  $\tilde{\pi}_2^{(1)}$ . The second term is maximised at  $d_1^{(2)} = d^{*,(2)}$  which is a solution of the equation

$$\int_{\Theta_2^{(2)}} \frac{1}{d} - \frac{2}{d + \theta_2^{(2)}} \tilde{\pi}_2^{(2)}(d\theta_2^{(2)}), \quad (\text{A.2})$$

in the interval  $[0, d_{max}^{(2)}]$ . Hence the design of the form  $\xi_\pi^a$  maximising the Bayesian  $D$ -optimality criterion function is  $\xi_\pi^{a,*} = (\xi_1^{a,*}, \xi_2^{a,*}, \mu^{a,*})$  where

$$\xi_1^{a,*} = \left\{ \begin{matrix} 0 & \tilde{d}^{*,(1)} & d_{max}^{(1)} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{matrix} \right\}, \quad \xi_2^{a,*} = \left\{ \begin{matrix} d^{*,(2)} \\ 1 \end{matrix} \right\}, \quad \mu^{a,*} = \left\{ \begin{matrix} 1 & 2 \\ \frac{3}{4} & \frac{1}{4} \end{matrix} \right\}.$$

Following similar arguments we have that the design of the form  $\xi_\pi^c = (\xi_1^c, \xi_2^c, \mu^c)$  maximising the Bayesian  $D$ -optimality criterion function is  $\xi_\pi^{c,*} = (\xi_1^{c,*}, \xi_2^{c,*}, \mu^{c,*})$  where

$$\xi_1^{c,*} = \left\{ \begin{matrix} d^{*,(1)} \\ 1 \end{matrix} \right\}, \quad \xi_2^{c,*} = \left\{ \begin{matrix} 0 & \tilde{d}^{*,(2)} & d_{max}^{(2)} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{matrix} \right\}, \quad \mu^{c,*} = \left\{ \begin{matrix} 1 & 2 \\ \frac{1}{4} & \frac{3}{4} \end{matrix} \right\},$$

and  $d^{*,(1)}$  and  $\tilde{d}^{*,(2)}$  are solutions of the equations (A.2) using the prior  $\tilde{\pi}_2^{(1)}$  and (2.8) using the prior  $\tilde{\pi}_2^{(2)}$  respectively.

For the  $D$ -optimality of the design  $\xi_\pi^b = (\xi_1^b, \xi_2^b, \mu^b)$  we note that the smallest support points  $d_1^{(1)}$  and  $d_1^{(2)}$  of the components  $\xi_1^b$  and  $\xi_2^b$  must satisfy  $d_1^{(1)} + d_1^{(2)} > 0$  otherwise the determinant of the information matrix vanishes. Hence there exist two possibilities corresponding to the cases  $d_1^{(1)} = 0$  or  $d_1^{(2)} = 0$ . That is,

$$\begin{aligned} \xi_1^{b1} &= \left\{ \begin{matrix} 0 & d_2^{(1)} \\ \frac{1}{2} & \frac{1}{2} \end{matrix} \right\}, \quad \xi_2^{b1} = \left\{ \begin{matrix} d_1^{(2)} & d_2^{(2)} \\ \frac{1}{2} & \frac{1}{2} \end{matrix} \right\}, \quad \mu^{b1} = \left\{ \begin{matrix} 1 & 2 \\ \frac{1}{2} & \frac{1}{2} \end{matrix} \right\}, \\ \xi_1^{b2} &= \left\{ \begin{matrix} d_1^{(1)} & d_2^{(1)} \\ \frac{1}{2} & \frac{1}{2} \end{matrix} \right\}, \quad \xi_2^{b2} = \left\{ \begin{matrix} 0 & d_2^{(2)} \\ \frac{1}{2} & \frac{1}{2} \end{matrix} \right\}, \quad \mu^{b2} = \left\{ \begin{matrix} 1 & 2 \\ \frac{1}{2} & \frac{1}{2} \end{matrix} \right\}. \end{aligned}$$

Now following the same arguments as for the previous candidate designs we conclude that the designs  $\xi_{\pi}^{b_{1,*}} = (\xi_1^{b_{1,*}}, \xi_2^{b_{1,*}}, \mu^{b_{1,*}})$  and  $\xi_{\pi}^{b_{2,*}} = (\xi_1^{b_{2,*}}, \xi_2^{b_{2,*}}, \mu^{b_{2,*}})$  maximising the Bayesian  $D$ -optimality criterion function, among all designs of the corresponding form, are given by

$$\xi_1^{b_{1,*}} = \left\{ \begin{array}{cc} 0 & d^{*,(1)} \\ \frac{1}{2} & \frac{1}{2} \end{array} \right\}, \quad \xi_2^{b_{1,*}} = \left\{ \begin{array}{cc} \tilde{d}^{*,(2)} & d_{max}^{(2)} \\ \frac{1}{2} & \frac{1}{2} \end{array} \right\}, \quad \mu^{b_{1,*}} = \left\{ \begin{array}{cc} 1 & 2 \\ \frac{1}{2} & \frac{1}{2} \end{array} \right\},$$

and

$$\xi_1^{b_{2,*}} = \left\{ \begin{array}{cc} \tilde{d}^{*,(1)} & d_{max}^{(1)} \\ \frac{1}{2} & \frac{1}{2} \end{array} \right\}, \quad \xi_2^{b_{2,*}} = \left\{ \begin{array}{cc} 0 & d^{*,(2)} \\ \frac{1}{2} & \frac{1}{2} \end{array} \right\}, \quad \mu^{b_{2,*}} = \left\{ \begin{array}{cc} 1 & 2 \\ \frac{1}{2} & \frac{1}{2} \end{array} \right\},$$

respectively, where  $d^{*,(i)}$  and  $\tilde{d}^{*,(i)}$  are solutions of the equations (A.2) and (2.8) respectively using the prior  $\tilde{\pi}_2^{(i)}$ ,  $i = 1, 2$ .

Finally the assertion of the theorem follows by straightforward calculations comparing the Bayesian  $D$ -optimality criterion function of the designs  $\xi_{\pi}^{a,*}$ ,  $\xi_{\pi}^{b_{1,*}}$ ,  $\xi_{\pi}^{b_{2,*}}$ ,  $\xi_{\pi}^{c,*}$  in the different scenarios. □

## References

- Bretz, F., Dette, H. and Pinheiro, J. (2010) Practical considerations for optimal designs in clinical dose finding studies. *Statistics in Medicine*, **29**, 731–742.
- Chaloner, K. (1989) Bayesian design for estimating the turning point of a quadratic regression. *Comm. Statist., Theory Methods*, **18**, 1385–1400.
- (1993) A note on optimal bayesian design for nonlinear problems. *Journal of Statistical Planning and Inference*, **37**, 229 – 235.
- Chaloner, K. and Larntz, K. (1989) Optimal bayesian design applied to logistic regression experiments. *Journal of Statistical Planning and Inference*, **21**, 191 – 208.
- (1992) Bayesian design for accelerated life testing. *Journal of Statistical Planning and Inference*, **33**, 245 – 259.
- Chaloner, K. and Verdinelli, I. (1995) Bayesian experimental design: A review. *Statistical Science*, **10**, 273–304.
- Clerc, M. (2006) *Particle Swarm Optimization*. London: Iste Publishing Company.
- Dette, H., Bretz, F., Pepelyshev, A. and Pinheiro, J. (2008) Optimal designs for dose-finding studies. *Journal of the American Statistical Association*, **103**, 1225–1237.
- Dette, H., Haines, L. M. and Imhof, L. A. (2007) Maximin and bayesian optimal designs for regression models. *Statistica Sinica*, **17**, 463–480.

- Dette, H., Kiss, C., Bevanda, M. and Bretz, F. (2010) Optimal designs for the EMAX, log-linear and exponential models. *Biometrika*, **97**, 513–518.
- Dragalin, V., Hsuan, F. and Padmanabhan, S. K. (2008) Two-stage design for dose-finding that accounts for both efficacy and safety. *Statistics in Medicine*, **27**, 5156–5176.
- Feller, C., Schorning, K., Dette, H., Bermann, G. and Bornkamp, B. (2017) Optimal designs for dose response curves with common parameters. *to appear in: The Annals of Statistics*, **45**, 2102–2132.
- Gabrielsson, J. and Weiner, D. (2007) *Pharmacokinetic and Pharmacodynamic Data Analysis: Concepts and Applications*. Stockholm: Swedish Pharmaceutical Press, 4th edn.
- Kiefer, J. (1974) General equivalence theory for optimum designs (approximate theory). *The Annals of Statistics*, **2**, 849–879.
- Pronzato, L. and Walter, E. (1985) Robust experiment design via stochastic approximation. *Mathematical Biosciences*, **75**, 103–120.
- Sacks, L. V., Shamsuddin, H. H., Yasinskaya, Y. I., Bouri, K., Lanthier, M. L. and Sherman, R. E. (2014) Scientific and regulatory reasons for delay and denial of fda approval of initial applications for new drugs, 2000-2012. *Journal of the American Medical Association*, **311**, 378–384.
- Thomas, N., Sweeney, K. and Somayaji, V. (2014) Meta-analysis of clinical dose-response in a large drug development portfolio. *Statistics in Biopharmaceutical Research*, **6**, 302–317.