The Annals of Statistics
2017, Vol. 45, No. 5, 2102–2132
DOI: 10.1214/16-AOS1520
© Institute of Mathematical Statistics, 2017

OPTIMAL DESIGNS FOR DOSE RESPONSE CURVES WITH COMMON PARAMETERS[1]

BY CHRYSTEL FELLER*, KIRSTEN SCHORNING†, HOLGER DETTE†, GEORGINA BERMANN* AND BJÖRN BORNKAMP*

*Novartis Pharma AG** *and Ruhr-Universität Bochum*†

A common problem in Phase II clinical trials is the comparison of dose response curves corresponding to different treatment groups. If the effect of the dose level is described by parametric regression models and the treatments differ in the administration frequency (but not in the sort of drug), a reasonable assumption is that the regression models for the different treatments share common parameters.

This paper develops optimal design theory for the comparison of different regression models with common parameters. We derive upper bounds on the number of support points of admissible designs, and explicit expressions for D-optimal designs are derived for frequently used dose response models with a common location parameter. If the location and scale parameter in the different models coincide, minimally supported designs are determined and sufficient conditions for their optimality in the class of all designs derived. The results are illustrated in a dose-finding study comparing monthly and weekly administration.

1. Introduction. Adequately describing the dose-response relationship of a pharmaceutical compound is of paramount importance for achieving a successful clinical development. Sacks et al. (2014) recently conducted a review of the reasons for delay or denial of approval of drugs by the Food and Drug Administration (FDA). For those drug submissions that were not approved in the first-time application, one of the most frequent deficiencies was a statistical uncertainty related to the selected dose, illustrating the importance of clearly determining an efficacious and safe dose in Phase II dose-finding trials.

Efforts to improve this situation have led to the introduction of dose-response modeling approaches in a prospective manner as the primary analysis method in dose finding studies, and have become increasingly widespread in the past few years [see among many others, Grieve and Krams (2005), Bretz, Pinheiro

Received March 2016; revised August 2016.

[1]This work has been supported in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823, Teilprojekt C2) of the German Research Foundation (DFG) and by a grant from the National Institute Of General Medical Sciences of the National Institutes of Health under Award Number R01GM107639.

MSC2010 subject classifications. Primary 62K05; secondary 62F03.

Key words and phrases. Admissible design, different treatment groups, D-optimal design, Loewner ordering, nonlinear regression.

and Branson (2005), Thomas (2006), Dragalin, Hsuan and Padmanabhan (2007), Bornkamp et al. (2007), Thomas, Sweeney and Somayaji (2014)]. These methods can more adequately address the main questions of interest in Phase II dose-finding studies (i.e., determination of the dose-response curve and estimation of target doses of interest) than AN(C)OVA based pairwise comparisons. Moreover, it was pointed out by numerous authors that an appropriate choice of the experimental conditions can improve the statistical accuracy in dose-finding studies substantially. For this reason, there exists a large amount of literature discussing the problem of constructing optimal experimental designs for regression models, which are commonly used to describe the dose relationships [see Dragalin, Hsuan and Padmanabhan (2007), Dette et al. (2008), Dragalin, Fedorov and Wu (2008), Fang and Hedayat (2008), Gilbert (2010), or Bretz, Dette and Pinheiro (2010) among many others].

For many compounds, a question closely related to "dose", the amount of drug, is the administration frequency of the drug. In most situations, it is not adequate to assume that the same amount of drug per time unit (e.g., total daily dose) administered at different dosing intervals (e.g., once a day or twice a day) will lead to the same pharmacological effect. For example, for once a day administration the drug exposure inside the body will generally be higher just after administration and lower just before the next administration, compared to a twice a day administration, where the same amount of drug is split into two doses in the morning and the evening, leading to more uniform drug exposure over the day.

These considerations often lead to the need of evaluating the question of finding the right dose as well as dosing frequency in dose-finding studies in Phase II. One way of modeling the dose-response curves in the different treatment groups is to estimate the dose response curve corresponding to each of them separately. This can, however, be wasteful as certain aspects of the dose-response curves for different groups can be similar for both groups, suggesting a borrowing of strength. When dose-response modeling is done in terms of parametric dose-response models, one can often assume that certain parameters of the dose-response curves for the two (or more) groups are shared, while other parameters might be assumed to be different between the curves. For example, if the Emax function

$$f(d,\theta_0^{(i)},\theta_1^{(i)},\theta_2^{(i)}) = \theta_0^{(i)} + \frac{\theta_1^{(i)} d}{\theta_2^{(i)} + d}, \qquad i = 1,2 \tag{1.1}$$

is used to model the dose response relationship for both groups [see Gabrielsson and Weiner (2007) or Thomas, Sweeney and Somayaji (2014)], it is often reasonable to assume that the placebo effect is the same between groups, that is, $\theta_0^{(1)} = \theta_0^{(2)} = \vartheta_{11}$. In some situations, it might also make sense to assume that the maximum efficacy for high doses is similar, that is, $\theta_1^{(1)} = \theta_1^{(2)} = \vartheta_{12}$, as a biological maximum attainable effect might exist. However, it might not be adequate to

assume that the dose providing half of the maximum efficacy is the same for different treatment frequencies, which means $\theta_2^{(1)} \neq \theta_2^{(2)}$. The common parameters can then be estimated more precisely allowing for a more accurate statistical analysis. An example motivating the research of this paper can be found in Section 5.

The major question when planning such a dose-finding study then is which doses to utilize in the different treatment groups and how to split the total sample size between the groups. Statistically, this corresponds to the construction of optimal designs for different regression models (modeling the effect of the drug in the different groups) which share some common parameters. To our best knowledge, design problems in this case have not been considered in the literature, and the goal of the present paper is to derive optimal designs for such situations. In Section 2, the model (in the context of M treatment groups) is introduced and the main differences between the situation considered in the paper and the common optimal design problems are explained. In Section 3, we derive some results on the comparison of different designs for regression models with common parameters with respect to the Loewner ordering. In particular, we generalize recent results of admissible designs as presented in Yang (2010), Dette and Melas (2011) and Yang and Stufken (2012) and derive upper bounds on the number of support points which cannot be improved upon in the Loewner ordering. Section 4 is devoted to the construction of D-optimal designs which are well suited for a "global" inference as they minimize the maximum confidence interval length around the predicted dose-response curve. Explicit expressions for locally D-optimal designs for the commonly used dose response models are derived, if some parameters of the models for the different groups coincide. We also discuss minimally supported optimal designs and investigate if these designs are optimal within the class of all designs. In Section 5, we illustrate the developed methods in a particular clinical dose-finding study investigating two different treatment groups. Some technical details are given in the Appendix while the Supplement (Feller et al., 2017) provides further proofs and some more background on the modeling problem discussed in Section 5.

For the sake of brevity and transparency, most parts of this paper consider locally optimal designs which require a-priori information about the unknown model parameters if the models are nonlinear [see Chernoff (1953)]. In several situations, preliminary knowledge regarding the unknown parameters of a nonlinear model is available but not in a form that is accurate enough to specify one parameter guess. As illustrated in Section 5, locally optimal designs can be used as benchmarks for commonly used designs and also serve as basis for constructing optimal designs with respect to more sophisticated optimality criteria, which are robust against a misspecification of the unknown parameters (and model) [see Pronzato and Walter (1985) or Chaloner and Verdinelli (1995), Dette (1997) among others]. Following this line of research the methodology introduced in the present paper can be further developed to address uncertainty in the preliminary information on the unknown parameters, and we will illustrate this approach in Section 5, where we also discuss robust designs for the data example under consideration.

2. Models with common parameters. Consider regression models

$$Y_{ij\ell} = f\big(d_j^{(i)}, \theta_1, \theta_2^{(i)}\big) + \varepsilon_{ij\ell} \tag{2.1}$$

($\ell = 1, \dots, n_{ij}$, $j = 1, \dots, k_i$, $i = 1, \dots, M$), where $\varepsilon_{ij\ell}$ are independent centered normally distributed random variables, that is, $\varepsilon_{ij\ell} \sim \mathcal{N}(0, \sigma_i^2)$. The assumption of a normal distribution in (2.1) is made for the sake of transparency. Other distributional assumptions can be treated exactly in the same way. This means that M different groups are investigated and in each group observations are taken at different experimental conditions $d_1^{(i)}, \dots, d_{k_i}^{(i)}$, which vary in possibly different design spaces, say $\mathcal{X}_i = [0, d_{\max}^{(i)}]$ ($i = 1, \dots, M$). At each dose level $d_j^{(i)}$, the experimenter can take n_{ij} observations and $n_i = \sum_{j=1}^{k_i} n_{ij}$ denotes the number of observations in the ith group ($i = 1, \dots, M$). Moreover, the total sample size is given by $n = \sum_{i=1}^{M} n_i$. In general, the regression model $f(\cdot, \theta_1, \theta_2^{(i)})$ with a $(p+q)$-dimensional parameter vector $\theta^{(i)} = (\theta_1^T, (\theta_2^{(i)})^T)^T$ is used to describe the dependency between the response and the effect in every group. We consider the same parametric form for all groups. Moreover, the parameter vector $\theta_1 \in \mathbb{R}^p$ is assumed to be the same in all groups ($i = 1, \dots, M$), while $\theta_2^{(i)} \in \mathbb{R}^q$ is different for different groups. Consequently, the vector of unknown parameters is given by $\theta = (\theta_1^T, (\theta_2^{(1)})^T, \dots, (\theta_2^{(M)})^T)^T \in \mathbb{R}^m$, where $m = p + qM$. The components of the vector are denoted by $\theta_1 = (\vartheta_1, \dots, \vartheta_p)^T$ and $\theta_2^{(i)} = (\vartheta_1^{(i)}, \dots, \vartheta_q^{(i)})^T$ ($i = 1, \dots, M$).

Following Kiefer (1974), we define for $i = 1, \dots, M$ approximate designs ξ_i (on the design space $\mathcal{X}_i$) as probability measures with masses ξ_{ij} at the experimental conditions $d_j^{(i)} \in \mathcal{X}_i$ ($j = 1, \dots, k_i$) and a design μ as a probability measure on the set $\{1, \dots, M\}$ assigning mass λ_i to the ith group. We collect these designs in the vector $\xi = (\xi_1, \dots, \xi_M, \mu)$, which is also called design (on the design space $\mathcal{X}_1 \times \cdots \times \mathcal{X}_M \times \{1, \dots, M\}$) throughout this paper. If an approximate design $\xi = (\xi_1, \dots, \xi_M, \mu)$ is given and N observations can be taken, a rounding procedure is applied to obtain integers n_i and n_{ij} ($i = 1, \dots M$, $j = 1, \dots, k_i$) from the not necessarily integer valued quantities $\lambda_i n$ and $\xi_{ij} n_i$, respectively. Then, under common assumptions of regularity and the assumption

$$\lim_{n_i \to \infty} \frac{n_{ij}}{n_i} = \xi_{ij} \in (0,1) \quad \text{and} \quad \lim_{n \to \infty} \frac{n_i}{n} = \lambda_i \in (0,1) \tag{2.2}$$

($i = 1, \dots, M$, $j = 1, \dots, k_i$), the maximum likelihood estimate $\hat{\theta}^T = (\hat{\theta}_1, \hat{\theta}_2^{(1)}, \dots, \hat{\theta}_2^{(M)})$ satisfies $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}(0, M^{-1}(\xi, \theta))$ as $n \to \infty$, where the symbol $\xrightarrow{\mathcal{D}}$ denotes weak convergence. Here, the matrix

$$M(\xi, \theta) = \int \int_{\mathcal{X}_z} h_z(d) h_z^T(d)\, d\xi_z(d)\, d\mu(z) = \sum_{i=1}^{M} \lambda_i M^{(i)}(\xi_i, \theta) \tag{2.3}$$

is called the information matrix of the design $\xi = (\xi_1, \dots, \xi_M, \mu)$ and will be derived in the Supplement (Feller et al., 2017). In (2.3), the matrices $M^{(i)}$ are defined by

$$M^{(i)}(\xi_i, \theta) = \int_{\mathcal{X}_i} h_i(d) h_i^T(d)\, d\xi_i(d) \tag{2.4}$$

$(i = 1, \dots, M)$ and

$$h_i(d) = \frac{1}{\sigma_i}\left(\left(\frac{\partial}{\partial\theta_1} f\big(d, \theta_1, \theta_2^{(i)}\big)\right)^T, 0_{q(i-1)}^T, \left(\frac{\partial}{\partial\theta_2^{(i)}} f\big(d, \theta_1, \theta_2^{(i)}\big)\right)^T, 0_{q(M-i)}^T\right)^T \in \mathbb{R}^m \tag{2.5}$$

is the gradient of the function $f(d, \theta_1, \theta_2^{(i)})$ with respect to the parameter $\theta \in \mathbb{R}^m$, where $m = p + qM$, $0_{q(i-1)} \in \mathbb{R}^{q(i-1)}$ and $0_{q(M-i)} \in \mathbb{R}^{q(M-i)}$ denote vectors with all entries equal to 0.

EXAMPLE 2.1. We assume that $M = 2$ and that the regression functions $f(\cdot, \theta_1, \theta_2^{(i)})$ can be written as

$$f\big(\cdot, \theta_1, \theta_2^{(i)}\big) = \vartheta_1 + \vartheta_2 f_0\big(\cdot, \theta_2^{(i)}\big) \tag{2.6}$$

with a given function f_0 [see Bretz, Pinheiro and Branson (2005)]. Here, the location and scale parameters $\theta_1 = (\vartheta_1, \vartheta_2)^T \in \mathbb{R}^2$ are the same for all groups, while the parameters $\theta_2^{(i)} \in \mathbb{R}^q$ are different. In this case, we have $p = 2$ and the vectors $h_1(d)$ and $h_2(d)$ are given by

$$h_1^T(d) = \frac{1}{\sigma_1}\left(1, f_0\big(d, \theta_2^{(1)}\big), \left(\frac{\partial}{\partial\theta_2^{(1)}} f_0\big(d, \theta_2^{(1)}\big)\right)^T, 0_q^T\right),$$

$$h_2^T(d) = \frac{1}{\sigma_2}\left(1, f_0\big(d, \theta_2^{(2)}\big), 0_q^T, \left(\frac{\partial}{\partial\theta_2^{(2)}} f_0\big(d, \theta_2^{(2)}\big)\right)^T\right).$$

As a further example, consider a regression function $f(\cdot, \theta_1, \theta_2^{(i)})$ of the form

$$f\big(\cdot, \theta_1, \theta_2^{(i)}\big) = \theta_1 + f_0\big(\cdot, \theta_2^{(i)}\big); \qquad i = 1, 2, \tag{2.7}$$

with a given function f_0. If the location parameter θ_1 is the same for the two groups and the parameters $\theta_2^{(i)} \in \mathbb{R}^q$ are different, we have $p = 1$ and the vectors $h_1^T(d)$ and $h_2^T(d)$ are given by $h_1^T(d) = \frac{1}{\sigma_1}(1, (\frac{\partial}{\partial\theta_2^{(1)}} f_0(d, \theta_2^{(1)}))^T, 0_q^T)$ and $h_2^T(d) = \frac{1}{\sigma_2}(1, 0_q^T, (\frac{\partial}{\partial\theta_2^{(2)}} f_0(d, \theta_2^{(2)}))^T)$.

3. Comparing designs in the Loewner ordering. An optimal design $\xi = (\xi_1, \ldots, \xi_M, \mu)$ maximizes a concave real valued function, say Φ, of the information matrix. Numerous criteria have been proposed in the literature [see Pukelsheim (2006) among others] which can be used to discriminate between competing designs and the particular case of D-optimality will be discussed in the subsequent section. The commonly used optimality criteria are monotone with respect to the Loewner ordering, that is the relation $M(\xi_1, \theta) \le M(\xi_2, \theta)$ implies $\Phi(M(\xi_1, \theta)) \le \Phi(M(\xi_2, \theta))$. For this reason we discuss at first some results for this ordering, which will be very helpful for the explicit determination of optimal designs in the following sections.

Throughout this paper, let $|\mathcal{A}|$ denote the cardinality of a set $\mathcal{A}$ and we denote by $\operatorname{supp}(\xi)$ the support of the design $\xi = (\xi_1, \ldots, \xi_M, \mu)$. Moreover, we define the index $I(\xi_i)$ of the design ξ_i on the interval $[0, d_{\max}^{(i)}]$ as the number of support points, where the boundary points 0 and $d_{\max}^{(i)}$ are only counted by $1/2$ if they are support points of the design ξ_i $(i = 1, \ldots, M)$.

Note that the gradient (2.5) can be rewritten in the form

$$\begin{aligned} h_i(d) &= \frac{1}{\sigma_i} \begin{pmatrix} I_{p\times p} & 0_{p\times q} \\ 0_{(i-1)q\times p} & 0_{(i-1)q\times q} \\ 0_{q\times p} & I_{q\times q} \\ 0_{(M-i)q\times p} & 0_{(M-i)q\times p} \end{pmatrix} \begin{pmatrix} \dfrac{\partial}{\partial\theta_1} f\big(d, \theta_1, \theta_2^{(i)}\big) \\ \dfrac{\partial}{\partial\theta_2^{(i)}} f\big(d, \theta_1, \theta_2^{(i)}\big) \end{pmatrix} \\ &:= P_i g\big(d, \theta^{(i)}\big), \end{aligned} \tag{3.1}$$

where P_i is a $(p + Mq) \times (p + q)$ block matrix, $I_{p\times p}$ is the p-dimensional identity matrix and $g(d, \theta^{(i)})$ is the $(p + q)$-dimensional gradient of $f(d, \theta_1, \theta_2^{(i)})$ with respect to $\theta^{(i)} = (\theta_1^T, (\theta_2^{(i)})^T)^T$ $(i = 1, \ldots, M)$. Consequently, for the information matrix (2.4) the representation

$$M^{(i)}(\xi_i, \theta) = P_i \int_{\mathcal{X}_i} g\big(d, \theta^{(i)}\big) g^T\big(d, \theta^{(i)}\big)\, d\xi_i(d)\, P_i^T := P_i C\big(\xi_i, \theta^{(i)}\big) P_i^T$$

holds, where the $(p + q) \times (p + q)$ matrix $C(\xi_i, \theta^{(i)})$ is given by

$$C\big(\xi_i, \theta^{(i)}\big) = \int_{\mathcal{X}_i} \begin{pmatrix} \psi_{1,1}\big(d, \theta^{(i)}\big) & \ldots & \psi_{1,p+q}\big(d, \theta^{(i)}\big) \\ \vdots & \ddots & \vdots \\ \psi_{p+q,1}\big(d, \theta^{(i)}\big) & \ldots & \psi_{p+q,p+q}\big(d, \theta^{(i)}\big) \end{pmatrix} d\xi_i(d)$$

$(i = 1, \ldots, M)$ and the functions $\psi_{s,t}(d, \theta^{(i)})$ are the entries of the matrix $g(d, \theta^{(i)}) g^T(d, \theta^{(i)})$.

In the following, we will present a generalization of the results in Yang (2010), Dette and Melas (2011) and Yang and Stufken (2012) for the situation where $M \ge 1$ groups are considered. To be precise for $i = 1, \ldots, M$, we define $\Psi_0(d) \equiv 1$ and choose a basis, say $\{\Psi_0(\cdot), \Psi_1^i(\cdot), \ldots, \Psi_{k-1}^i(\cdot), \Psi_k^i(\cdot)\}$ for the space

$$\mathcal{L} = \operatorname{span}\big(\big\{\psi_{s,t}\big(\cdot, \theta^{(i)}\big) \big| 1 \le s, t \le p + q\big\} \cup \{1\}\big),$$

where the dependence on the parameters is reflected by the upper index i for the sake of a transparent notation. Note that k is the dimension of $\mathcal{L}$ and depends on the considered regression model. We further assume that the function $\Psi_k^i(\cdot)$ is a diagonal element of the matrix $C(\xi_i, \theta^{(i)})$, does not coincide with any of the other elements $\psi_{s,t}(\cdot, \theta^{(i)})$ and that $\{\Psi_0(\cdot), \Psi_1^i(\cdot), \ldots, \Psi_{k-1}^i(\cdot)\}$ is a basis of the space

$$\tilde{\mathcal{L}} = \operatorname{span}\big(\{\psi_{s,t}(\cdot, \theta^{(i)}) \mid s, t \in \{1, \ldots, p+q\}; \psi_{s,t}(\cdot, \theta^{(i)}) \neq \Psi_k^i(\cdot)\} \cup \{1\}\big).$$

Note again that the spaces $\mathcal{L}$ and $\tilde{\mathcal{L}}$ determine the value k, which is basically a constant determined by the regression model.

For our first results, we require the notation of a Chebyshev system [see Karlin and Studden (1966)]. A set of k real valued functions $\Psi_0, \ldots, \Psi_{k-1} : [A, B] \to \mathbb{R}$ is called *Chebychev system* on the interval $[A, B]$ if and only if it fulfills the inequality

$$\det\big(\Psi_i(x_j)\big)_{i,j=0,\ldots,k-1} > 0$$

for any points $x_0, \ldots, x_{k-1}$ with $A \leq x_0 < x_1 < \cdots < x_{k-1} \leq B$.

LEMMA 3.1. (1) *If* $\{\Psi_0(\cdot), \Psi_1^i(\cdot), \ldots, \Psi_{k-1}^i(\cdot)\}$ *and* $\{\Psi_0(\cdot), \Psi_1^i(\cdot), \ldots, \Psi_{k-1}^i(\cdot), \Psi_k^i(\cdot)\}$ *are Chebychev systems on the interval* $\mathcal{X}_i = [0, d_{\max}^{(i)}]$ *(for all* $i = 1, \ldots, M$*), then for any design* ξ *there exists a design* $\xi^+ = (\xi_1^+, \ldots, \xi_M^+, \mu)$ *such that* $|\operatorname{supp}(\xi_i^+)| \leq \frac{k+2}{2}$ $(i = 1, \ldots, M)$ *and* $M(\xi^+, \theta) \geq M(\xi, \theta)$. *If the index of the design* ξ_i *satisfies* $I(\xi_i) < \frac{k}{2}$ *the design coincides with the design* ξ. *In the case* $I(\xi_i) \geq \frac{k}{2}$, *the following two assertions are valid*:

(1a) *If* k *is odd, then* ξ_i^+ *has at most* $\frac{k+1}{2}$ *support points and* ξ_i^+ *can be chosen such that its support contains* $d_{\max}^{(i)}$ $(i = 1, \ldots, M)$.

(1b) *If* k *is even, then* ξ_i^+ *has at most* $\frac{k+2}{2}$ *support points and* ξ_i^+ *can be chosen such that its support contains the points* 0 *and* $d_{\max}^{(i)}$ $(i = 1, \ldots, M)$.

(2) *If* $\{\Psi_0(\cdot), \Psi_1^i(\cdot), \ldots, \Psi_{k-1}^i(\cdot)\}$ *and* $\{\Psi_0(\cdot), \Psi_1^i(\cdot), \ldots, \Psi_{k-1}^i(\cdot), -\Psi_k^i(\cdot)\}$ *are Chebychev systems on the interval* $\mathcal{X}_i = [0, d_{\max}^{(i)}]$ *(for all* $i = 1, \ldots, M$*), then for any design* ξ *there exists a design* $\xi^- = (\xi_1^-, \ldots, \xi_M^-, \mu)$ *such that* $|\operatorname{supp}(\xi_i^-)| \leq \frac{k+2}{2}$ $(i = 1, \ldots, M)$ *and* $M(\xi^-, \theta) \geq M(\xi, \theta)$. *If the index of the design* ξ_i *satisfies* $I(\xi_i) < \frac{k}{2}$, *the design coincides with the design* ξ. *In the case* $I(\xi_i) \geq \frac{k}{2}$, *the following two assertions are valid.*

(2a) *If* k *is odd, then* ξ_i^- *has at most* $\frac{k+1}{2}$ *support points and* ξ_i^- *can be chosen such that its support contains* 0.

(2b) *If* k *is even, then* ξ_i^- *has at most* $\frac{k}{2}$ *support points*.

Lemma 3.1 provides an upper bound for the maximal number of support points if $\{\Psi_0(\cdot), \Psi_1^i(\cdot), \ldots, \Psi_{k-1}^i(\cdot)\}$ and $\{\Psi_0(\cdot), \Psi_1^i(\cdot), \ldots, \Psi_{k-1}^i(\cdot), \Psi_k^i(\cdot)\}$ are Chebychev systems for the different groups $i = 1, \ldots, M$. Note that this bound is the

same independently from the dimension of the common parameter vector θ_1, since the number of support points is bounded in every group $1, \dots, M$ separately. For the case $M = 1$, we obtain the classic results of Yang (2010), Dette and Melas (2011) and Yang and Stufken (2012) by setting $\mu = \delta_1$, where δ_t is the dirac measure in t.

Note that Lemma 3.1 refers to the principal representations of the elements of the moment space $\{\int_{\mathcal{X}}(1, \Psi_1^{(i)}(d), \dots, \Psi_k^{(i)}(d))\, d\xi_i(d) | \xi_i \text{ measure on } \mathcal{X}_i\}$ [see Karlin and Studden (1966), pages 51 and 53]. For the case $M = 1$, Dette and Melas (2011) proved that for every design ξ with more than $(k+2)/2$ support points there exists a design ξ^+ with at most $(k+2)/2$ support points and $M(\xi^+, \theta) \geq M(\xi, \theta)$. The support points of the design ξ^+ are given either by the lower or the upper principal representation of the element $\int_{\mathcal{X}}(1, \Psi_1(d), \dots, \Psi_k(d))\, d\xi(d)$ depending both on k and on the conditions given in Lemma 3.1. The second part of Lemma 3.1 refers to the lower principal representations and if k is even this lower principal representation neither includes the lower boundary nor the upper boundary point of $\mathcal{X}$.

The next lemma shows that (for the commonly used dose response models) it is sufficient to allocate only patients from the group with the smallest population variance to placebo.

LEMMA 3.2. *Assume that the design spaces are given by* $\mathcal{X}_i = [0, d_{\max}^{(i)}]$ $(i = 1, \dots, M)$ *and that the regression models are given by* (2.6) *or by* (2.7), *where the function* f_0 *is differentiable with respect to* θ_2 $(i = 1, \dots, M)$. *Moreover, assume that* $f_0(0, \theta_2) = 0$ *and* $\frac{\partial}{\partial \theta_2} f_0(0, \theta_2) = 0_q$.

If $\eta = (\eta_1, \dots, \eta_M, \nu)$ *denotes a design with* $0 \in \operatorname{supp}(\eta_j)$ *for (at least) one index* j, *then there exists a design* $\xi = (\xi_1, \dots, \xi_M, \mu)$ *such that* $M(\eta, \theta) \leq M(\xi, \theta)$ *and* ξ *has the following properties*:

$$0 \in \operatorname{supp}(\xi_{j^*}), \qquad 0 \notin \operatorname{supp}(\xi_j) \qquad \text{for all } j \neq j^*,\ j^* \in \arg\min_{i=1,\dots,M} \sigma_i^2.$$

Note that Lemma 3.2 remains correct in the case of more than one group with minimal variance. Moreover, if for example two groups have minimal variances one can improve the design with respect to the Loewner ordering by shifting the placebo dose 0 to each of these two groups (for details, see the proof of Lemma 3.2).

In the following discussion, we will apply the previous results to some of the commonly used dose response models, namely the Emax model, linear-in-log and exponential model [see Gabrielsson and Weiner (2007)], which are listed in Table 1. In this table, we also illustrate our notation again. The left part of the table corresponds to a model with a common location parameter (namely θ_1), while the right part of the table shows a model with a common location (ϑ_1) and scale parameter (ϑ_2). We note that all these models satisfy the conditions of Lemma 3.1 and Lemma 3.2.

TABLE 1
Commonly used dose response models for $i = 1, \dots, M$. Left column: The placebo effect is the same in every group (common location). Right column: Both the placebo effect and the scale parameter coincide in every group (common location and scale)

		location		location and scale
Emax	(3.2)	$\theta_1 + \vartheta_1^{(i)} \frac{d}{\vartheta_2^{(i)} + d}$	(3.5)	$\vartheta_1 + \vartheta_2 \frac{d}{\theta_2^{(i)} + d}$
Linear-in-log	(3.3)	$\theta_1 + \vartheta_1^{(i)} \log(\frac{d}{\vartheta_2^{(i)}} + 1)$	(3.6)	$\vartheta_1 + \vartheta_2 \log(\frac{d}{\theta_2^{(i)}} + 1)$
Exponential	(3.4)	$\theta_1 + \vartheta_1^{(i)} (\exp(\frac{d}{\vartheta_2^{(i)}}) - 1)$	(3.7)	$\vartheta_1 + \vartheta_2 (\exp(\frac{d}{\theta_2^{(i)}}) - 1)$

COROLLARY 3.3. *Let $\xi = (\xi_1, \dots, \xi_M, \mu)$ denote an arbitrary design with $|\operatorname{supp}(\xi_i)| \geq 3$ $(i = 1, \dots, M)$ and assume (w.l.o.g) that $\sigma_1^2 = \min_{i=1,\dots,M} \sigma_i^2$.*

(1) *If the regression model is given by the Emax model (3.2) or (3.5), then there exists a design $\xi^+ = (\xi_1^+, \dots, \xi_M^+, \mu)$ with at most $2M + 1$ support points such that $M(\xi^+, \theta) \geq M(\xi, \theta)$. Moreover, ξ^+ can be chosen such that $|\operatorname{supp}(\xi_1^+)| = 3$ with $0, d_{\max}^{(1)} \in \operatorname{supp}(\xi_1^+)$ and $|\operatorname{supp}(\xi_i^+)| = 2$ with $d_{\max}^{(i)} \in \operatorname{supp}(\xi_i^+)$ $(i = 2, \dots, M)$.*

(2) *If the regression model is given by the linear-in-log model (3.3) or (3.6), then there exists a design $\xi^+ = (\xi_1^+, \dots, \xi_M^+, \mu)$ with at most $2M + 1$ support points such that $M(\xi^+, \theta) \geq M(\xi, \theta)$. Moreover, ξ^+ can be chosen such that $|\operatorname{supp}(\xi_1^+)| = 3$ with $0, d_{\max}^{(1)} \in \operatorname{supp}(\xi_1^+)$ and $|\operatorname{supp}(\xi_i^+)| = 2$ with $d_{\max}^{(i)} \in \operatorname{supp}(\xi_i^+)$ $(i = 2, \dots, M)$.*

(3) *If the regression model is given by the exponential model (3.4) or (3.7), then there exists a design $\xi^+ = (\xi_1^+, \dots, \xi_M^+, \mu)$ with at most $3M$ support points such that $M(\xi^+, \theta) \geq M(\xi, \theta)$. Moreover, ξ_i^+ can be chosen such that $|\operatorname{supp}(\xi_i^+)| = 3$ and $d_{\max}^{(i)} \in \operatorname{supp}(\xi_i^+)$ $(i = 1, \dots, M)$.*

4. *D*-optimal designs. When one of the major purposes of the study is to determine the dose-response curve, D-optimal designs are well suited as they minimize the maximum confidence interval length around the predicted dose-response curve [see Silvey (1980)]. Following Chernoff (1953), a design $\xi = (\xi_1, \dots, \xi_M, \mu)$ is called (locally) D-optimal for the information matrix given in (2.3) if it maximizes the determinant of the information matrix $\det(M(\xi, \theta))$ in the class of all designs ξ on $\mathcal{X}_1 \times \cdots \times \mathcal{X}_M \times \{1, \dots, M\}$. A main tool of optimal design theory are equivalence theorems which, on one hand provide a simple checking condition for the optimality of a given design, and on other hand, are the basis of many procedures for their numerical construction. Moreover, these characterizations of optimality can also be used to derive structural properties of

optimal designs. The following result provides the equivalence theorem for the D-optimality criterion corresponding to the matrix given in (2.3). The proof can be found in the Supplement (Feller et al., 2017).

THEOREM 4.1. *The design $\xi^\star = (\xi_1^\star, \dots, \xi_M^\star, \mu^\star)$ is D-optimal if and only if the M inequalities*

$$\kappa_i(d, \xi^\star, \theta) = h_i^T(d) M^{-1}(\xi^\star, \theta) h_i(d) \le m = p + qM, \tag{4.1}$$

are satisfied for all $d \in \mathcal{X}_i$, $i = 1, \dots, M$. Equality holds in (4.1) for any points $(d_1, \dots, d_M, z) \in \operatorname{supp}(\xi_1^\star) \times \cdots \times \operatorname{supp}(\xi_M^\star) \times \operatorname{supp}(\mu^\star)$.

Denote

$$\Xi_m^M = \left\{ \xi = (\xi_1, \dots, \xi_M, \mu) \;\Big|\; \sum_{i=1}^M |\operatorname{supp}(\xi_i)| = m \right\} \tag{4.2}$$

as the set of all designs on $\mathcal{X}_1 \times \cdots \times \mathcal{X}_M \times \{1, \dots, M\}$ with exactly m different dose levels in the M groups. The proof of the next lemma follows by similar arguments as in the standard case [see Silvey (1980) among others], and can be found in the Supplement (Feller et al., 2017).

LEMMA 4.2. *Let $\xi = (\xi_1, \dots, \xi_M, \mu) \in \Xi_m^M$ denote a design on $\mathcal{X}_1 \times \cdots \times \mathcal{X}_M \times \{1, \dots, M\}$ and m_i denote the number of support points of ξ_i $(i = 1, \dots, M)$. Assume that the $m = \sum_{i=1}^M m_i$ vectors $h_1(d_1^{(1)}), \dots, h_1(d_{m_1}^{(1)}), \dots, h_M(d_1^{(M)}), \dots, h_M(d_{m_M}^{(M)})$ are linearly independent where $d_j^{(i)} \in \operatorname{supp}(\xi_i)$, $j = 1, \dots, m_i$, $i = 1, \dots, M$.*

If ξ is locally D-optimal in the class Ξ_m^M, then each component ξ_i has equal weights at its support points. Moreover, the weights of μ at the points $1, \dots, M$ are given by $\frac{m_1}{m}, \dots, \frac{m_M}{m}$, respectively.

In the following two sections, we present some locally D-optimal designs for the Emax, the exponential and the linear-in-log model. The proofs of these results are complicated and, therefore, deferred either to the Appendix or the Supplement (Feller et al., 2017).

4.1. *Models with the same location parameter.* First, we consider the case where only the location parameter is the same in the different models. In applications, this reflects the situation of a common placebo effect for all groups (cf. the first column of Table 1), and we are able to identify the locally D-optimal design explicitly. We begin with a general result for regression functions of the form (2.7) where the unknown parameter vectors are given by $\theta = (\theta_1, (\theta_2^{(1)})^T, \dots, (\theta_2^{(M)})^T)^T \in \mathbb{R}^m$ with $m = 1 + Mq$. The following result provides a solution of the D-optimal design problem if the D-optimal design for the single models are known.

THEOREM 4.3. *Let* $\sigma_1^2 = \min_{i=1,\dots,M} \sigma_i^2$ *and consider the model given by* (2.7), *which satisfies*

$$f_0(0, \theta_2^{(i)}) = 0, \qquad \eta_0(0, \theta_2^{(i)}) = \frac{\partial}{\partial \theta_2^{(i)}} f_0(d, \theta_2^{(i)}) \Big|_{d=0} = 0_q \tag{4.3}$$

$(i = 1, \dots, M)$. *If the design*

$$\tilde{\xi}^{(i)} = \begin{pmatrix} 0 & d_1^{(i)} & \dots & d_q^{(i)} \\ 1 & 1 & & 1 \\ \overline{q+1} & \overline{q+1} & \dots & \overline{q+1} \end{pmatrix} \tag{4.4}$$

is locally D-optimal for the single model $f(d, \theta_1, \theta_2^{(i)})$ $(i = 1, \dots, M)$, *then the locally D-optimal design for model* (2.7) *is given by* $\xi^\star = (\xi_1^\star, \dots, \xi_M^\star, \mu^\star)$ *where*

$$\begin{aligned} \xi_1^\star &= \tilde{\xi}^{(1)}, \qquad \xi_i^\star = \begin{pmatrix} d_1^{(i)} & \dots & d_q^{(i)} \\ 1 & & 1 \\ \overline{q} & \dots & \overline{q} \end{pmatrix}, \qquad i = 2, \dots, M, \\ \mu^\star &= \begin{pmatrix} 1 & 2 & \dots & M \\ \overline{q+1} & \overline{q} & \dots & \overline{q} \\ m & m & & m \end{pmatrix}. \end{aligned} \tag{4.5}$$

Using Theorem 4.3, the placebo effect θ_1 is estimated in the group where the variance is smallest (see also Lemma 3.2 and Corollary 3.3). Moreover, it follows from the proof of Lemma 3.2 that the D-optimal design given by Theorem 4.3 is not unique if there exist more than one group with minimal variance. For example, if both the first and the second group have minimal variance ($\sigma_1^2 = \sigma_2^2$), then another D-optimal design [besides the one stated in (4.5)] is given by $\xi_2^\star = \tilde{\xi}^{(2)}$,

$$\begin{aligned} \mu^\star &= \begin{pmatrix} 1 & 2 & \dots & M \\ \overline{q} & \overline{q+1} & \dots & \overline{q} \\ m & m & & m \end{pmatrix}, \\ \xi_i^\star &= \begin{pmatrix} d_1^{(i)} & \dots & d_q^{(i)} \\ 1 & & 1 \\ \overline{q} & \dots & \overline{q} \end{pmatrix}, \qquad i = 1, 3, \dots, M. \end{aligned}$$

We now use Theorem 4.3 to determine D-optimal designs for the Emax, exponential and linear-in-log model explicitly.

COROLLARY 4.4. *Let* $\sigma_1^2 = \min_{i=1,\dots,M} \sigma_i^2$. *The locally D-optimal design for the Emax, exponential and linear-in-log model* (3.2) *is of the form* $\xi^\star = (\xi_1^\star, \dots, \xi_M^\star, \mu^\star)$, *where*

$$\xi_1^\star = \begin{pmatrix} 0 & x^{\star,(1)} & d_{\max}^{(1)} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}, \qquad \xi_i^\star = \begin{pmatrix} x^{\star,(i)} & d_{\max}^{(i)} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}, \qquad i = 2, \dots, M,$$

$$\mu^\star = \begin{pmatrix} 1 & 2 & \dots & M \\ \frac{3}{m} & \frac{2}{m} & \dots & \frac{2}{m} \end{pmatrix}$$

and the point $x^{\star,(i)}$ *is given by*

$$x^{\star,(i)} = x^{\star,(i)}_{\text{emax}} = \frac{\vartheta_2^{(i)} d^{(i)}_{\max}}{d^{(i)}_{\max} + 2\vartheta_2^{(i)}}, \qquad i = 1, \dots, M \tag{4.6}$$

for the Emax model, by

$$x^{\star,(i)} = x^{\star,(i)}_{\text{exp}} = \frac{(d^{(i)}_{\max} - \vartheta_2^{(i)}) \exp(d^{(i)}_{\max}/\vartheta_2^{(i)}) + \vartheta_2^{(i)}}{\exp(d^{(i)}_{\max}/\vartheta_2^{(i)}) - 1}, \qquad i = 1, \dots, M \tag{4.7}$$

for the exponential model and by

$$x^{\star,(i)} = x^{\star,(i)}_{\log} = \frac{(d^{(i)}_{\max} + \vartheta_2^{(i)})\vartheta_2^{(i)} \log(d^{(i)}_{\max}/\vartheta_2^{(i)} + 1) - \vartheta_2^{(i)} d^{(i)}_{\max}}{d^{(i)}_{\max}}, \qquad i = 1, \dots, M \tag{4.8}$$

for the linear-in-log model.

It is worthwhile to mention that the locally D-optimal design for model (2.7) with an Emax curve consists of the designs which are locally D-optimal for the models given by an individual Emax model with parameter $\theta^{(1)} = (\theta_1, \theta_2^{(1)})^T$ and by an Emax model with location parameter equal to zero and parameter $\theta_2^{(i)}$, $i = 2, \dots, M$. This effect can also be observed for the exponential and the linear-in-log model.

4.2. *Models with the same location and scale parameters.* In this section, we consider model (2.6) and assume that the location and scale parameter coincide across the different models (cf. the second column in Table 1). It turns out that in this case the D-optimal design problem is substantially harder, and for the sake of a transparent presentation, we restrict ourselves to the case of $M = 2$ groups. Similar results can be obtained in the case $M > 2$ with an additional amount of notation. We begin with some general properties of locally D-optimal designs for the model (2.6) in the case of an Emax, linear-in-log and exponential curve. For this purpose, we define $r = \frac{\sigma_1^2}{\sigma_2^2}$ as the ratio of the two population variances.

LEMMA 4.5.

(A) *The locally D-optimal design* $\xi^\star = (\xi_1^\star, \xi_2^\star, \mu)$ *for the Emax model* (3.5) *and the linear-in-log* (3.6) *have the following properties*:

(A1) $|\operatorname{supp}(\xi_1^\star)| + |\operatorname{supp}(\xi_2^\star)| \in \{4, 5\}$.

(A2) *If* $|\operatorname{supp}(\xi_1^\star)| + |\operatorname{supp}(\xi_2^\star)| = 5$, *then* $d_{\max}^{(i)} \in \operatorname{supp}(\xi_i^\star)$, $i = 1, 2$.

(A3) *If* $|\operatorname{supp}(\xi_1^\star)| + |\operatorname{supp}(\xi_2^\star)| = 4$, *then* $d_{\max}^{(1)} \in \operatorname{supp}(\xi_1^\star)$ *or* $d_{\max}^{(2)} \in \operatorname{supp}(\xi_2^\star)$.

(B) *The locally D-optimal design* $\xi^\star = (\xi_1^\star, \xi_2^\star, \mu)$ *for the exponential model* (3.7) *satisfies*

$$\bigl|\operatorname{supp}(\xi_1^\star)\bigr| + \bigl|\operatorname{supp}(\xi_2^\star)\bigr| \in \{4, 5, 6\}.$$

By the previous lemma, the number of support points of the locally D-optimal designs is at most 5 for the Emax and linear-in-log model and at most 6 for the exponential model. On the other hand, at least four support points are required to estimate all parameters in both models (note that the scale and location are assumed to be the same throughout this section). In the following discussion, we determine such "minimally" supported D-optimal designs explicitly for the Emax, exponential and linear-in-log model.

4.2.1. *Minimally supported designs.* Recall the definition of the set Ξ_m^M in (4.2). We call a design $\xi = (\xi_1, \xi_2, \mu)$ minimally supported (for the Emax, linear-in-log and exponential model) if $\xi \in \Xi_4^2$ (note that for these models the information matrix is of size 4×4 as the scale and location parameter coincide in both models). It turns out the minimally supported D-optimal designs for the three models under consideration have a very similar structure. On the other hand, the question, if these designs are D-optimal in the class of all designs does not have a simple answer and will be discussed in the following section.

THEOREM 4.6. *Let* $r = \frac{\sigma_1^2}{\sigma_2^2}$, $\bar\theta_2^{(i)} = \frac{\theta_2^{(i)}}{d_{\max}^{(i)}}$, $i = 1, 2$ *and* $0 < \bar\theta_2^{(1)} < \bar\theta_2^{(2)} < 1$, *define* $y^\star = \theta_2^{(2)}$, $z^\star = \theta_2^{(1)}$ *and* $x^{\star,(i)} = x_{emax}^{\star,(i)}$ *by* (4.6) $(i = 1, 2)$.

(1) *If* $r \le 1$, *the locally D-optimal design for model* (3.5) *in the class* Ξ_4^2 *is given by*

$$\xi_1^{a,\star} = \begin{pmatrix} 0 & x^{\star,(1)} & d_{\max}^{(1)} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}, \qquad \xi_2^{a,\star} = \begin{pmatrix} y^\star \\ 1 \end{pmatrix}, \qquad \mu^{a,\star} = \begin{pmatrix} 1 & 2 \\ \frac{3}{4} & \frac{1}{4} \end{pmatrix}. \tag{4.9}$$

(2) *If* $1 < r \le (\frac{1+\bar\theta_2^{(2)}}{1+\bar\theta_2^{(1)}})^6$, *the locally D-optimal design for model* (3.5) *in the class* Ξ_4^2 *is given by*

$$\xi_1^{b,\star} = \begin{pmatrix} x^{\star,(1)} & d_{\max}^{(1)} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}, \qquad \xi_2^{b,\star} = \begin{pmatrix} 0 & y^\star \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}, \qquad \mu^{b,\star} = \begin{pmatrix} 1 & 2 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}. \tag{4.10}$$

(3) *If* $r > (\frac{1+\bar{\theta}_2^{(2)}}{1+\bar{\theta}_2^{(1)}})^6$, *the locally D-optimal design for model* (3.5) *in the class* Ξ_4^2 *is given by*

$$\xi_1^{c,\star} = \begin{pmatrix} z^\star \\ 1 \end{pmatrix}, \qquad \xi_2^{c,\star} = \begin{pmatrix} 0 & x^{\star,(2)} & d_{\max}^{(2)} \\ 1 & 1 & 1 \\ \overline{3} & \overline{3} & \overline{3} \end{pmatrix}, \qquad \mu^{c,\star} = \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ \overline{4} & \overline{4} \end{pmatrix}. \tag{4.11}$$

We can also obtain the minimally supported D-optimal designs for the exponential model and the linear-in-log with common location and scale parameter.

THEOREM 4.7. *Let* $r = \frac{\sigma_1^2}{\sigma_2^2}$, $\bar{\theta}_2^{(i)} = \frac{\theta_2^{(i)}}{d_{\max}^{(i)}}$, $i = 1, 2$, $0 < \bar{\theta}_2^{(1)} < \bar{\theta}_2^{(2)} < 1$, *define*

$$g(\theta, x) = 1 + (x-1)\exp\left(\frac{x}{\theta}\right) - x\exp\left(\frac{x-1}{\theta}\right)$$

and $y^\star = d_{\max}^{(2)}$, $z^\star = d_{\max}^{(1)}$ *and the point* $x^{\star,(i)} = x_{\exp}^{\star,(i)}$ *by* (4.7) *for* $i = 1, 2$.

(1) *If* $r \le 1$, *the D-optimal design for model* (3.7) *in the class* Ξ_4^2 *is given by* (4.9).

(2) *If* $1 < r \le \frac{g^2(\theta_2^{(1)}, x_{\exp}^{\star,(1)})}{g^2(\theta_2^{(2)}, x_{\exp}^{\star,(2)})}$, *the D-optimal design for model* (3.7) *in the class* Ξ_4^2 *is given by* (4.10).

(3) *If* $r > \frac{g^2(\theta_2^{(1)}, x_{\exp}^{\star,(1)})}{g^2(\theta_2^{(2)}, x_{\exp}^{\star,(2)})}$, *the D-optimal design for model* (3.7) *in the class* Ξ_4^2 *is given by* (4.11).

THEOREM 4.8. *Let* $r = \frac{\sigma_1^2}{\sigma_2^2}$, $\bar{\theta}_2^{(i)} = \frac{\theta_2^{(i)}}{d_{\max}^{(i)}}$ $i = 1, 2$, $0 < \bar{\theta}_2^{(1)} < \bar{\theta}_2^{(2)} < 1$, *define*

$$g(\theta, x) = \left(\frac{x}{x+\theta}\right)\left(\log\left(\frac{1}{\theta}+1\right)(1+\theta) - \log\left(\frac{x}{\theta}+1\right)\left(1+\frac{\theta}{x}\right)\right)$$

and $y^\star = d_{\max}^{(2)}$, $z^\star = d_{\max}^{(1)}$ *and the point* $x^{\star,(i)} = x_{\log}^{\star,(i)}$ *by* (4.8) *for* $i = 1, 2$.

(1) *If* $r \le 1$, *the D-optimal design for model* (3.6) *in the class* Ξ_4^2 *is given by* (4.9).

(2) *If* $1 < r \le \frac{g^2(\theta_2^{(1)}, x_{\log}^{\star,(1)})}{g^2(\theta_2^{(2)}, x_{\log}^{\star,(2)})}$, *the D-optimal design for model* (3.6) *in the class* Ξ_4^2 *is given by* (4.10).

(3) *If* $r > \frac{g^2(\theta_2^{(1)}, x_{\log}^{\star,(1)})}{g^2(\theta_2^{(2)}, x_{\log}^{\star,(2)})}$, *the D-optimal design for model* (3.6) *in the class* Ξ_4^2 *is given by* (4.11).

The proof of Theorem 4.6 can be found in the Supplement (Feller et al., 2017), the proofs of Theorems 4.7 and 4.8 are similar.

4.2.2. *D-optimal designs in the class of all designs*. The question if a minimally supported D-optimal design for one of the models considered in Section 4.2.1 is in fact D-optimal in the class of all designs is an extremely difficult one. Its answer depends sensitively on the particular parameters in the model under consideration and differs for the three dose response models under consideration. We exemplarily state a result for the Emax model, which provides sufficient conditions for the D-optimality of a minimally supported D-optimal design, and illustrates the general structure and difficulties in results of this type. The proof is based on the equivalence Theorem 4.1 and given in the Appendix. Similar but substantially more complicated statements can also be obtained of the linear-in-log and the exponential model (note that in contrast to the Emax model these models contain transcendental functions).

THEOREM 4.9. *Let* $r = \frac{\sigma_1^2}{\sigma_2^2}$, $\bar{\theta}_2^{(i)} = \frac{\theta_2^{(i)}}{d_{\max}^{(i)}}$, $i = 1, 2$ *and assume* $0 < \bar{\theta}_2^{(1)} < \bar{\theta}_2^{(2)} < 1$.

(1) *Let* $r \leq 1$. *The design* $\xi^{a,\star}$ *defined in* (4.9) *is locally D-optimal for model* (3.5) *if the condition*

$$\bar{\theta}_2^{(2)} \geq \frac{r(6\bar{\theta}_2^{(1)}(\bar{\theta}_2^{(1)}+1)(2\bar{\theta}_2^{(1)}+1)^2) - (1-r)}{(6+2r\bar{\theta}_2^{(1)}(1+2\bar{\theta}_2^{(1)}))} \tag{4.12}$$

is satisfied.

(2) *Let* $r > 1$. *The design* $\xi^{b,\star}$ *defined in* (4.10) *is locally D-optimal for model* (3.5) *if and only if the condition*

$$\bar{\theta}_2^{(2)} \geq \frac{(\bar{\theta}_2^{(1)})^2(1+2\bar{\theta}_2^{(1)})^2 + r(1+\bar{\theta}_2^{(1)})^2(1+4\bar{\theta}_2^{(1)}+20(\bar{\theta}_2^{(1)})^2) - 1}{6+2\bar{\theta}_2^{(1)}(1+2\bar{\theta}_2^{(1)})} \tag{4.13}$$

is satisfied.

(3) *Let* $r > 1$. *The design* $\xi^{c,\star}$ *defined in* (4.11) *is locally D-optimal for model* (3.5) *if the condition*

$$\bar{\theta}_2^{(1)} \geq \frac{\frac{1}{r}(6\bar{\theta}_2^{(2)}(\bar{\theta}_2^{(2)}+1)(2\bar{\theta}_2^{(2)}+1)^2) - (1-\frac{1}{r})}{(6+2\frac{1}{r}\bar{\theta}_2^{(2)}(1+2\bar{\theta}_2^{(2)}))} \tag{4.14}$$

is satisfied.

Figure 1 illustrates the parameter domains for different ratios $r = \frac{\sigma_1^2}{\sigma_2^2}$. The case where the variance is equal in both groups is presented in the third panel. Obviously, there are several parameter constellations $\theta_2^{(1)} \geq \theta_2^{(2)}$ where the minimally supported D-optimal design $\xi^{a,\star}$ is not D-optimal in the class of all designs.

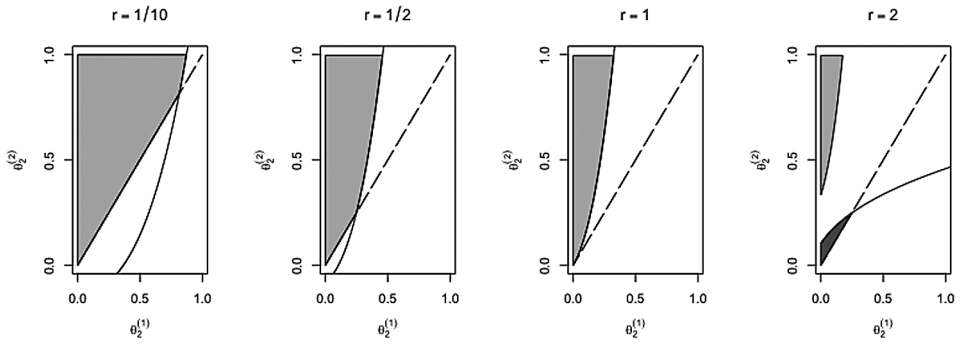

FIG. 1. *The marked regions describe the parameter spaces, where the minimally supported D-optimal design is optimal in the Emax model* (3.5) (*see Theorem* 4.9). *The different figures correspond to different values of* $r = \sigma_1^2/\sigma_2^2$. *The domain for the first case of Theorem* 4.9 *is represented in gray for the case* $r = 1/10$, $r = 1/2$ *and* $r = 1$ (*see the first three panels from the left*). *In the right panel, we display the case* $r = 2$ *of Theorem* 4.9 [*here the gray region corresponds to case* (2), *while the dark gray region corresponds to case* (3)].

5. Application to a dose-finding study. In this section, we illustrate the application of the results of the previous sections and discuss the problem of designing experiments for a dose finding study with different treatment groups. Our example refers to a Phase II study on a drug that works by increasing the level of a biomarker that induces a beneficial clinical effect in patients. The dosing groups under consideration are monthly and weekly administration. The primary objective of the study is the characterization of dose-response relationships at a given time-point, say T, after initiation of treatment for each of these two dosing groups. This will support the selection of an appropriate dose level and group to be used in phase III clinical trials. To maintain the confidentiality of the trial the dose-range has been rescaled and the considered range (in terms of total monthly dose) is $[0, 400]$ for the weekly group and $[0, 1000]$ for the once-a-month group. The natural questions for the design of this study are (i) which doses should be studied in each treatment group and (ii) how to split the total sample size between the two treatment groups. Here, the objective of the study is addressed by deriving the best estimates of the dose response curves, a task for which a D-optimal design is best suited.

To arrive at a suitable design for the Phase II study, we need to quantify the available information. This quantification can generate a best guess for the dose-response curves, but even better, it can be used to obtain a candidate set of dose-group-response scenarios to reflect the uncertainty about the true dose-group-response relationship. In this case, information was available from a very small early trial, which was used to develop a nonlinear mixed effects pharmacokinetics (PK)/pharmacodynamics (PD) model linking drug concentrations to biomarker levels. Using this model, data of the new trial were predicted for the time-point T of the dose-response analysis and dose-group-response models were fitted to

the data. Under the assumption of a normal distribution for the logarithm of the biomarker level, it turned out that the Emax function was able to adequately describe the population average predicted by the PK/PD model. The Emax model utilized total monthly dose as input and had different ED_{50} parameters in the two groups ($\theta_2^{(1)}$ and $\theta_2^{(2)}$), but the same placebo ϑ_{11} and Emax parameter ϑ_{12}, so that the model function in the weekly and monthly group is given by

$$f\big(d, \theta_1, \theta_2^{(i)}\big) = \vartheta_{11} + \vartheta_{12} \frac{d}{\theta_2^{(i)} + d}, \qquad i = 1, 2.$$

Here, group $i = 1$ contains patients receiving monthly administration and the group $i = 2$ the weekly administration. The parameter estimates can be found in Table 2 as model 1, which can be considered as population average fit. We now use these estimates as a guess and determine the locally D-optimal design for these values. The variability is expected to be the same in both treatment groups. Recalling the design spaces for the monthly and weekly doses are $\mathcal{X}_1 = [0, 1000]$ and $\mathcal{X}_2 = [0, 400]$, respectively, we obtain from Theorem 4.6 and Theorem 4.9 the (locally) D-optimal design $\xi^\star = (\xi_1^\star, \xi_2^\star, \mu^\star)$, where $\xi_1^\star$ puts equal masses at the dose levels 0, 13.45, 1000, the design $\xi_2^\star$ is a one-point design at the point 10.46 and the design $\mu^\star$ has masses $\frac{3}{4}$ and $\frac{1}{4}$ at the points 1 and 2, respectively. It can be seen that based on the population average fit, it is sufficient to investigate the low dose-range in both groups and a high dose in one of the two groups. Here, the maximum dose is placed in the monthly group because $\theta_2^{(1)}/d_{\max}^{(1)} < \theta_2^{(2)}/d_{\max}^{(2)}$, so relative to the allowed maximum dose a larger ED_{50} parameter exists for the weekly group and thus patients are allocated to the monthly group.

In practice, it is not realistic to assume that the data and model from previous small trials completely represent the underlying truth (otherwise no further study would need to be conducted). So it is important to derive ranges covering the uncertainty about the available information to use for the design of the new study. First, uncertainty with respect to the model parameters in the Emax model family, but also in terms of uncertainty on the parametric form of the model function. This is important, because the population to be included in the Phase II trial will cover a broader range of characteristics than in the small proof of concept trial. For this purpose, the PK/PD model was used to predict individual dose-response curves and the Emax model was fitted to the individual dose-response curves to derive a range of plausible dose-response parameters. Quantiles of the derived parameter distributions were used to derive four additional candidate model shapes. More details on how these candidate shapes were derived can be found in the Supplement (Feller et al., 2017). The parameters for these four additional candidate models can be found in Table 2 under the numbers 2–5. These models are depicted in the first row of Figure 2. To robustify the design calculation with respect to this set of candidate models, the design maximizing the mean efficiency

$$g_c(\xi, s) = \sum_{i=1}^{s} \pi_i \mathrm{Eff}_i(\xi) \tag{5.1}$$

TABLE 2
Set of candidate models used in the robust criterion (5.1)

Model type	**id**	ϑ_{11}	ϑ_{12}		$\theta_2^{(1)}$	$\theta_2^{(2)}$	γ
Emax	1	5.48	0.90		13.82	10.46	1
Emax	2	5.47	0.93		2.93	2.39	1
Emax	3	5.47	0.93		2.93	40.40	1
Emax	4	5.47	0.93		53.49	2.39	1
Emax	5	5.47	0.93		53.49	40.40	1
Sigmoid Emax	6	5.48	0.90		13.82	10.46	3
Model type	**id**	θ_1	$\vartheta_1^{(1)}$	$\vartheta_1^{(2)}$	$\vartheta_2^{(1)}$	$\vartheta_2^{(2)}$	γ
Emax	7	5.48	0.85	0.95	13.82	10.46	1
Sigmoid Emax	8	5.48	0.65	0.75	2.93	2.39	3
Sigmoid Emax	9	5.48	0.95	1.05	53.49	40.40	3
Log	10	5.44	0.13	0.14	0.32	0.41	

can be calculated, where s is the number of candidate models (here 5 or 10), $\pi_1, \ldots, \pi_s$ are nonnegative model weights chosen to reflect prior probability associated the model function $1, \ldots, s$ (throughout this paper we will use $\pi_i = 1/s$, $i = 1, \ldots, s$). The efficiencies $\mathrm{Eff}_i(\xi)$ of the experimental design ξ with respect

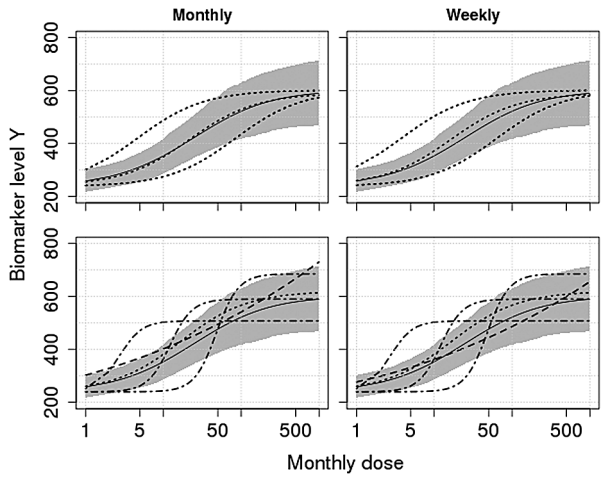

FIG. 2. *Candidate models for the dose-response curve in monthly and weekly group. In the first row, models* 1–5 *are depicted and in the second row models* 6–10 (*see Table* 2). *The solid line represents the population average of the new trial data generated with the PK/PD model. The grey area represents the biomarker level between the* 25*th and* 75*th quantiles of the patient responses of the new trial data. Dotted curves correspond to the Emax models. Sigmoid Emax models are depicted as dotted-dashed lines and the linear-in-log models as dashed lines.*

to the (locally) D-optimal design $\xi^{\star,i}$ associated to the model i is defined as $\mathrm{Eff}_i(\xi) = (|M_i(\xi,\theta_i)|/|M_i(\xi^{\star,i},\theta_i)|)^{1/m_i}$, where M_i is the Fisher information matrix associated to the model i with parameter specification θ_i and m_i is the number of parameters of this model. The criterion (5.1) is called *Bayesian* or *compound* optimality criterion in the literature [see Dette (1990), Cook and Wong (1994) or Tsai and Zen (2004), Zen and Tsai (2004) among many others]. In the following, we will denote the designs maximizing (5.1) by $\xi^\star_{c,s} = (\xi^\star_{1,c,s}, \xi^\star_{2,c,s}, \mu^\star_{c,s})$ and call it compound optimal design. We emphasize that the definition of the criterion (5.1) requires knowledge of the locally optimal designs $\xi^{\star,i}$, which have been determined in Section 4.

The compound optimal design based on the first 5 models in Table 2 can be calculated numerically and is given by $\xi^\star_{c,5} = (\xi^\star_{1,c,5}, \xi^\star_{2,c,5}, \mu^\star_{c,5})$, where $\xi^\star_{1,c,5}$ puts the weights 0.26, 0.24, 0.25, 0.25 at the dose levels 0, 3.02, 43.67, 1000, the design $\xi^\star_{2,c,5}$ puts the weights 0.48, 0.52 at the points 2.53, 37.51 and the design $\mu^\star$ has masses 0.67 and 0.33 at the points 1 and 2, respectively. Its optimality can be proved by an analogue of Theorem 4.1 for the Bayesian optimality criterion (5.1). Compared to the design using only the best guess model, now the low dose-range is investigated in finer granularity by using two instead of one dose (safeguarding against different possible values of the ED_{50}). In addition, still more patients are evaluated for the monthly group, as the high dose is only used there.

Based on general plausibility considerations, five further candidate shapes were included as example of models different from the Emax function (e.g., the sigmoid Emax and the linear-in-log function), or of models where the maximum efficacy differed between the two groups. These models are shown in the second row of Figure 2 and the corresponding parameters are given in the rows with ids 6–10 in Table 2. First, a sigmoid Emax model

$$f(d,\theta_1,\theta_2^{(i)},\gamma) = \vartheta_{11} + \vartheta_{12}\frac{d^\gamma}{(\theta_2^{(i)})^\gamma + d^\gamma}, \qquad i = 1,2,$$

with Hill coefficient $\gamma = 3$ (model 6) is also considered as a possible dose response function. Note that this model provides a steeper dose-response curve compared to the Emax model, but with the same ED_{50} values as model 1. Furthermore, an Emax and a sigmoid Emax model

$$f(d,\theta_1,\theta_2^{(i)},\gamma) = \theta_1 + \frac{\vartheta_1^{(i)} d^\gamma}{(\vartheta_2^{(i)})^\gamma + d^\gamma}, \qquad i = 1,2$$

is added that allows for different Emax parameters in the two treatment group (models 7, 8, 9). In addition, a linear-in-log model (id 10) is utilized. The locally D-optimal designs for these models are computed using the results of Section 4. For the sigmoid Emax models, a transformation has to be used to reduce it to the case of an Emax model, such that the derived theory is applicable (note that the parameter γ is assumed to be fixed).

TABLE 3
Efficiency $Eff_i(\xi^\star_{c,s})$ of the two compound optimal designs compared to each of the locally D-optimal designs for the 10 *models*

	$g_c(\cdot, s)$	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
$\xi^\star_{c,5}$	0.82	0.71	0.84	0.88	0.85	0.85	0.10	0.80	0.93	0.91	0.63
$\xi^\star_{c,10}$	0.75	0.83	0.75	0.78	0.77	0.79	0.75	0.90	0.76	0.75	0.75

When using all $s = 10$ candidate models, we obtain $\xi^\star_{c,10} = (\xi^\star_{1,c,10}, \xi^\star_{2,c,10}, \mu^\star_{c,10})$ where $\xi^\star_{1,c,10}$ puts the weights 0.27, 0.13, 0.22, 0.13, 0.24 at the dose levels 0, 2.90, 12.98, 41.91, 1000, the design $\xi^\star_{2,c,10}$ puts the weights 0.33, 0.21, 0.31, 0.15 at the dose levels 3.01, 13.16, 49.46, 400 and the design $\mu^\star_{c,10}$ has masses 0.58, 0.42 at the points 1 and 2. This design investigates the lower dose range comparably to the previous design based on the first five candidate models, but the maximum dose is studied in both groups. The efficiencies of the two designs $\xi^\star_{c,5}$ and $\xi^\star_{c,10}$ in the different models are displayed in Table 3. We observe that the design $\xi^\star_{c,5}$ has reasonable efficiencies in all models except in the sigmoid Emax (6). Note that this design has been constructed on the basis of the models (1)–(5). On the other hand, the the design $\xi^\star_{c,10}$ maximizes the criterion (5.1), where uncertainty with respect to all models (1)–(10) is addressed. As a consequence, this design has efficiencies varying between 75%–90% in all competing models under consideration. Moreover, it can be used for a goodness-of-fit test of the Emax model, as both components have more than 3 support points. For these reasons, we recommend this design for the Phase II study under consideration.

APPENDIX: TECHNICAL DETAILS

PROOF OF LEMMA 3.1. We only discuss the first part of the proof. The second assertion follows by similar argument. Let $\xi = (\xi_1, \ldots, \xi_M, \mu)$ be an arbitrary design with $|\operatorname{supp}(\xi_i)| \geq \frac{(k+2)}{2}$ $(i = 1, \ldots, M)$ and assume that there exists a design $\xi^+ = (\xi_1^+, \ldots, \xi_M^+, \mu)$ such that $C(\xi^+, \theta^{(i)}) \geq C(\xi, \theta^{(i)})$ for all $i = 1, \ldots, M$. Recalling the definition of the matrix P_i in (3.1) it then follows by Theorem 14.2.9 of Harville (1997) that

$$M^{(i)}(\xi_i^+, \theta) = P_i C(\xi_i^+, \theta^{(i)}) P_i^T \geq P_i C(\xi_i, \theta^{(i)}) P_i^T = M^{(i)}(\xi_i, \theta), \quad i = 1, \ldots, M.$$

This implies

$$M(\xi^+, \theta) = \sum_{i=1}^M \lambda_i M^{(i)}(\xi_i^+, \theta) \geq \sum_{i=1}^M \lambda_i M^{(i)}(\xi_i, \theta) = M(\xi, \theta)$$

and the design ξ^+ increases the information matrix $M(\cdot, \theta)$ with respect to the Loewner ordering.

It now follows from Theorem 3.1 of Dette and Melas (2011) that there exists a design ξ^+ with components ξ_i^+ with at most $\frac{k+2}{2}$ support points ($i = 1, \dots, M$). The statements (1a) and (1b) in Lemma 3.1 also follows from Theorem 3.1 in Dette and Melas (2011). $\square$

PROOF OF LEMMA 3.2. We only prove the lemma for the model given by (2.6). The proof for model (2.7) is analogous. Note that in the model under consideration we have $\frac{\partial}{\partial\theta_1} f(d, \theta_1, \theta_2^{(i)}) = (1, f_0(d, \theta_2^{(i)}))^T$ for the gradient in (2.5). Consequently, if δ_0 denotes the Dirac measure at the point 0, it follows for the matrices $M^{(i)}$ defined in (2.4) that

$$\sigma_i^2 M^{(i)}(\delta_0, \theta) = \sigma_1^2 M^{(1)}(\delta_0, \theta), \qquad i = 1, \dots, M. \tag{A.1}$$

Now, we consider the design $\eta = (\eta_1, \dots, \eta_M, \nu)$ and represent its components as

$$\eta_i = \omega_0^{(i)}\delta_0 + (1 - \omega_0^{(i)})\eta_i^0, \qquad i = 1, \dots, M, \qquad \nu = \sum_{i=1}^M \lambda_i \delta_i.$$

Here, δ_t is the Dirac measure at the point t, $\lambda_i, \omega_0^{(i)} \in [0, 1]$, $i = 1, \dots, M$ and $\eta_1^0, \dots, \eta_M^0$ denote designs with $0 \notin \operatorname{supp}(\eta_i^0)$. Moreover, at least for one $i \in \{1, \dots, M\}$ we have $\lambda_i \omega_0^{(i)} > 0$.

We now assume without loss of generality that $j^* = 1$ and construct a "better" design $\xi = (\xi_1, \dots, \xi_M, \mu)$ as follows:

$$\xi_1 = \omega^\star\delta_0 + (1 - \omega^\star)\eta_1^0, \qquad \xi_i = \eta_i^0, \qquad i = 2, \dots, M, \qquad \mu = \sum_{i=1}^M \lambda_i^\star \delta_i,$$

where $\lambda_1^\star = \lambda_1 + \sum_{i=2}^M \lambda_i \omega_0^{(i)} \in [0, 1]$, $\lambda_i^\star = \lambda_i(1 - \omega_0^{(i)})$ $(i = 2, \dots, M)$, $\omega^\star = (\sum_{i=1}^M \lambda_i \omega_0^{(i)})/(\lambda_1 + \sum_{i=2}^M \lambda_i \omega_0^{(i)})$. Note that we shift the weights of the measures η_i at the point 0 to the design for the group with the smallest population variance. Observing (A.1) gives for the difference $M(\xi, \theta) - M(\eta, \theta)$

$$\begin{aligned}
&\omega^\star\lambda_1^\star M^{(1)}(\delta_0, \theta) + (1 - \omega^\star)\lambda_1^\star M^{(1)}(\eta_1^0, \theta) + \sum_{i=2}^M \lambda_i^\star M^{(i)}(\eta_i^0, \theta) \\
&\quad - \left(\sum_{i=1}^M \omega_i^{(0)} \lambda_i M^{(i)}(\delta_0, \theta) + \sum_{i=1}^M \lambda_i (1 - \omega_i^{(0)}) M^{(i)}(\eta_i^0, \theta) \right) \\
&\quad = \left(\omega^\star\lambda_1^\star - \sum_{i=1}^M \frac{\sigma_1^2}{\sigma_i^2} \omega_i^{(0)} \lambda_i \right) M^{(1)}(\delta_0, \theta) \\
&\quad \geq \left(\omega^\star\lambda_1^\star - \sum_{i=1}^M \lambda_i \omega_0^{(i)} \right) M^{(1)}(\delta_0, \theta) = 0,
\end{aligned}$$

since $\sigma_1^2 \leq \sigma_i^2$ $(i = 1, \dots, M)$. $\square$

PROOF OF COROLLARY 3.3. Lemma 3.1 can be applied in the case of an Emax model or linear-in-log model with $k = 4$ [see Yang (2010)]. Consequently, there exists a design ξ^+ with $3M$ support points and each component ξ_i^+ contains the placebo 0 and $d_{\max}^{(i)}$, $i = 1, \dots, M$. Now we apply Lemma 3.2 with $\eta = \xi^+$ and we allocate the placebo 0 in the group with the smallest variance. For the exponential model, Lemma 3.1 can be applied with $k = 5$ [see Yang (2010)]. Consequently, there exists a design ξ^+ with $3M$ support points and each component ξ_i^+ contains $d_{\max}^{(i)}$, $i = 1, \dots, M$. $\square$

PROOF OF THEOREM 4.3. For the sake of transparency, we restrict ourselves to the case $M = 2$ such that $m = Mq + 1 = 2q + 1$.We use the equivalence Theorem 4.1 to establish the D-optimality of the design $\xi^\star$. In the present situation, this means that the D-optimality of the design $\xi^\star$ defined in (4.5) for model (2.7) with assumption (4.3) can be proved by checking the two inequalities:

$$\begin{aligned}\kappa_1(t, \xi^\star, \theta) &= \frac{1}{\sigma_1^2}\big(1, \eta_0^T(t, \theta_2^{(1)}), 0_q^T\big) M^{-1}(\xi^\star, \theta)\big(1, \eta_0^T(t, \theta_2^{(1)}), 0_q^T\big)^T \\ &\le 2q + 1,\end{aligned} \tag{A.2}$$

$$\begin{aligned}\kappa_2(t, \xi^\star, \theta) &= \frac{1}{\sigma_2^2}\big(1, 0_q^T, \eta_0^T(t, \theta_2^{(2)})\big) M^{-1}(\xi^\star, \theta)\big(1, 0_q^T, \eta_0^T(t, \theta_2^{(2)})\big)^T \\ &\le 2q + 1,\end{aligned} \tag{A.3}$$

for $t \in [0, d_{\max}^{(1)}]$, $t \in [0, d_{\max}^{(2)}]$, respectively, where $\eta_0(t, \theta_2^{(i)}) = \frac{\partial}{\partial \theta_2^{(i)}} f_0(t, \theta_2^{(i)})$ $(i = 1, 2)$. A straightforward calculation shows that the information matrix of the design $\xi^\star$ can be represented as

$$M(\xi^\star, \theta) = \frac{1}{m} X\big(\sigma_1, \theta_2^{(1)}, \sigma_2, \theta_2^{(2)}\big) X^T\big(\sigma_1, \theta_2^{(1)}, \sigma_2, \theta_2^{(2)}\big), \tag{A.4}$$

where the matrix $X(\sigma_1, \theta_2^{(1)}, \sigma_2, \theta_2^{(2)})$ is given by

$$\begin{pmatrix} X_{11}(\sigma_1, 0, d_1^{(1)}, \dots, d_q^{(1)}, \theta_2^{(1)}) & X_{12}(\sigma_2) \\ 0 & X_{22}(\sigma_2, d_1^{(2)}, \dots, d_q^{(2)}, \theta_2^{(2)}) \end{pmatrix},$$

and the matrices $X_{11}(\sigma, 0, d_1^{(1)}, \dots, d_q^{(1)}, \theta_2^{(1)}) \in \mathbb{R}^{(q+1)\times(q+1)}$, $X_{22}(\sigma, d_1^{(2)}, \dots, d_q^{(2)}, \theta_2^{(2)}) \in \mathbb{R}^{q\times q}$ and $X_{12}(\sigma) \in \mathbb{R}^{(q+1)\times q}$ are defined by

$$X_{22}(\sigma, d_1^{(2)}, \dots, d_q^{(2)}, \theta_2^{(2)}) = \frac{1}{\sigma}(\eta_0(d_1^{(2)}, \theta_2^{(2)}), \dots, \eta_0(d_q^{(2)}, \theta_2^{(2)})),$$

$$X_{11}(\sigma, 0, d_1^{(1)}, \dots, d_q^{(1)}, \theta_2^{(1)}) = \frac{1}{\sigma}\begin{pmatrix} 1 & 1_q^T \\ 0_q & X_{22}(\sigma, d_1^{(1)}, \dots, d_q^{(1)}, \theta_2^{(1)}) \end{pmatrix},$$

$$X_{12}(\sigma) = \frac{1}{\sigma}\begin{pmatrix} 1 & \dots & 1 \\ 0_q & \dots & 0_q \end{pmatrix}.$$

Consequently, the inverse of $M(\xi^\star, \theta)$ is obtained as

$$M^{-1}(\xi^\star, \theta) = m\big(X^T(\sigma_1, \theta_2^{(1)}, \sigma_2, \theta_2^{(2)})\big)^{-1} X^{-1}(\sigma_1, \theta_2^{(1)}, \sigma_2, \theta_2^{(2)}),$$

where the matrix $X^{-1}(\sigma_1, \theta_2^{(1)}, \sigma_2, \theta_2^{(2)})$ is given by

$$\begin{pmatrix} X_{11}^{-1}(\sigma, 0, d_1^{(1)}, \ldots, d_q^{(1)}, \theta_2^{(1)}) & X_{12}^{\mathrm{inv}} \\ 0 & X_{22}^{-1}(\sigma, d_1^{(2)}, \ldots, d_q^{(2)}, \theta_2^{(2)}) \end{pmatrix}$$

with

$$X_{12}^{\mathrm{inv}} = -X_{11}^{-1}(\sigma, 0, d_1^{(1)}, \ldots, d_q^{(1)}, \theta_2^{(1)}) X_{12}(\sigma_2)\, X_{22}^{-1}(\sigma, d_1^{(2)}, \ldots, d_q^{(2)}, \theta_2^{(2)}, \theta_2^{(2)}).$$

Using these block structures, the function $\kappa_1(t, \xi^\star, \theta)$ defined in (A.2) reduces for the design $\xi^\star = (\xi_1^\star, \xi_2^\star, \mu^\star)$ to

$$\begin{aligned} &\kappa_1(t, \xi^\star, \theta) \\ &\quad = \frac{m}{\sigma_1^2} g^T(t, \theta^{(1)}) \big(X_{11}^{-1}(\sigma, 0, d_1^{(1)}, \ldots, d_q^{(1)})\big)^T \\ &\qquad \times X_{11}^{-1}(\sigma, 0, d_1^{(1)}, \ldots, d_q^{(1)}) g(t, \theta^{(1)}) \\ &\quad = \frac{m}{(q+1)\sigma_1^2} g^T(t, \theta^{(1)}) M_1^{-1}(\xi_1^\star, \theta^{(1)}) g(t, \theta^{(1)}), \end{aligned}$$

where

$$g^T(t, \theta^{(1)}) = (1, \eta_0^T(t, \theta_2^{(1)}))$$

and

$$M_1(\xi_1^\star, \theta^{(1)}) = \frac{1}{\sigma_1^2} \int_0^1 g(t, \theta^{(1)}) g^T(t, \theta^{(1)})\, d\xi_1^\star(t)$$

denote the gradient and the information matrix of the design $\xi_1^\star$ in the single model with parameter $\theta^{(1)} = (\theta_1, (\theta_2^{(1)})^T)^T$. Consequently, the function $\kappa_1(t, \xi^\star, \theta_1, \theta_2^{(1)})$ only depends on the first component $\xi_1^\star$ and is proportional to the left-hand side of the standard equivalence theorem for D-optimality for the single model. The inequality $\kappa_1(t, \xi^\star, \theta) \le m$ for all $t \in [0, d_{\max}^{(1)}]$ follows from the fact that the design $\xi_1^\star$ given in (4.4) is locally D-optimal for the single model with parameter $\theta^{(1)} = (\theta_1, (\theta_2^{(1)})^T)^T$. This proves (A.2).

In order to show the remaining inequality (A.3) for all $t \in [0, d_{\max}^{(2)}]$, we use the fact that the information matrix in (A.4) can be represented as

$$\begin{aligned} M(\xi^\star, \theta) = SX(\sigma_2, \theta_2^{(2)}, \sigma_1, \theta_2^{(1)}) \operatorname{diag}\Big(\frac{\sigma_2^2}{m\sigma_1^2}, \frac{1}{m}, \ldots, \frac{1}{m}\Big) \\ \times X^T(\sigma_2, \theta_2^{(2)}, \sigma_1, \theta_2^{(1)}) S, \end{aligned}$$

where S denotes a $m \times m$ permutation matrix, defined by

$$S = \begin{pmatrix} 1 & 0_q^T & 0_q^T \\ 0_q & 0_{q\times q} & I_{q\times q} \\ 0_q & I_{q\times q} & 0_{q\times q} \end{pmatrix},$$

$0_{q\times q}$ denotes a matrix with all entries equal to zero and $I_{q\times q}$ the $q \times q$ identity matrix. Observing that $Sh_2(t) = \frac{1}{\sigma_2}(1, \eta_0^T(t, \theta_2^{(2)}), 0_q^T)^T$, it follows that the function $\kappa_2(t, \xi^\star, \theta)$ in (A.3) can be represented as

$$\begin{aligned}
&\frac{1}{\sigma_2^2}(1, \eta_0^T(t, \theta_2^{(2)}))(X_{11}^{-1}(\sigma_2, 0, d_1^{(2)}, \ldots, d_q^{(2)}, \theta_2^{(2)}))^T \\
&\quad \times \left[\operatorname{diag}(m \cdot 1_{q+1} - m\left(1 - \frac{\sigma_1^2}{\sigma_2^2}\right)\operatorname{diag}(1, 0_q)\right] \\
&\quad \times X_{11}^{-1}(\sigma_2, 0, d_1^{(2)}, \ldots, d_q^{(2)}, \theta_2^{(2)})(1, \eta_0^T(t, \theta_2^{(2)}))^T \\
&\qquad = \frac{m}{(q+1)\sigma_2^2}(1, \eta_0^T(t, \theta_2^{(2)}))M_2^{-1}(\tilde{\xi}_2, \theta_1, \theta_2^{(2)})(1, \eta_0^T(t, \theta_2^{(2)}))^T \\
&\qquad\quad - m\left(1 - \frac{\sigma_1^2}{\sigma_2^2}\right)\frac{1}{\sigma_2^2} \\
&\qquad\quad \times [(1, 0_q^T)X_{11}^{-1}(\sigma_2, 0, d_1^{(2)}, \ldots, d_q^{(2)}, \theta_2^{(2)})(1, \eta_0^T(t, \theta_2^{(2)}))^T]^2,
\end{aligned}$$

where $M_2(\tilde{\xi}_2, \theta_1, \theta_2^{(2)})$ is the information matrix of the design $\tilde{\xi}_2$ given by (4.4) for the single model. The first term of this expression is proportional to the left-hand side of the equivalence theorem corresponding to the D-optimality in the single model with parameter $\theta^{(2)} = (\theta_1, (\theta_2^{(2)})^T)^T$. Moreover, it follows that the design $\tilde{\xi}_2$ is D-optimal for the single model with parameter $\theta^{(2)} = (\theta_1, (\theta_2^{(2)})^T)^T$, which implies that the first term is always smaller than m. By the assumption $\sigma_1^2 \le \sigma_2^2$, we obtain that the second term of this expression is nonpositive, which shows $\kappa_2(t, \xi^\star, \theta) \le m$ for all $t \in [0, d_{\max}^{(2)}]$. This proves the inequality (A.3) and completes the proof of Theorem 4.3 in the case $M = 2$. □

PROOF OF COROLLARY 4.4. The locally D-optimal designs for the (single) Emax, the linear-in-log and the exponential model were calculated by Dette et al. (2010). The corollary now follows by an application of Theorem 4.3. □

PROOF OF LEMMA 4.5. Let $\xi^\star = (\xi_1^\star, \xi_2^\star, \mu^\star)$ denote the locally D-optimal design for the Emax, the linear-in-log or the exponential model. Since the information matrix $M(\xi^\star, \theta)$ of a locally D-optimal design must be nonsingular, one

can easily deduce $|\operatorname{supp}(\xi_1^\star)| + |\operatorname{supp}(\xi_2^\star)| \geq 4$ and

$$\begin{aligned} &\text{If } \big|\operatorname{supp}\big(\xi_1^\star\big)\big| + \big|\operatorname{supp}\big(\xi_2^\star\big)\big| = 4, \qquad \text{then } 0 \notin \operatorname{supp}\big(\xi_1^\star\big) \cap \operatorname{supp}\big(\xi_2^\star\big), \\ &\text{If } \big|\operatorname{supp}\big(\xi_i^\star\big)\big| = 1, \qquad \text{then } 0 \notin \operatorname{supp}\big(\xi_i^\star\big), \qquad i = 1, 2. \end{aligned} \tag{A.5}$$

Moreover, it follows by Corollary 3.3 that the locally D-optimal design has at most 5 support points for the Emax and the linear-in-log model and at most 6 support points for the exponential model. This proves Assertion (A1) and (B). Assertion (A2) also follows by Corollary 3.3.

For a proof of (A3), we note that $(|\operatorname{supp}(\xi_1^\star)|, |\operatorname{supp}(\xi_2^\star)|) \in \{(1,3), (2,2), (3,1)\}$ if the locally D-optimal design is given by a design in Ξ_2^4. If $(|\operatorname{supp}(\xi_1^\star)|, |\operatorname{supp}(\xi_2^\star)|) = (1,3)$, $\xi_2^\star$ must contain the boundary points $0, d_{\max}^{(2)}$, otherwise it could be improved with respect to the Loewner ordering (see Theorem 3.1). If $(|\operatorname{supp}(\xi_1^\star)|, |\operatorname{supp}(\xi_2^\star)|) = (2,2)$, both designs must contain at least one of the boundary points, otherwise $I(\xi_i^\star) = 2$ $(i = 1, 2)$ and the designs could be improved with respect to the Loewner ordering (see again Theorem 3.1). Using (A.5), it follows that at least one of the designs contains the corresponding upper boundary point. If $(|\operatorname{supp}(\xi_1^\star)|, |\operatorname{supp}(\xi_2^\star)|) = (3,1)$, $\xi_1^\star$ must contain the boundary points $0, d_{\max}^{(1)}$, otherwise it could be improved with respect to the Loewner ordering (see Theorem 3.1). Assertion (A3) now follows. □

PROOF OF THEOREM 4.9. By similar arguments as given in the proof of Theorem 4.6 [see Supplement (Feller et al., 2017)], we obtain that it is sufficient to consider the case $d_{\max}^{(1)} = d_{\max}^{(2)} = 1$.

(1) In the case $r \leq 1$, it follows from Theorem 4.1 that the design $\xi^{a,\star}$ defined in (4.9) is locally D-optimal for model (3.5) if and only if the two inequalities

$$\begin{aligned} &\frac{1}{\sigma_1^2}\left(1, \frac{t}{t+\theta_2^{(1)}}, \frac{-t}{(t+\theta_2^{(1)})^2}, 0\right) M^{-1}\big(\xi^{a,\star}, \theta\big) \\ &\qquad \times \left(1, \frac{t}{t+\theta_2^{(1)}}, \frac{-t}{(t+\theta_2^{(1)})^2}, 0\right)^T \leq 4, \end{aligned} \tag{A.6}$$

$$\begin{aligned} &\frac{1}{\sigma_2^2}\left(1, \frac{t}{t+\theta_2^{(2)}}, 0, \frac{-t}{(t+\theta_2^{(2)})^2}\right) M^{-1}\big(\xi^{a,\star}, \theta\big) \\ &\qquad \times \left(1, \frac{t}{t+\theta_2^{(2)}}, 0, \frac{-t}{(t+\theta_2^{(2)})^2}\right)^T \leq 4 \end{aligned} \tag{A.7}$$

hold for all $t \in [0,1]$ [here (A.6) and (A.7) represent the functions κ_1 and κ_2 in Theorem 4.1], respectively. The information matrix of the design $\xi^{a,\star}$ can be rep-

resented as $M(\xi^{a,\star},\theta)=\frac{1}{4}\tilde{X}(\sigma_1,\theta_2^{(1)},\sigma_2,\theta_2^{(2)})\tilde{X}^T(\sigma_1,\theta_2^{(1)},\sigma_2,\theta_2^{(2)})$, where

$$\tilde{X}\big(\sigma_1,\theta_2^{(1)},\sigma_2,\theta_2^{(2)}\big)=\begin{pmatrix} X_{11}\Big(\sigma_1,0,\dfrac{\theta_2^{(1)}}{2\theta_2^{(1)}+1},1\Big) & \tilde{X}_{12}\big(\sigma_2,\theta_2^{(2)}\big)\\ 0 & \tilde{X}_{22}\big(\sigma_2,\theta_2^{(2)}\big)\end{pmatrix}$$

and the matrices X_{11}, $\tilde{X}_{12}$ and $\tilde{X}_{22}$ are defined by

$$X_{11}\big(\sigma_1,d_1^{(1)},d_2^{(1)},d_3^{(1)}\big)=\begin{pmatrix} 1 & 1 & 1\\ \dfrac{d_1^{(1)}}{d_1^{(1)}+\theta_2^{(1)}} & \dfrac{d_2^{(1)}}{d_2^{(1)}+\theta_2^{(1)}} & \dfrac{d_3^{(1)}}{d_3^{(1)}+\theta_2^{(1)}}\\ \dfrac{-d_1^{(1)}}{(d_1^{(1)}+\theta_2^{(1)})^2} & \dfrac{-d_2^{(1)}}{(d_2^{(1)}+\theta_2^{(1)})^2} & \dfrac{-d_3^{(1)}}{(d_3^{(1)}+\theta_2^{(1)})^2}\end{pmatrix},$$

$\tilde{X}_{12}(\sigma_2,\theta_2^{(2)})=(\frac{1}{\sigma_2},\frac{1}{2\sigma_2})^T$, $\tilde{X}_{12}(\sigma_2,\theta_2^{(2)})=(\frac{-1}{4\theta_2^{(2)}\sigma_2})$. A straightforward calculation of the inverse of the matrix $\tilde{X}$ yields

$$\tilde{X}^{-1}\big(\sigma_1,\theta_2^{(1)},\sigma_2,\theta_2^{(2)}\big)=\begin{pmatrix} X_{11}^{-1}\Big(\sigma_1,0,\dfrac{\theta_2^{(1)}}{2\theta_2^{(1)}+1},1\Big) & \tilde{X}_{12}^{\mathrm{inv}}\\ 0 & \tilde{X}_{22}^{-1}\big(\sigma_2,\theta_2^{(2)}\big)\end{pmatrix},$$

with $\tilde{X}_{12}^{\mathrm{inv}}=-X_{11}^{-1}(\sigma_1,0,\frac{\theta_2^{(1)}}{2\theta_2^{(1)}+1},1)\tilde{X}_{12}(\sigma_2,\theta_2^{(2)})\tilde{X}_{22}^{-1}(\sigma_2,\theta_2^{(2)})$ and we obtain for the left-hand side of (A.6) the representation

$$\begin{aligned}&\frac{4}{3\sigma_1^2}g^T\big(t,\theta^{(1)}\big)3X_{11}^{-T}\Big(\sigma_1,0,\frac{\theta_2^{(1)}}{2\theta_2^{(1)}+1},1\Big)X_{11}^{-1}\Big(\sigma_1,0,\frac{\theta_2^{(1)}}{2\theta_2^{(1)}+1},1\Big)g\big(t,\theta^{(1)}\big)\\ &\quad=\frac{4}{3\sigma_1^2}g^T\big(t,\theta^{(1)}\big)M_1^{-1}\big(\xi_1^{a,\star},\theta^{(1)}\big)g\big(t,\theta^{(1)}\big),\end{aligned}$$

where $g^T(t,\theta^{(1)})=(1,\frac{t}{t+\theta_2^{(1)}},\frac{-t}{(t+\theta_2^{(1)})^2})$ and $M_1(\xi_1^{a,\star},\theta^{(1)})$ are the gradient and the information matrix of the design $\xi_1^{a,\star}$ in the Emax model with parameter vector $\theta^{(1)}=(\theta_1^T,\theta_2^{(1)})^T$. Because the design $\xi_1^{a,\star}$ given in (4.9) is in fact locally D-optimal for this model, it follows that $\kappa_1(t,\xi^{a,\star},\theta)\le 4$, which proves the first inequality of the equivalence theorem.

In order to show that the inequality in (A.7) holds for all $t\in[0,1]$, we note that this inequality is equivalent to

$$P(t)=\alpha_{21}t^4+\alpha_{22}t^3+\alpha_{23}t^2+\alpha_{24}t+\alpha_{25}\le 0, \tag{A.8}$$

where the last identity defines the coefficients α_{2j} in an obvious manner. For example, the leading coefficient and the intercept are given by

$$\begin{aligned}\alpha_{21} &= \frac{1}{\sigma_2^2}(1,1,0,0)M^{-1}\big(\xi^{a,\star},\theta\big)(1,1,0,0)^T - 4\\ &= 24r\theta_2^{(1)}\big(\theta_2^{(1)}+1\big)\big(2\theta_2^{(1)}+1\big)^2 - 4(1-r),\\ \alpha_{25} &= \big(\theta_2^{(2)}\big)^4\bigg(\frac{1}{\sigma_2^2}(1,0,0,0)M^{-1}\big(\xi^{a,\star},\theta\big)(1,0,0,0)^T - 4\bigg) = 4\big(\theta_2^{(2)}\big)^4(r-1),\end{aligned}$$

respectively. Consider the case $r<1$ (the case $r\le 1$ is finally obtained considering the corresponding limit) and note that $P(0)=\alpha_{25}<0$. Consequently, (A.8) holds if either there are no roots of P in the interval $(0,1)$ or all roots of P in the interval $(0,1)$ have multiplicity 2. The roots of $P(t)$ are easily calculated as

$$d_1^{(2)} = \theta_2^{(2)}, \qquad \tilde{d}_1 = \theta_2^{(2)}\frac{3+r\theta_2^{(1)}(1+2\theta_2^{(1)}) - \sqrt{s(\theta_2^{(1)})}}{0.25\,\alpha_{21}},$$

$$\tilde{d}_2 = \theta_2^{(2)}\frac{3+r\theta_2^{(1)}(1+2\theta_2^{(1)}) + \sqrt{s(\theta_2^{(1)})}}{0.25\,\alpha_{21}},$$

where we use the notation

$$\begin{aligned}s\big(\theta_2^{(1)}\big) &= 8 - r^2\big(1+\theta_2^{(1)}\big)^2\big(1+4\theta_2^{(1)}+20\big(\theta_2^{(1)}\big)^2\big)\\ &\quad +2r\big(1+6\theta_2^{(1)}+21\big(\theta_2^{(1)}\big)^2+24\big(\theta_2^{(1)}\big)^3+12\big(\theta_2^{(1)}\big)^4\big).\end{aligned}$$

Note that $s(\theta_2^{(1)})$ is positive (because $\theta_2^{(1)}>0$ and $r\le 1$) and that $\theta_2^{(2)}\in(0,1)$ is a root of multiplicity 2. Moreover, $P(-\theta_2^{(2)})>0$ [since $M^{-1}(\xi^{a,\star},\theta)$ is positive definite], and it follows from $P(0)<0$ that P has a root in the interval $(-\theta_2^{(2)},0)$. This is either $\tilde{d}_1$ or $\tilde{d}_2$ depending on the sign of the leading coefficient α_{21}. The inequality (A.8) holds, if the other root is not in $(0,1)$.

In order to check the location of the roots $\tilde{d}_1$ and $\tilde{d}_2$, we consider the condition (4.12) and the case that the right-hand side of (4.12) is positive. This implies that the leading coefficient α_{21} is positive and the root $\tilde{d}_2$ is also positive. We obtain the inequality $3+r\,\theta_2^{(1)}(1+2\theta_2^{(1)})<\sqrt{s(\theta_2^{(1)})}$ from the condition $\tilde{d}_1\in(-\theta_2^{(2)},0)$. This gives for the second root $\tilde{d}_2>4\theta_2^{(2)}(6+2r\,\theta_2^{(1)}(1+2\theta_2^{(1)}))\alpha_{21}$. Therefore, it follows from (4.12) (with positive right-hand side) that the inequality $\tilde{d}_2\ge 1$ is satisfied.

If the right-hand side of (4.12) is negative, the leading coefficient α_{21} is negative and the conditions $P(0)<0$ and $P(-\theta_2^{(2)})>0$ imply that both roots $\tilde{d}_1$ and $\tilde{d}_2$ must be negative, because otherwise the polynomial P does not satisfy (A.8). Observing that $\tilde{d}_2<\tilde{d}_1$ in this case, it is easy to see that the condition (4.12) (with negative right-hand side) implies $\tilde{d}_1<0$.

Summarizing, in the case $r \leq 1$ the inequality (4.12) implies (A.8) for all $t \in [0,1]$ and the D-optimality of the designs $\xi^{a,\star}$ follows by an application of Theorem 4.1.

(2) At first, we show that the condition (4.13) and $r > 1$ imply that $1 < r \leq \frac{(1+\theta_2^{(2)})^6}{(1+\theta_2^{(1)})^6}$. The last inequality is equivalent to $\theta_2^{(2)} \geq r^{1/6}(1+\theta_2^{(1)}) - 1$ and we have to show that

$$\begin{aligned} &\frac{(\theta_2^{(1)})^2(1+2\theta_2^{(1)})^2 + r(1+\theta_2^{(1)})^2(1+4\theta_2^{(1)}+20(\theta_2^{(1)})^2) - 1}{6+2\theta_2^{(1)}(1+2\theta_2^{(1)})} \\ &\qquad > r^{1/6}\big(1+\theta_2^{(1)}\big) - 1. \end{aligned} \tag{A.9}$$

This inequality can be rewritten by

$$\begin{aligned} &(20r+4)\big(\theta_2^{(1)}\big)^4 + \big(44r - \sqrt[6]{r} + 1\big)\big(\theta_2^{(1)}\big)^3 + \big(29r - 6\sqrt[6]{r} + 5\big)\big(\theta_2^{(1)}\big)^2 \\ &\quad + \big(6r - 8\sqrt[6]{r} + 2\big)\big(\theta_2^{(1)}\big) + \big(r - 6\sqrt[6]{r} + 5\big) > 0. \end{aligned}$$

Note that the coefficients of the polynomial are positive for all $r > 1$. It follows by the rule of Decartes that this polynomial has no positive roots, and consequently, (A.9) is satisfied for all positive $\theta_2^{(1)}$.

Thus, if $r \geq 1$ and the inequality (4.13) holds, we investigate the D-optimality of the design $\xi^{b,\star}$ defined by (4.10) checking the two inequalities

$$\begin{aligned} &\frac{1}{\sigma_1^2}\Big(1, \frac{t}{t+\theta_2^{(1)}}, \frac{-t}{(t+\theta_2^{(1)})^2}, 0\Big) M^{-1}(\xi^{b,\star}, \theta) \\ &\qquad \times \Big(1, \frac{t}{t+\theta_2^{(1)}}, \frac{-t}{(t+\theta_2^{(1)})^2}, 0\Big)^T \leq 4, \end{aligned} \tag{A.10}$$

$$\begin{aligned} &\frac{1}{\sigma_2^2}\Big(1, 0, \frac{t}{t+\theta_2^{(2)}}, 0, \frac{-t}{(t+\theta_2^{(2)})^2}\Big) M^{-1}(\xi^{b,\star}, \theta) \\ &\qquad \times \Big(1, \frac{t}{t+\theta_2^{(2)}}, 0, \frac{-t}{(t+\theta_2^{(2)})^2}\Big)^T \leq 4, \end{aligned} \tag{A.11}$$

for $t \in [0,1]$ [here (A.10) and (A.11) represent the functions κ_1 and κ_2 in Theorem 4.1]. Analogously to the proof of part (1) it can be shown that the first inequality (A.10) is satisfied for all $t \in [0,1]$. In order to establish the inequality (A.11) for all $t \in [0,1]$, f we consider the polynomial

$$\begin{aligned} P(t) &= \big(t+\theta_2^{(2)}\big)^4\big(\kappa_2\big(t, \xi^{b,\star}, \theta\big) - 4\big) \\ &= \alpha_{21}t^4 + \alpha_{22}t^3 + \alpha_{23}t^2 + \alpha_{24}t + \alpha_{25}, \end{aligned}$$

where the intercept and leading coefficient the are now given by $\alpha_{25} = 0$.

$$\alpha_{21} = 4\big((\theta_2^{(1)})^2(1+2\theta_2^{(1)})^2 + r(1+\theta_2^{(1)})^2(1+4\theta_2^{(1)}+20(\theta_2^{(1)})^2)\big) - 4.$$

Moreover, $P(-\theta_2^{(2)}) > 0$ [since $M^{-1}(\xi^{b,\star}, \theta)$ is positive definite] and the leading coefficient α_{21} is always positive, since $\alpha_{21}(0) = 4r - 4 > 0$ and α_{21} is increasing for $\theta_2^{(1)} \geq 0$. The roots of $P(t)$ are given by $d_1^{(2)} = 0$, $d_2^{(2)} = \theta_2^{(2)}$, and $\tilde{d}_1 = 4\theta_2^{(2)}(6 + 2\theta_2^{(1)}(1 + 2\theta_2^{(1)}))/\alpha_{21}$ where $d_2^{(2)}$ is a root of second order. Now the inequality $P(t) \leq 0$ holds for all $t \in [0, 1]$ if and only if $\tilde{d}_1 \geq 1$. It is easy to see that this condition is equivalent to (4.13).

(3) At first, one can show that condition (4.14) and $r > 1$ imply that $r \geq \frac{(1+\theta_2^{(2)})^6}{(1+\theta_2^{(1)})^6}$. The result follows by similar arguments as given in the proof of part (1), which are omitted for the sake of brevity. □

Acknowledgements. The authors would like to thank Martina Stein, who typed parts of this manuscript with considerable technical expertise. We are also grateful to Katrin Kettelhake for computational assistance and to Antoine Soubret, who built the original PK/PD model that was used to derive the candidate models for the designs calculated in Section 5. The constructive comments of the reviewers yielded to a substantial improvement of an earlier version of this paper. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

SUPPLEMENTARY MATERIAL

Supplement to "Optimal designs for dose response curves with common parameters" (DOI: 10.1214/16-AOS1520SUPP; .pdf). This supplement file contains the additional proofs omitted in the main paper and some additional comments on the derivation of the candidate models considered in Section 5.

REFERENCES

BORNKAMP, B., BRETZ, F., DMITRIENKO, A., ENAS, G., GAYDOS, B., HSU, C.-H., KÖNIG, F., KRAMS, M., LIU, Q., NEUENSCHWANDER, B., PARKE, T., PINHEIRO, J. C., ROY, A., SAX, R. and SHEN, F. (2007). Innovative approaches for designing and analyzing adaptive dose-ranging trials. *J. Biopharm. Statist.* **17** 965–995.

BRETZ, F., DETTE, H. and PINHEIRO, J. (2010). Practical considerations for optimal designs in clinical dose finding studies. *Stat. Med.* **29** 731–742. MR2752038

BRETZ, F., PINHEIRO, J. C. and BRANSON, M. (2005). Combining multiple comparisons and modeling techniques in dose-response studies. *Biometrics* **61** 738–748. MR2196162

CHALONER, K. and VERDINELLI, I. (1995). Bayesian experimental design: A review. *Statist. Sci.* **10** 273–304. MR1390519

CHERNOFF, H. (1953). Locally optimal designs for estimating parameters. *Ann. Math. Stat.* **24** 586–602.

COOK, R. D. and WONG, W. K. (1994). On the equivalence of constrained and compound optimal designs *J. Amer. Statist. Assoc.* **89** 687–692.

DETTE, H. (1990). A generalization of D- and D_1-optimal designs in polynomial regression. *Ann. Statist.* **18** 1784–1805. MR1074435

DETTE, H. (1997). Designing experiments with respect to "standardized" optimality criteria. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **59** 97–110. MR1436556

DETTE, H. and MELAS, V. B. (2011). A note on the de la Garza phenomenon for locally optimal designs. *Ann. Statist.* **39** 1266–1281. MR2816354

DETTE, H., BRETZ, F., PEPELYSHEV, A. and PINHEIRO, J. (2008). Optimal designs for dose-finding studies. *J. Amer. Statist. Assoc.* **103** 1225–1237. MR2462895

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., FEDOROV, V. V. and WU, Y. (2008). Two-stage design for dose-finding that accounts for both efficacy and safety. *Stat. Med.* **27** 5156–5176. MR2516748

DRAGALIN, V., HSUAN, F. and PADMANABHAN, S. K. (2007). Adaptive designs for dose-finding studies based on the sigmoid emax model. *J. Biopharm. Statist.* **17** 1051–1070.

FANG, X. and HEDAYAT, A. S. (2008). Locally D-optimal designs based on a class of composed models resulted from blending $E_{\max}$ and one-compartment models. *Ann. Statist.* **36** 428–444. MR2387978

FELLER, C., SCHORNING, K., DETTE, H., BERMANN, G. and BORNKAMP, B. (2017). Supplement to "Optimal designs for dose response curves with common parameters." DOI:10.1214/16-AOS1520SUPP.

GABRIELSSON, J. and WEINER, D. (2007). *Pharmacokinetic and Pharmacodynamic Data Analysis*: *Concepts and Applications*, 4th ed. Swedish Pharmaceutical Press, Stockholm.

GILBERT, P. B. (2010). Some design issues in phase 2B vs phase 3 prevention trials for testing efficacy of products or concepts. *Stat. Med.* **29** 1061–1071. MR2756810

GRIEVE, A. P. and KRAMS, M. (2005). ASTIN: A Bayesian adaptive dose-response trial in acute stroke. *Clin. Trials* **2** 340–351.

HARVILLE, D. A. (1997). *Matrix Algebra from a Statistician's Perspective*. Springer, New York. MR1467237

KARLIN, S. and STUDDEN, W. J. (1966). *Tchebycheff Systems*: *With Applications in Analysis and Statistics*. *Pure and Applied Mathematics*, *Vol. XV*. Wiley, New York. MR0204922

KIEFER, J. (1974). General equivalence theory for optimum designs (approximate theory) *Ann. Statist.* **2** 849–879. MR0356386

PRONZATO, L. and WALTER, E. (1985). Robust experiment design via stochastic approximation. *Math. Biosci.* **75** 103–120. MR0800967

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

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. *J. Am. Med. Assoc.* **311** 378–384.

SILVEY, S. D. (1980). *Optimal Design*. Chapman & Hall, London. MR0606742

THOMAS, N. (2006). Hypothesis testing and Bayesian estimation using a sigmoid $E_{\max}$ model applied to sparse dose-response designs. *J. Biopharm. Statist.* **16** 657–677. MR2252313

THOMAS, N., SWEENEY, K. and SOMAYAJI, V. (2014). Meta-analysis of clinical dose-response in a large drug development portfolio. *Stat. Biopharm. Res.* **6** 302–317.

TSAI, M.-H. and ZEN, M.-M. (2004). Criterion-robust optimal designs for model discrimination and parameter estimation: Multivariate polynomial regression case. *Statist. Sinica* **14** 591–601. MR2059298

YANG, M. (2010). On the de la Garza phenomenon. *Ann. Statist.* **38** 2499–2524. MR2676896

YANG, M. and STUFKEN, J. (2012). Identifying locally optimal designs for nonlinear models: A simple extension with profound consequences. *Ann. Statist.* **40** 1665–1681. MR3015039

ZEN, M. M. and TSAI, M. H. (2004). Criterion-robust optimal designs for model discrimination and parameter estimation in Fourier regression models. *J. Statist. Plann. Inference* **124** 475–487.

C. FELLER
G. BERMANN
B. BORNKAMP
BIOSTATISTICAL SCIENCES & PHARMACOMETRICS
NOVARTIS PHARMA AG
4002 BASEL
SWITZERLAND
E-MAIL: chrystel.feller@novartis.com
georgina.bermann@novartis.com
bjoern.bornkamp@novartis.com

K. SCHORNING
H. DETTE
FAKULTÄT FÜR MATHEMATIK
RUHR-UNIVERSITÄT BOCHUM
44780 BOCHUM
GERMANY
E-MAIL: kirsten.schorning@ruhr-uni-bochum.de
holger.dette@ruhr-uni-bochum.de