*Biometrika* (2015), **102**, 4, *pp.* 937–950
*Printed in Great Britain*
doi: 10.1093/biomet/asv041
Advance Access publication 4 November 2015

# Designing dose-finding studies with an active control for exponential families

By H. DETTE, K. KETTELHAKE

*Ruhr-Universität Bochum, Fakultät für Mathematik, 44780 Bochum, Germany*
holger.dette@ruhr-uni-bochum.de Katrin.Kettelhake@rub.de

and F. BRETZ

*Statistical Methodology, Novartis Pharma AG, 4002 Basel, Switzerland*
frank.bretz@novartis.com

## Summary

Optimal design of dose-finding studies with an active control has only been considered in the literature for regression models with normally distributed errors and known variances, where the focus is on estimating the smallest dose that achieves the same treatment effect as the active control. This paper discusses such dose-finding studies from a broader perspective. We consider a general class of optimality criteria and models arising from an exponential family. Optimal designs are constructed for several situations and their efficiency is illustrated with examples.

*Some key words*: Active control; Dose-finding; Dose-response study; Optimal design.

## 1. Introduction

Dose-finding studies are important for investigating the effect of a compound on a response and have applications in various fields. They are of particular importance in drug development, because marketed doses have to be safe and provide clinically relevant efficacy (Ting, 2006). In most of the statistical literature on dose-response studies, a placebo is included as a control group (Bretz et al., 2008), and numerous authors have worked on optimal designs in such applications because the use of efficient designs can substantially increase the accuracy of statistical analysis (Dragalin et al., 2007; Miller et al., 2007; Bornkamp et al., 2011). Dose-response studies that include a marketed drug as an active control are becoming more popular, especially in preparation for active-controlled confirmatory noninferiority trials, where the use of a placebo may be unethical. Hence, there is growing interest in active-controlled studies. For example, Helms et al. (2015) investigated the finite-sample properties of maximum likelihood estimators of the target dose in an active-controlled study which achieves the same treatment effect as the active control. However, optimal design problems for such studies have only been considered in one paper (Dette et al., 2014). These authors investigated optimal designs for estimating the target dose under the assumption of a normal distribution with known variances. In particular, they demonstrated the superiority of the optimal designs over standard designs used in pharmaceutical practice.

In this paper we investigate optimal design problems for dose-finding studies with an active control from a more general perspective. First, we consider a general class of optimality criteria. Second, we study exponential families for modelling the distribution of the responses of the new

drug and the active control, as in practice the assumption of normally distributed responses is often hard to justify. This enables the design of experiments for active-controlled studies with discrete data, as motivated by the consulting projects described in the next paragraph. Third, we demonstrate that even when the assumption of a normal distribution is justifiable, the estimation of the variances has a nontrivial effect on the optimal designs for an active-controlled study.

The research in the present paper is motivated by two examples where the assumption of normally distributed responses made by Dette et al. (2014) is hard to justify. The first example involves a 24-week dose-ranging Phase II study in patients with gouty arthritis to determine the target dose of a compound in preventing signs and symptoms of flares in chronic gout patients starting allopurinol therapy. The primary endpoint is the number of flares occurring per subject within 16 weeks of randomization, which is modelled using a negative binomial distribution for all treatment arms. The second example is a Phase IIb multicentre, randomized, double-blind, active-controlled dose-finding study in the treatment of acute migraine, as measured by the percentage of patients reporting freedom from pain at two hours post-dose.

For brevity, in this paper we restrict our attention to locally optimal designs (Chernoff, 1953). Following Chaloner and Verdinelli (1995) and Dette (1997), the methodology introduced in the present paper can be further developed to address uncertainty in the preliminary information for the unknown parameters.

## 2. MODELLING ACTIVE-CONTROLLED STUDIES USING EXPONENTIAL FAMILIES

Consider a clinical trial in which patients are treated either with an active control, i.e., a standard treatment administered at a fixed dose level, or with a new drug using different dose levels in order to investigate a dose-response relationship. Let $n_1$ and $n_2 = N - n_1$ denote the numbers of patients randomized to the new drug and to the active control, respectively. We determine the optimal number of different dose levels for the new drug, the dose levels themselves, and the optimal number of patients allocated to each dose level.

More formally, we assume that $k$ different dose levels, $d_1, \ldots, d_k$, are chosen in a dose range, $\mathcal{D} \subset \mathbb{R}_0^+$, for the new drug, and that at each dose level $d_i$ the experimenter can investigate $n_{1i}$ patients ($i = 1, \ldots, k$), where $n_1 = \sum_{i=1}^k n_{1i}$. The optimal numbers $n_{1i}$, or more precisely the optimal proportions $n_{1i}/n_1$, will be determined by the design. The responses are modelled as realizations of independent real-valued random variables $Y_{ij}$ ($j = 1, \ldots, n_{1i}; i = 1, \ldots, k$). Similarly, the responses of patients treated with the active control are modelled as realizations of independent real-valued random variables $Z_1, \ldots, Z_{n_2}$, where all responses are assumed to be independent. We further assume that the random variables $Z_j$ and $Y_{ij}$ have distributions from an exponential family, with the densities of the $Y$ variables depending on the dose levels $d_i$:

$$f_1(y \mid d_i, \theta_1) = \exp\{c_1^{\mathrm{T}}(d_i, \theta_1) T_1(y) - b_1(d_i, \theta_1)\}\, h_1(y), \tag{1}$$

$$f_2(z \mid \theta_2) = \exp\{c_2^{\mathrm{T}}(\theta_2) T_2(z) - b_2(\theta_2)\}\, h_2(z). \tag{2}$$

Here, $\theta_1 \in \Theta_1 \subset \mathbb{R}^{s_1}$ and $\theta_2 \in \Theta_2 \subset \mathbb{R}^{s_2}$ are unknown parameters, and we use common terminology for exponential families (Brown, 1986). In particular, the functions $c_1 : \mathcal{D} \times \Theta_1 \to \mathbb{R}^{\ell_1}$, $b_1 : \mathcal{D} \times \Theta_1 \to \mathbb{R}$, $c_2 : \Theta_2 \to \mathbb{R}^{\ell_2}$ and $b_2 : \Theta_2 \to \mathbb{R}$ are assumed to be twice continuously differentiable, with $\partial c_1/\partial\theta_1, \partial c_2/\partial\theta_2 \not\equiv 0$, and $T_1$ and $T_2$ denote $\ell_1$- and $\ell_2$-dimensional statistics defined on the corresponding sample spaces. The functions $h_1$ and $h_2$ are assumed to be positive and measurable.

Throughout the paper, let $\kappa$ denote a variable indicating whether a patient receives the new drug, $\kappa = 0$, or the active control, $\kappa = 1$. Further, let

$$\mathcal{X} = (\mathcal{D} \times \{0\}) \cup \{(C, 1)\} \tag{3}$$

denote the design space of the experiment, where $\mathcal{D}$ is the dose range for the new drug, $C$ is the dose level of the active control, and the second component of an experimental condition $(d, \kappa) \in \mathcal{X}$ determines the treatment, $\kappa = 0$ or 1. The Fisher information matrix at the point $(d, \kappa) \in \mathcal{X}$ is

$$I\{(d, \kappa), \theta\} = \begin{bmatrix} 1_{\{\kappa=0\}} I_1(d, \theta_1) & 0 \\ 0 & 1_{\{\kappa=1\}} I_2(\theta_2) \end{bmatrix}, \tag{4}$$

where $0$ denotes a matrix of appropriate dimension with all entries equal to 0, $\theta = (\theta_1^{\mathrm{T}}, \theta_2^{\mathrm{T}})^{\mathrm{T}} \in \Theta_1 \times \Theta_2 \subset \mathbb{R}^{s_1+s_2}$ is the vector of all parameters, $1_{\{\kappa=0\}}$ is the indicator function of the event $\{\kappa = 0\}$, and the matrices $I_1$ and $I_2$ are the Fisher information matrices of the two models (1) and (2), that is,

$$\begin{aligned} I_1(d, \theta_1) &= E\left[\left\{\frac{\partial}{\partial \theta_1} \log f_1(Y \mid d, \theta_1)\right\}\left\{\frac{\partial}{\partial \theta_1} \log f_1(Y \mid d, \theta_1)\right\}^{\mathrm{T}}\right], \\ I_2(\theta_2) &= E\left[\left\{\frac{\partial}{\partial \theta_2} \log f_2(Z \mid \theta_2)\right\}\left\{\frac{\partial}{\partial \theta_2} \log f_2(Z \mid \theta_2)\right\}^{\mathrm{T}}\right], \end{aligned} \tag{5}$$

where the random variables $Y$ and $Z$ have densities (1) and (2). The Fisher information matrix in (4) is block-diagonal because of the independence of the samples, given that a patient is randomized to either the new drug or the active control.

*Example* 1. To demonstrate the different structures of the Fisher information matrix arising from different distributions of the exponential family, we consider several examples. Here we restrict our attention to the normal and negative binomial distributions. Further examples can be found in the Supplementary Material.

Dette et al. (2014) investigated normally distributed responses with known variances $\sigma_1^2$ for the new drug and $\sigma_2^2$ for the active control. For the expectation of the response of the new drug at dose level $d$ they assumed a nonlinear regression model, $\eta(d, \vartheta)$ where $\vartheta = (\vartheta_0, \ldots, \vartheta_s)^{\mathrm{T}}$, while the expectation of the response of the active control is assumed to equal $\mu$ for the active control. If the variances must be estimated from the data, we have $\theta_1 = (\vartheta_0, \ldots, \vartheta_s, \sigma_1^2)^{\mathrm{T}}$ and $\theta_2 = (\mu, \sigma_2^2)^{\mathrm{T}}$ for the parameters in models (1) and (2), respectively. The Fisher information matrix at a point $(d, \kappa) \in \mathcal{X}$ is given by (4) with

$$I_1(d, \theta_1) = \begin{bmatrix} \dfrac{1}{\sigma_1^2}\left\{\dfrac{\partial}{\partial \vartheta}\eta(d, \vartheta)\right\}\left\{\dfrac{\partial}{\partial \vartheta}\eta(d, \vartheta)\right\}^{\mathrm{T}} & 0 \\ 0 & \dfrac{1}{2\sigma_1^4} \end{bmatrix}, \quad I_2(\theta_2) = \begin{pmatrix} \dfrac{1}{\sigma_2^2} & 0 \\ 0 & \dfrac{1}{2\sigma_2^4} \end{pmatrix}.$$

Next, assume a negative binomial distribution with parameter $r_1 \in \mathbb{N}$ and a function $\pi(d, \theta_1) \in (0, 1)$ for the probability of a success of the new drug at dose level $d$, and parameters $r_2 \in \mathbb{N}$ and

$\mu \in (0, 1)$ for the active control. Then we have $\theta_2 = \mu$, and the Fisher information matrix is given by (4) with

$$I_1(d, \theta_1) = \frac{r_1 \left\{ \frac{\partial}{\partial \theta_1} \pi(d, \theta_1) \right\} \left\{ \frac{\partial}{\partial \theta_1} \pi(d, \theta_1) \right\}^{\mathrm{T}}}{\pi^2(d, \theta_1)\{1 - \pi(d, \theta_1)\}}, \quad I_2(\theta_2) = \frac{r_2}{\mu^2(1-\mu)}.$$

Here the parameters $r_1, r_2 \in \mathbb{N}$ are assumed to be known.

Throughout this paper we consider approximate designs in the sense of Kiefer (1974), which are defined as probability measures with finite support on the design space $\mathcal{X}$ in (3). Therefore, an experimental design is given by

$$\xi = \begin{Bmatrix} (d_1, 0) & \dots & (d_k, 0) & (C, 1) \\ w_1 & \dots & w_k & w_{k+1} \end{Bmatrix}, \tag{6}$$

where $w_1, \dots, w_{k+1}$ are positive weights such that $\sum_{i=1}^{k+1} w_i = 1$. Here, $w_i$ denotes the relative proportion of patients treated at dose level $d_i$ $(i = 1, \dots, k)$ or, for $i = k + 1$, the active control. If $N$ observations are taken, a rounding procedure is applied to obtain integers $n_{1i}$ $(i = 1, \dots, k)$ and $n_2$ from the possibly real-valued quantities $w_i N$ $(i = 1, \dots, k + 1)$. The information matrix of an approximate design $\xi$ of the form (6) is defined by the $(s_1 + s_2) \times (s_1 + s_2)$ matrix

$$M(\xi, \theta) = \int_{\mathcal{X}} I\{(d, \kappa), \theta\} \, \mathrm{d}\xi(d, \kappa) = \begin{bmatrix} (1 - \omega_{k+1}) M_1(\tilde{\xi}, \theta_1) & 0 \\ 0 & \omega_{k+1} I_2(\theta_2) \end{bmatrix}. \tag{7}$$

Here, the $s_1 \times s_1$ matrix $M_1(\tilde{\xi}, \theta_1)$ and the $s_2 \times s_2$ matrix $I_2(\theta_2)$ are given by

$$M_1(\tilde{\xi}, \theta_1) = \int_{\mathcal{D}} I_1(d, \theta_1) \, \mathrm{d}\tilde{\xi}(d)$$

and (5), respectively, where

$$\tilde{\xi} = \begin{pmatrix} d_1 & \dots & d_k \\ \tilde{w}_1 & \dots & \tilde{w}_k \end{pmatrix} \tag{8}$$

denotes the design on the design space $\mathcal{D}$ for the new drug, which is induced by the design $\xi$ in (6) defining the weights $\tilde{w}_i = w_i(1 - w_{k+1})^{-1}$ $(i = 1, \dots, k)$. If observations are taken according to an approximate design, then the covariance of the maximum likelihood estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ in models (1) and (2) is given approximately by $M^{-1}(\xi, \theta)/N$, and consequently optimal designs maximize an appropriate functional of the information matrix defined in (7).

In order to discriminate between competing designs, we consider Kiefer's $\phi_p$-criteria (Kiefer, 1974). To be precise, let $p \in [-\infty, 1)$ and let $K \in \mathbb{R}^{(s_1+s_2)\times t}$ denote a matrix of full column rank $t$. Then a design $\xi^*$ is said to be locally $\phi_p$-optimal for estimating the linear combination $K^{\mathrm{T}}\theta$ in a dose-response model with an active control if $K^{\mathrm{T}}\theta$ is estimable by the design $\xi^*$, i.e., $K^{\mathrm{T}}\theta \in \mathrm{range}\{M(\xi^*, \theta)\}$, and $\xi^*$ maximizes the functional

$$\phi_p(\xi) = \left( \frac{1}{t} \mathrm{tr}\left[ \{K^{\mathrm{T}} M^{-}(\xi, \theta) K\}^{-p} \right] \right)^{1/p} \tag{9}$$

among all designs for which $K^{\mathrm{T}}\theta$ is estimable, where $\mathrm{tr}(A)$ and $A^{-}$ denote the trace and a generalized inverse of the matrix $A$, respectively. The cases $p=0$ and $p=-\infty$ correspond to the D- and E-optimality criteria, i.e., $\phi_0(\xi)=\det[\{K^{\mathrm{T}}M^{-}(\xi,\theta)K\}^{-1/t}]$ and $\phi_{-\infty}(\xi)=\lambda_{\min}[\{K^{\mathrm{T}}M^{-}(\xi,\theta)K\}^{-1}]$, where $\lambda_{\min}(A)$ denotes the smallest eigenvalue of the matrix $A$. An application of the general equivalence theorem (Pukelsheim, 2006, §§ 7.19 and 7.21) yields the following result.

LEMMA 1. *If $p\in(-\infty,1)$, a design $\xi^*$ with $K^{\mathrm{T}}\theta\in\mathrm{range}\{M(\xi^*,\theta)\}$ is locally $\phi_p$-optimal for estimating the linear combination $K^{\mathrm{T}}\theta$ in a dose-response model with an active control if and only if there exists a generalized inverse $G$ of the information matrix $M(\xi^*,\theta)$ such that the inequality*

$$\mathrm{tr}\big[I\{(d,\kappa),\theta\}GK\{K^{\mathrm{T}}M^{-}(\xi^*,\theta)K\}^{-p-1}K^{\mathrm{T}}G^{\mathrm{T}}\big]\leqslant\mathrm{tr}\big[\{K^{\mathrm{T}}M^{-}(\xi^*,\theta)K\}^{-p}\big] \tag{10}$$

*holds for all $(d,\kappa)\in\mathcal{X}$.*

*If $p=-\infty$, a design $\xi^*$ with $K^{\mathrm{T}}\theta\in\mathrm{range}\{M(\xi^*,\theta)\}$ is locally $\phi_{-\infty}$-optimal for estimating the linear combination $K^{\mathrm{T}}\theta$ if and only if there exist a generalized inverse $G$ of the information matrix $M(\xi^*,\theta)$ and a nonnegative-definite matrix $E\in\mathbb{R}^{t\times t}$ with $\mathrm{tr}(E)=1$ such that the inequality*

$$\begin{aligned}\mathrm{tr}\big[I\{(d,\kappa),\theta\}GK\{K^{\mathrm{T}}M^{-}(\xi^*,\theta)K\}^{-1}E\{K^{\mathrm{T}}M^{-}(\xi^*,\theta)K\}^{-1}K^{\mathrm{T}}G^{\mathrm{T}}\big]\\ \leqslant\lambda_{\min}\big[\{K^{\mathrm{T}}M^{-}(\xi^*,\theta)K\}^{-1}\big]\end{aligned} \tag{11}$$

*holds for all $(d,\kappa)\in\mathcal{X}$.*

*Moreover, there is equality in (10) and (11) for all support points of the design $\xi^*$.*

Below we assume that either $p=-1$ or the matrix $K$ has the form

$$K=\begin{pmatrix}K_{11}&0\\0&K_{22}\end{pmatrix}\in\mathbb{R}^{(s_1+s_2)\times(t_1+t_2)} \tag{12}$$

with elements $K_{11}\in\mathbb{R}^{s_1\times t_1}$ and $K_{22}\in\mathbb{R}^{s_2\times t_2}$, such that $t_1+t_2=t$. Roughly speaking, the choice $p=-1$ or a block-diagonal structure of the matrix $K$ in (12) leads to a separation of the parameters from models (1) and (2) in the corresponding optimality criterion. Hence, optimal designs for dose-finding studies with an active control can be obtained from optimal designs for dose-finding studies that include a placebo group, which maximize the criterion

$$\tilde{\phi}_p(\tilde{\xi})=\left(\frac{1}{t_1}\mathrm{tr}\big[\{K_{11}^{\mathrm{T}}M_1^{-}(\tilde{\xi},\theta_1)K_{11}\}^{-p}\big]\right)^{1/p} \tag{13}$$

in the class of all designs $\tilde{\xi}$ for which $K_{11}^{\mathrm{T}}\theta_1$ is estimable, i.e., $K_{11}^{\mathrm{T}}\theta_1\in\mathrm{range}\{M_1(\tilde{\xi},\theta_1)\}$. We call these designs $\tilde{\phi}_p$-optimal for estimating the parameter $K_{11}^{\mathrm{T}}\theta_1$ in the dose-response model (1). The proof can be found in the Appendix.

THEOREM 1. *Suppose that $p\in[-\infty,1)$, the matrix $K$ is given by (12), and*

$$\tilde{\xi}_p^*=\begin{pmatrix}d_1^*&\dots&d_k^*\\\tilde{w}_1^*&\dots&\tilde{w}_k^*\end{pmatrix} \tag{14}$$

*is a locally $\tilde{\phi}_p$-optimal design for estimating $K_{11}^{\mathrm{T}}\theta_1$ in (1). Then the design*

$$\xi_p^* = \begin{Bmatrix} (d_1^*, 0) & \dots & (d_k^*, 0) & (C, 1) \\ w_1^* & \dots & w_k^* & w_{k+1}^* \end{Bmatrix}$$

*is locally $\phi_p$-optimal for estimating $K^{\mathrm{T}}\theta$ in the dose-response model with an active control, where the weights are*

$$w_{k+1}^* = \frac{1}{1+\rho_p}, \quad w_i^* = \frac{\rho_p}{1+\rho_p}\tilde{w}_i^* \quad (i = 1, \dots, k), \tag{15}$$

*with*

$$\rho_p = \frac{(\mathrm{tr}[\{K_{22}^{\mathrm{T}} I_2^-(\theta_2) K_{22}\}^{-p}])^{1/(p-1)}}{(\mathrm{tr}[\{K_{11}^{\mathrm{T}} M_1^-(\tilde{\xi}_p^*, \theta_1) K_{11}\}^{-p}])^{1/(p-1)}}. \tag{16}$$

*The case $p = -\infty$ is interpreted as the corresponding limit.*

In the $p = -1$ case a more general statement is available, without the restriction to block matrices of the form (12). The proof is obtained using similar arguments to those in the proof of Theorem 1; we therefore omit it.

THEOREM 2. *Assume that $K^{\mathrm{T}} = (K_{11}^{\mathrm{T}}, K_{22}^{\mathrm{T}}) \in \mathbb{R}^{t\times(s_1+s_2)}$ with $K_{11}^{\mathrm{T}} \in \mathbb{R}^{t\times s_1}$ and $K_{22}^{\mathrm{T}} \in \mathbb{R}^{t\times s_2}$, and let $\tilde{\xi}_{-1}^*$ denote the $\tilde{\phi}_{-1}$-optimal design for estimating the parameter $K_{11}^{\mathrm{T}}\theta_1$ in the dose-response model (1). Then the design $\xi_{-1}^*$ defined in Theorem 1 is locally $\phi_{-1}$-optimal for estimating $K^{\mathrm{T}}\theta$ in the dose-response model with an active control.*

The final result of this section concerns the special case of $p = 0$. The result is a direct consequence of Theorem 1, upon considering a corresponding limit and observing that the quantity $\rho_p$ defined in (16) satisfies $\lim_{p\to 0} \rho_p = t_1/t_2$.

COROLLARY 1. *Assume that the matrix $K$ is given by (12), and let $\tilde{\xi}_0^*$ denote the locally D-optimal design of the form (14) for estimating the parameter $K_{11}^{\mathrm{T}}\theta_1$ in the dose-response model (1), which maximizes $\det[\{K_{11}^{\mathrm{T}} M_1^-(\tilde{\xi}, \theta_1) K_{11}\}^{-1}]$ in the class of all designs for which $K_{11}^{\mathrm{T}}\theta_1$ is estimable. Then the design $\xi_0^*$ with masses $t_1(t_1+t_2)^{-1}\tilde{w}_1^*, \dots, t_1(t_1+t_2)^{-1}\tilde{w}_k^*$ and $t_2(t_1+t_2)^{-1}$ at the points $(d_1^*, 0), \dots, (d_k^*, 0)$ and $(C, 1)$, respectively, is locally D-optimal for estimating the parameter $K^{\mathrm{T}}\theta$ in the dose-response model with an active control.*

*Remark* 1. The assumption of a block matrix $K$ in Theorem 1 and Corollary 1 cannot be omitted. A counterexample is given in the Supplementary Material.

## 3. D-OPTIMAL DESIGNS FOR THE MICHAELIS–MENTEN AND $E_{\max}$ MODELS

In this section we determine some D-optimal designs under different distributional assumptions for the Michaelis–Menten model $\vartheta_1 d(\vartheta_2 + d)^{-1}$ and the $E_{\max}$ model $\vartheta_0 + \vartheta_1 d(\vartheta_2 + d)^{-1}$, where the dose range is the interval $[L, R] \subset \mathbb{R}_0^+$. Both models are widely used in practice and are therefore treated in detail. If the dose-response function describes a probability, some restrictions on the parameters are required. For example, if $\pi(d, \theta_1) = \vartheta_1 d(\vartheta_2 + d)^{-1}$ denotes the success probability for the negative binomial distribution in Example 1, then we implicitly assume that $\vartheta_1 R(\vartheta_2 + R)^{-1} < 1$ in the following discussion. In other models similar assumptions have to be made, but for brevity we do not mention them explicitly. In the following, $x \vee y$ denotes the maximum of $x, y \in \mathbb{R}$.

THEOREM 3 (Michaelis–Menten model).

(i) *If the distributions of the responses corresponding to the new drug and active control are normal with parameters* $\{\vartheta_1 d(\vartheta_2 + d)^{-1}, \sigma_1^2\}$ *and* $(\mu, \sigma_2^2)$*, respectively, then the locally D-optimal design for the dose-response model with an active control allocates* 30% *of patients to each of the dose levels* $L \vee \vartheta_2 R(2\vartheta_2 + R)^{-1}$ *and* $R$ *of the new drug and* 40% *of patients to the active control.*

(ii) *If negative binomial distributions with probabilities* $\pi(d, \theta) = \vartheta_1 d(\vartheta_2 + d)^{-1}$ *and* $\mu$ *are used, then the locally D-optimal design for the dose-response model with an active control allocates* $33{\cdot}\bar{3}\%$ *of patients to each of the dose levels* $L$ *and* $R$ *of the new drug and* $33{\cdot}\bar{3}\%$ *of patients to the active control.*

(iii) *If binomial distributions with probabilities* $\pi(d, \theta) = \vartheta_1 d(\vartheta_2 + d)^{-1}$ *and* $\mu$ *are used, then the locally D-optimal design for the dose-response model with an active control allocates* $33{\cdot}\bar{3}\%$ *of patients to each of the dose levels* $L \vee \{\vartheta_2 R + 3\vartheta_2^2 - \vartheta_2(9R^2 - 8R^2\vartheta_1 + 18R\vartheta_2 - 8R\vartheta_1\vartheta_2 + 9\vartheta_2^2)^{1/2}\}(4\vartheta_1\vartheta_2 - 4R + 4R\vartheta_1 - 6\vartheta_2)^{-1}$ *and* $R$ *of the new drug and* $33{\cdot}\bar{3}\%$ *of patients to the active control.*

(iv) *If Poisson distributions with parameters* $\lambda(d, \theta_1) = \vartheta_1 d(\vartheta_2 + d)^{-1}$ *and* $\mu$ *are used in* (1) *and* (2)*, then the locally D-optimal design for the dose-response model with an active control allocates* $33{\cdot}\bar{3}\%$ *of patients to each of the dose levels* $L \vee \vartheta_2 R(3\vartheta_2 + 2R)^{-1}$ *and* $R$ *of the new drug and* $33{\cdot}\bar{3}\%$ *of patients to the active control.*

The proof of Theorem 3 follows directly from Corollary 1, if the locally D-optimal designs for model (1) are known. For example, in the case of a normal distribution, it follows from Dette et al. (2010) that the D-optimal design for the Michaelis–Menten model has equal masses at the points $L \vee \vartheta_2 R(2\vartheta_2 + R)^{-1}$ and $R$, which yields statement (i) of Theorem 3. In the other cases, the D-optimal designs for model (1) are not known and the proof can be found in the Appendix.

The differences in the D-optimal designs derived under different distributional assumptions can be substantial. For example, if the design space is $[0, R]$ with a large right boundary $R$, the nontrivial dose level for the new drug is approximately $\vartheta_2$ under the assumption of a normal distribution and $\vartheta_2/2$ under a Poisson distribution.

We now give the corresponding results for the $E_{\max}$ model. The proof uses similar arguments and is therefore omitted.

THEOREM 4 ($E_{\max}$ model).

(i) *If the distributions of responses corresponding to the new drug and active control are normal with parameters* $\{\vartheta_0 + \vartheta_1 d(\vartheta_2 + d)^{-1}, \sigma_1^2\}$ *and* $(\mu, \sigma_2^2)$*, respectively, then the locally D-optimal design for the dose-response model with an active control allocates* $22{\cdot}\bar{2}\%$ *of patients to each of the dose levels* $L$*,* $d^* = \{R(L + \vartheta_2) + L(R + \vartheta_2)\}\{(L + \vartheta_2) + (R + \vartheta_2)\}^{-1}$ *and* $R$ *of the new drug and* $33{\cdot}\bar{3}\%$ *of patients to the active control.*

(ii) *If negative binomial distributions with probabilities* $\pi(d, \theta) = \vartheta_0 + \vartheta_1 d(\vartheta_2 + d)^{-1}$ *and* $\mu$ *are used, then the locally D-optimal design for the dose-response model with an active control allocates* 25% *of patients to each of the dose levels* $L$*,* $d^*$ *and* $R$ *of the new drug and* 25% *of patients to the active control, where* $d^*$ *is the solution of the equation*

$$\frac{2}{d - L} + \frac{2}{d - R} - \frac{\vartheta_0 + \vartheta_1 - 1}{d(\vartheta_0 + \vartheta_1 - 1) + (\vartheta_0 - 1)\vartheta_2} - \frac{2(\vartheta_0 + \vartheta_1)}{\vartheta_0(\vartheta_2 + d) + \vartheta_1 d} - \frac{1}{\vartheta_2 + d} = 0.$$

Table 1. *D-optimal designs in the two examples under different distributional assumptions, together with the efficiencies of the designs which were actually used in the study*

| | Gouty arthritis example | | | | | Acute migraine example | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Distribution | D-optimal design | | | | $\text{eff}_D$ | D-optimal design | | | | $\text{eff}_D$ |
| Normal | (0, 0) | (9·81, 0) | (300, 0) | ($C$, 1) | 0·25 | (0, 0) | (10·95, 0) | (200, 0) | ($C$, 1) | 0·84 |
| | 22·2̄% | 22·2̄% | 22·2̄% | 33·3̄% | | 22·2̄% | 22·2̄% | 22·2̄% | 33·3̄% | |
| nB/Bin | (0, 0) | (8·23, 0) | (300, 0) | ($C$, 1) | 0·11 | (0, 0) | (9·05, 0) | (200, 0) | ($C$, 1) | 0·86 |
| | 25% | 25% | 25% | 25% | | 25% | 25% | 25% | 25% | |

nB, negative binomial; Bin, binomial.

(iii) *If binomial distributions with probabilities* $\pi(d,\theta) = \vartheta_0 + \vartheta_1 d(\vartheta_2 + d)^{-1}$ *and* $\mu$ *are used, then the locally D-optimal design is of the same form as in* (ii)*, where* $d^*$ *is the solution of the equation*

$$\frac{2}{d-L} + \frac{2}{d-R} - \frac{\vartheta_0 + \vartheta_1 - 1}{d(\vartheta_0 + \vartheta_1 - 1) + (\vartheta_0 - 1)\vartheta_2} - \frac{\vartheta_0 + \vartheta_1}{\vartheta_0(\vartheta_2 + d) + \vartheta_1 d} - \frac{2}{\vartheta_2 + d} = 0.$$

(iv) *If Poisson distributions with parameters* $\lambda(d,\theta_1) = \vartheta_0 + \vartheta_1 d(\vartheta_2 + d)^{-1}$ *and* $\mu$ *are used in* (1) *and* (2)*, respectively, then the locally D-optimal design is of the same form as in* (ii)*, where*

$$d^* = \vartheta_2 \frac{4m(L)m(R) - \vartheta_1\{Lm(R) + Rm(L)\} - \vartheta_0\sqrt{\iota}}{-4m(L)m(R) - \vartheta_1\vartheta_2\{m(R) + m(L)\} + (\vartheta_1 + \vartheta_0)\sqrt{\iota}},$$

*with* $\iota = \{(\vartheta_2 + L)m(R) + (\vartheta_2 + R)m(L)\}^2 + 12(\vartheta_2 + L)(\vartheta_2 + R)m(R)m(L)$ *and* $m(d) = \vartheta_0\vartheta_2 + \vartheta_1 d + \vartheta_0 d$.

*Example* 2. We now discuss D-optimal designs for the two clinical trials described in § 1.

First let us consider the gouty arthritis example. The primary endpoint is modelled by a negative binomial distribution with parameters $r_1$ and $\pi(d,\theta_1) = \vartheta_0 + \vartheta_1 d(\vartheta_2 + d)^{-1}$ for the new drug and parameters $r_2$ and $\theta_2$ for the comparator. The dose range is [0, 300] mg, and we obtained from the clinical team the following preliminary information on the unknown parameters: $\vartheta_0 = 0{\cdot}26$, $\vartheta_1 = 0{\cdot}73$, $\vartheta_2 = 10{\cdot}5$, $\sigma_1 = 0{\cdot}05$, $\theta_2 = 0{\cdot}9206$ and $\sigma_2 = 0{\cdot}05$. In addition, $r_1 = r_2 = 10$ are fixed. The D-optimal design is obtained from Theorem 4 and summarized in Table 1. The standard design actually used in this study allocates 14·3% of patients to each of the dose levels 25, 50, 100, 200 and 300 mg of the new drug and 28·5% of patients to the active control. To compare these designs, we also show in Table 1 the D-efficiency $\text{eff}_D(\xi,\theta) = \phi_0(\xi,\theta)/\phi_0(\xi_D^*,\theta)$, where $\xi_D^*$ is the locally D-optimal design. We observe that, in this example, the optimal design is substantially more efficient than the standard design.

Next, consider the acute migraine example, which measured the percentage of patients reported to be free of pain two hours post-dose. We assume a binomial distribution for this case. The probabilities of success are $\pi(d,\theta_1) = \vartheta_0 + \vartheta_1 d(\vartheta_2 + d)^{-1}$ for the new compound, where the dose level varies in the interval [0, 200] mg, and $\theta_2$ for the active control. The sample

sizes are $n_1 = 517$ and $n_2 = 100$, and the preliminary information obtained from the clinical team consists of $\vartheta_0 = 0{\cdot}098$, $\vartheta_1 = 0{\cdot}2052$, $\vartheta_2 = 12{\cdot}3$, $\sigma_1 = 0{\cdot}05$, $\theta_2 = 0{\cdot}2505$ and $\sigma_2 = 0{\cdot}05$. The locally D-optimal designs under normal and binomial distributional assumptions are listed in Table 1. The design actually used for this study allocated 21, 5, 7, 10, 10, 11, 10 and 10% of the patients to dose levels 0, 2·5, 5, 10, 20, 50, 100 and 200 mg of the new drug, respectively, and 16% of the patients to the active control. Again, a substantial improvement can be observed under both distributions.

## 4. Optimal designs for estimating the target dose

### 4·1. *AC-optimal designs*

In this section we investigate the construction of locally optimal designs for estimating the smallest dose of the new compound that achieves the same treatment effect as the active control. We consider a dose range of the form $\mathcal{D} = [L, R] \subset \mathbb{R}_0^+$ and write

$$E_{\theta_1}(Y_{ij} \mid d_i) = \eta(d_i, \theta_1) \quad (j = 1, \ldots, n_{1i};\ i = 1, \ldots, k), \tag{17}$$

$$E_{\theta_2}(Z_i) = \Delta \quad (i = 1, \ldots, n_2), \tag{18}$$

for the expected values of responses corresponding to the new drug for dose level $d_i$ and the active control, respectively. We assume that the function $\eta$ in (17) is strictly increasing with respect to $d \in \mathcal{D}$ and that $d^*(\theta) = \eta^{-1}(\Delta, \theta_1)$ is an element of the dose range $\mathcal{D}$ for the new drug. The expectation $\Delta$ in (18) is a function of the $s_2$-dimensional parameter $\theta_2$, say $\Delta = k(\theta_2)$. Consequently, a natural estimator of $d^*$ is given by $\hat{d}^* = d^*(\hat{\theta}) = \eta^{-1}(\hat{\Delta}, \hat{\theta}_1)$, where $\hat{\Delta} = k(\hat{\theta}_2)$ and $\hat{\theta} = (\hat{\theta}_1^{\mathrm{T}}, \hat{\theta}_2^{\mathrm{T}})^{\mathrm{T}}$ denotes the vector of the maximum likelihood estimators of the parameters $\theta_1$ and $\theta_2$ in models (1) and (2), respectively. Standard calculations show that the variance of this estimator is $\mathrm{var}\{d^*(\hat{\theta})\} \approx N^{-1}\psi(\xi, \theta)$, where

$$\begin{aligned}\psi(\xi, \theta) = {} & \frac{1}{1 - \omega_{k+1}} \left\{ \frac{\partial}{\partial \theta_1} d^*(\theta) \right\}^{\mathrm{T}} M_1^-(\tilde{\xi}, \theta_1) \left\{ \frac{\partial}{\partial \theta_1} d^*(\theta) \right\} \\ & + \frac{1}{\omega_{k+1}} \left\{ \frac{\partial}{\partial \theta_2} d^*(\theta) \right\}^{\mathrm{T}} I_2^-(\theta_2) \left\{ \frac{\partial}{\partial \theta_2} d^*(\theta) \right\}. \end{aligned} \tag{19}$$

Here, $\tilde{\xi}$ denotes the design for the new drug induced by the design $\xi$, see (8), and $M_1^-(\tilde{\xi}, \theta_1)$ and $I_2^-(\theta_2)$ are generalized inverses of the information matrices $M_1(\tilde{\xi}, \theta_1)$ and $I_2(\theta_2)$, respectively. Following Dette et al. (2014), we say that a design $\xi_{\mathrm{AC}}^*$ is locally AC-optimal if $\partial d^*(\theta)/\partial\theta_1 \in \mathrm{range}\{M_1(\tilde{\xi}, \theta_1)\}$, $\partial d^*(\theta)/\partial\theta_2 \in \mathrm{range}\{I_2(\theta_2)\}$, and $\xi_{\mathrm{AC}}^*$ minimizes the function $\psi(\xi, \theta)$ among all designs satisfying the above estimability conditions. Criterion (19) corresponds to a $\phi_{-1}$-optimal design for estimating the parameter $K^{\mathrm{T}}\theta$ in a dose-response model with an active control, where $K = [\{\partial d^*(\theta)/\partial\theta_1\}^{\mathrm{T}}, \{\partial d^*(\theta)/\partial\theta_2\}^{\mathrm{T}}]^{\mathrm{T}}$. In particular, Theorem 2 is applicable and locally AC-optimal designs can be derived from the corresponding optimal designs for model (1). The following result provides an alternative representation of the criterion (19) in the case where $s_2 = 1$. As a consequence, the design $\tilde{\xi}$ required in Theorem 2 is a locally $\tilde{c}$-optimal design in model (1) for a specific vector $\tilde{c}$, i.e., the design minimizing $\tilde{c}^{\mathrm{T}} M_1^-(\tilde{\xi}, \theta_1)\tilde{c}$ where $\tilde{c} = \partial\eta(d^*, \theta_1)/\partial\theta_1$. The proof can be obtained using arguments similar to those in Dette et al. (2014) and is given in the Supplementary Material.

THEOREM 5. *In the case where $s_2 = 1$, the function in (19) can be expressed as*

$$\psi(\xi,\theta) = \frac{\left\{\frac{\partial}{\partial\theta_2}d^*(\theta)\right\}^2}{\left\{\frac{\partial}{\partial\theta_2}k(\theta_2)\right\}^2}\left[\frac{1}{1-w_{k+1}}\left\{\frac{\partial}{\partial\theta_1}\eta(d^*,\theta_1)\right\}^{\mathrm{T}} M_1^-(\tilde{\xi},\theta_1)\left\{\frac{\partial}{\partial\theta_1}\eta(d^*,\theta_1)\right\}\right.$$
$$\left. + \left\{\frac{\partial}{\partial\theta_2}k(\theta_2)\right\}^2 \frac{I_2^-(\theta_2)}{w_{k+1}}\right].$$

### 4·2. *Some explicit results for two-dimensional models*

In this subsection we present some examples illustrating different structures of locally AC-optimal designs. We suppose that the Fisher information matrix $I_1(d,\theta_1)$ defined in (5) is of the form $I_1(d,\theta_1) = \mathrm{diag}\{f(d,\theta_1)f^{\mathrm{T}}(d,\theta_1), \Sigma(\theta_1)\} \in \mathbb{R}^{s_1\times s_1}$, where $f(d,\theta_1) = \{f_1(d,\theta_1), f_2(d,\theta_1)\}^{\mathrm{T}}$ denotes a two-dimensional vector and $\Sigma(\theta_1)$ a $(s_1-2)\times(s_1-2)$ matrix, which does not depend on the dose level. By Theorem 2, the locally AC-optimal design can be determined from the design $\tilde{\xi}^*$ that minimizes the expression $\tilde{c}^{\mathrm{T}}M_1^-(\tilde{\xi},\theta_1)\tilde{c}$ in the class of all designs defined on the dose range $\mathcal{D}$ for the new drug, where the vector $\tilde{c}$ is given by $\tilde{c} = \partial d^*(\theta)/\partial\theta_1$. Because of the block structure of the Fisher information matrix $I_1(d,\theta_1)$, with a lower block not depending on the dose level, we may assume without loss of generality that $s_1 = 2$, i.e.,

$$M_1(\tilde{\xi},\theta_1) = \int_{\mathcal{D}} f(d,\theta_1)f^{\mathrm{T}}(d,\theta_1)\,\mathrm{d}\tilde{\xi}(d). \tag{20}$$

By Elfving's theorem (Elfving, 1952) a design $\tilde{\xi}^*$ with weights $\tilde{w}_i^*$ at the points $d_i^*$ $(i = 1,\ldots,k)$ minimizes $\tilde{c}^{\mathrm{T}}M_1^-(\tilde{\xi},\theta_1)\tilde{c}$ if and only if there exist a constant $\gamma > 0$ and $\varepsilon_1,\ldots,\varepsilon_k \in \{-1,1\}$ such that the point $\gamma\tilde{c}$ is a boundary point of the Elfving set

$$\mathcal{R} = \mathrm{conv}\{\varepsilon f(d,\theta_1) : d\in\mathcal{D}, \varepsilon\in\{-1,1\}\} \tag{21}$$

and the representation $\gamma\tilde{c} = \sum_{i=1}^k \varepsilon_i\tilde{w}_i^* f(d_i^*,\theta_1)$ is valid. Note that $\mathcal{R} = \mathrm{conv}\{\mathcal{C}\cup(-\mathcal{C})\}$, where the curve $\mathcal{C}$ is defined by $\mathcal{C} = \{f(d,\theta_1) : d\in\mathcal{D}\}$. The structure of the Elfving set $\mathcal{R}$ depends sensitively on the distributional assumptions, and we now consider several examples in the context of the Michaelis–Menten model.

*Example* 3. If the dependence on the dose in model (1) is described by the Michaelis–Menten model, then the vector $f$ in (20) has the form $v(d,\theta_1)\{d(\vartheta_2+d)^{-1}, -\vartheta_1 d(\vartheta_2+d)^{-2}\}^{\mathrm{T}}$, where the function $v$ varies with the distributional assumption.

If the responses are normally distributed, we have $v(d,\theta_1) = 1$, and it follows by an obvious generalization of Theorem 5 that we must consider a $\tilde{c}$-optimal design problem in model (1), where the vector $\tilde{c}$ is now given by $\tilde{c} = \partial\eta(d^*,\vartheta)/\partial\vartheta = \{d^*(\vartheta_2+d^*)^{-1}, -\vartheta_1 d^*(\vartheta_2+d^*)^{-2}\}^{\mathrm{T}}$. From Fig. 1(a) we see that the line $\{\gamma\tilde{c} : \gamma > 0\}$ intersects the boundary of the Elfving set $\mathcal{R}$ at some point $\mathcal{C}\cup(-\mathcal{C})$ whenever $L \leqslant x^* \leqslant d^* < R$, where

$$x^* = L \vee \frac{\sqrt{2}R^2\vartheta_2 + (\sqrt{2}-1)R\vartheta_2^2}{2R^2 + 4R\vartheta_2 + \vartheta_2^2}.$$

A typical situation is shown for the vector $\tilde{c}_2$ in Fig. 1(a) for $\vartheta_1 = \vartheta_2 = 2$ and $\mathcal{D} = [0{\cdot}1, 50]$. Consequently, Elfving's theorem shows that a one-point design minimizes $\tilde{c}^{\mathrm{T}}M_1^-(\tilde{\xi},\theta_1)\tilde{c}$. An application of Theorem 2 yields a locally AC-optimal design which allocates $\sigma_1$

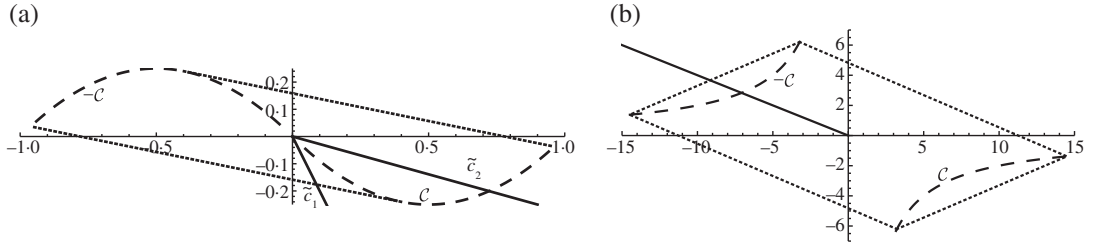


Fig. 1. The Elfving set (21) in model (1), where the expected response is given by the Michaelis–Menten model: (a) normal distribution; (b) negative binomial distribution.

$(\sigma_1 + \sigma_2)^{-1}100\%$ of the patients to dose level $d^* = \eta^{-1}(\Delta, \vartheta)$ for the new drug and the remaining patients to the active control. On the other hand, if $L < d^* \leqslant x^* < R$, the line $\{\gamma\tilde{c} : \gamma > 0\}$ does not intersect the set $\mathcal{C} \cup (-\mathcal{C})$ at the boundary of the Elfving set $\mathcal{R}$, and the situation is more complicated. A typical situation in this case is shown in Fig. 1(a) for the vector $\tilde{c}_1$. The locally AC-optimal design allocates $\rho\tilde{\omega}_1 100\%$ and $\rho\tilde{\omega}_2 100\%$ of patients to dose levels $x^*$ and $R$ of the new drug, respectively, where $\rho = \sqrt{\delta}\sigma_1(\sqrt{\delta}\sigma_1 + \sigma_2)^{-1}$, and allocates the remaining patients to the active control; here

$$\tilde{\omega}_1 = \frac{v(R, \theta_1)R(R - d^*)(\vartheta_2 + x^*)^2}{v(R, \theta_1)R(R - d^*)(\vartheta_2 + x^*)^2 + v(x^*, \theta_1)x^*(x^* - d^*)(\vartheta_2 + R)^2}, \tag{22}$$

$\tilde{\omega}_2 = 1 - \tilde{\omega}_1$, $\delta = \tilde{c}^{\mathrm{T}} M_1^{-1}(\tilde{\xi}^*, \theta_1)\tilde{c}$ and $d^* = \eta^{-1}(\Delta, \theta_1)$.

As a further example, consider the Michaelis–Menten model for the probability of a negative binomially distributed response. We have $s_1 = 2$, $s_2 = 1$, $\pi(d, \theta_1) = \vartheta_1 d(\vartheta_2 + d)^{-1}$, $\tilde{c} = \partial\eta(d^*, \theta_1)/\partial\theta_1 = r_1(\vartheta_1 d^*)^{-1}\{-(\vartheta_2 + d^*)\vartheta_1^{-1}, 1\}^{\mathrm{T}}$ and $v(d, \theta_1) = (\{r_1(d + \vartheta_2)^3\}/[d^2\vartheta_1^2\{d(1 - \vartheta_1) + \vartheta_2\}])^{1/2}$. A corresponding Elfving set is depicted in Fig. 1(b) for $\vartheta_1 = 1$, $\vartheta_2 = 0{\cdot}5$ and $\mathcal{D} = [0, 10]$; the locally AC-optimal design is always supported at three points. A straightforward calculation shows that the locally AC-optimal design allocates $\rho\tilde{\omega}_1 100\%$ and $\rho\tilde{\omega}_2 100\%$ of the patients to dose levels $L$ and $R$ of the new drug, respectively, where $\rho = [\delta\theta_2^2 - \{(1 - \theta_2)\delta\theta_2^2 r_2\}^{1/2}]\{\delta\theta_2^2 - (1 - \theta_2)r_2\}^{-1}$, and allocates the remaining patients to the active control; in this case

$$\tilde{\omega}_1 = \frac{v(R, \theta_1)R(R - d^*)(\vartheta_2 + L)^2}{v(R, \theta_1)R(R - d^*)(\vartheta_2 + L)^2 + v(L, \theta_1)L(d^* - L)(\vartheta_2 + R)^2}, \tag{23}$$

$\tilde{\omega}_2 = 1 - \tilde{\omega}_1$, $\delta = \tilde{c}^{\mathrm{T}} M_1^{-1}(\tilde{\xi}^*, \theta_1)\tilde{c}$ and $d^* = \eta^{-1}(\Delta, \theta_1)$.

Next, consider the Michaelis–Menten model for binomially distributed responses. In this case we have $s_1 = 2$, $s_2 = 1$, $\pi(d, \theta_1) = \vartheta_1 d(\vartheta_2 + d)^{-1}$, $\tilde{c} = \partial\pi(d^*, \theta_1)/\partial\theta_1$ and $v(d, \theta_1) = |d + \vartheta_2|[d\vartheta_1\{d(1 - \vartheta_1) + \vartheta_2\}]^{-1/2}$. The corresponding Elfving set is depicted in Fig. 2(a) for $\vartheta_1 = 1$, $\vartheta_2 = 0{\cdot}1$ and $\mathcal{D} = [0{\cdot}02, 2]$; we have to distinguish three different situations. Observe that the line $\{\gamma\tilde{c} : \gamma > 0\}$ intersects the boundary of the Elfving set $\mathcal{R}$ at some point $\mathcal{C} \cup (-\mathcal{C})$ if and only if $L \leqslant x_1^* \leqslant d^* \leqslant x_2^* \leqslant R$, where

$$x_1^* = L \vee \frac{\vartheta_2[1 - \{1 - \pi(R, \theta_1)\}^{1/2}]}{2\vartheta_1 - 1 + \{1 - \pi(R, \theta_1)\}^{1/2}}, \quad x_2^* = R \wedge \frac{\vartheta_2[1 + \{1 - \pi(R, \theta_1)\}^{1/2}]}{2\vartheta_1 - 1 - \{1 - \pi(R, \theta_1)\}^{1/2}}.$$

A typical situation is shown for the vector $\tilde{c}_1$ in Fig. 2(a). Consequently, the same arguments as in the previous examples show that in this case the locally AC-optimal design allocates $\rho 100\%$

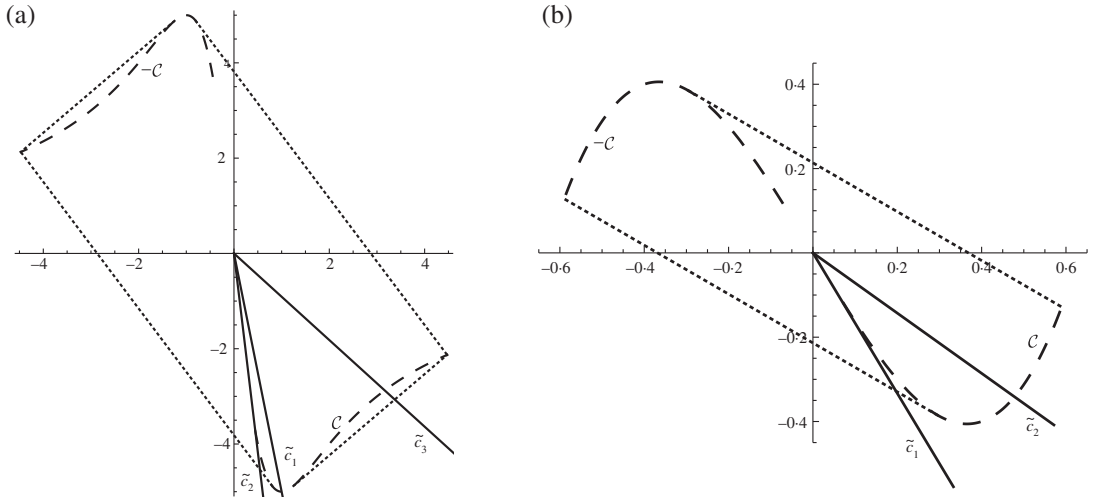


Fig. 2. The Elfving set (21) in model (1), where the expected response is given by the Michaelis–Menten model: (a) binomial distribution; (b) Poisson distribution.

of the patients to the dose level $d^*$ of the new drug, where $\rho = [\delta - \{\delta(1-\theta_2)\theta_2\}^{1/2}]\{\delta - (1-\theta_2)\theta_2\}^{-1}$ with $\delta = \tilde{c}^{\mathrm{T}} M_1^{-1}(\tilde{\xi}^*, \theta_1)\tilde{c}$ and $d^* = \eta^{-1}(\Delta, \theta_1)$, and allocates the remaining patients to the active control. If $L < d^* \leqslant x_1^*$, the locally AC-optimal design allocates $\rho\tilde{\omega}_{11}100\%$ and $\rho(1-\tilde{\omega}_{11})100\%$ of patients to dose levels $x_1^*$ and $R$ of the new drug, respectively, and allocates the remaining patients to the active control, where $\tilde{\omega}_{11}$ is of the form (22) with $x^* = x_1^*$. A typical situation is shown for the vector $\tilde{c}_2$ in Fig. 2(a). The case $L \leqslant x_2^* \leqslant d^* \leqslant R$ corresponds to the vector $\tilde{c}_3$. Here the locally AC-optimal design allocates $\rho\tilde{\omega}_{21}100\%$ and $\rho(1-\tilde{\omega}_{21})100\%$ of patients to dose levels $x_2^*$ and $R$ of the new drug and the remaining patients to the active control, where, with $L = x_2^*$, $\tilde{\omega}_{21}$ is of the form (23) and $d^* = \eta^{-1}(\Delta, \theta_1)$.

Finally, we consider the case of Poisson-distributed responses. We have $s_1 = 2$, $s_2 = 1$, $\lambda(d, \theta_1) = \vartheta_1 d(\vartheta_2 + d)^{-1}$ and $v(d, \theta_1) = \{(\vartheta_2 + d)/(\vartheta_1 d)\}^{1/2}$; by Theorem 5, we have to solve a $\tilde{c}$-optimal design problem with $\tilde{c} = \partial\lambda(d^*, \theta_1)/\partial\theta_1 = \{d^*(\vartheta_2 + d^*)^{-1}, -\vartheta_1 d^*(\vartheta_2 + d^*)^{-2}\}^{\mathrm{T}}$. It is easy to see that the line $\{\gamma\tilde{c} : \gamma > 0\}$ intersects the boundary of the Elfving set $\mathcal{R}$ at some point $\mathcal{C} \cup (-\mathcal{C})$ if and only if $L \leqslant x^* \leqslant d^* < R$, where $x^* = L \vee R\vartheta_2(3R + 4\vartheta_2)^{-1}$; see Fig. 2(b) for illustration of the case with $\vartheta_1 = 2{\cdot}5$, $\vartheta_2 = 1{\cdot}5$, $\mathcal{D} = [0{\cdot}02, 10]$ and the vector $\tilde{c}_2$. Consequently, the same arguments as in the previous examples show that in this case the locally AC-optimal design allocates $\rho 100\%$ of patients to dose level $d^*$ of the new drug, where $\rho = \sqrt{\delta}(\sqrt{\delta} + \sqrt{\theta_2})^{-1}$ with $\delta = d^*\vartheta_1(\vartheta_2 + d^*)^{-1}$ and $d^* = \lambda^{-1}(\Delta, \theta_1)$, and allocates the remaining patients to the active control. If $L < d^* \leqslant x^* < R$, the locally AC-optimal design allocates $\rho\tilde{\omega}_1 100\%$ and $\rho(1-\tilde{\omega}_1)100\%$ of patients to dose levels $x^*$ and $R$ of the new drug and allocates the remaining patients to the active control, where $\tilde{\omega}_1$ is of the form (22) with $\delta = \{\partial\eta(d^*, \theta_1)/\partial\theta_1\}^{\mathrm{T}} M_1^{-}(\tilde{\xi}^*, \theta_1)\{\partial\eta(d^*, \theta_1)/\partial\theta_1\}$. A typical situation is shown for the vector $\tilde{c}_1$ in Fig. 2(b).

### 4·3. *Locally AC-optimal designs in the $E_{\max}$ model*

Explicit expressions for the AC-optimal designs in the $E_{\max}$ model are complicated, so for brevity we conclude this paper by discussing AC-optimal designs for the two examples from § 1. All designs presented in this subsection were calculated numerically using particle swarm optimization (Clerc, 2006), and the optimality has been checked by applying Lemma 1.

Table 2. *AC-optimal designs in the two examples under different distributional assumptions, together with the efficiencies of the designs used in the study*

| | Gouty arthritis example (target dose $d^* = 100$ mg) | | | | Acute migraine example (target dose $d^* = 35{\cdot}6$ mg) | | | |
|---|---|---|---|---|---|---|---|---|
| Distribution | AC-optimal design | | | $\text{eff}_{\text{AC}}$ | AC-optimal design | | | $\text{eff}_{\text{AC}}$ |
| Normal | (101·06, 0) 49·99% | | ($C$, 1) 50·01% | 0·66 | (35·739, 0) 49·99% | | ($C$, 1) 50·01% | 0·48 |
| nB/Bin | (5·44, 0) 7·6% | (300, 0) 35·6% | ($C$, 1) 56·8% | 0·48 | (0, 0) 7·34% | (200, 0) 41·95% | ($C$, 1) 50·71% | 0·47 |

nB, negative binomial; Bin, binomial.

We begin with the gouty arthritis clinical trial, for which we use the same prior information as in Example 2. AC-optimal designs under the assumption of normal and negative binomial distributions are shown in Table 2. For instance, under the assumption of normally distributed endpoints, the AC-optimal design allocates almost half of the patients to the dose level 101·06 mg and the rest to the active control. In order to compare these results with the standard design introduced in Example 2, we also report values of the efficiency $\text{eff}_{\text{AC}}(\xi, \theta) = \psi(\xi^*_{\text{AC}}, \theta)/\psi(\xi, \theta)$, where $\psi(\xi, \theta)$ is defined in (19) and $\xi^*_{\text{AC}}$ is the locally AC-optimal design. For example, the efficiency of the standard design for estimating the target dose under the assumption of a normal or negative binomial distribution is 66% or 48%, respectively.

For the acute migraine clinical trial, we again use the prior information from Example 2. AC-optimal designs for normally and binomially distributed responses are summarized in Table 2. The efficiencies of the standard designs are 48% and 47% under the assumptions of normal and binomial distributions, respectively.

## Acknowledgement

This work was supported in part by the German Science Foundation and the U.S. National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Frank Bretz is also affiliated with the Shanghai University of Finance and Economics, People's Republic of China.

## Supplementary material

Supplementary material available at *Biometrika* online includes further examples and the proof of Theorem 5.

## Appendix

*Proof of Theorem* 1. If the matrix $K$ is of the form (12), we obtain from (7) and (12) that $K^{\mathrm{T}}M^{-}(\xi, \theta)K = [\{(1 - w_{k+1})^{-1}K_{11}^{\mathrm{T}}M_1^{-}(\tilde{\xi}, \theta_1)K_{11}, 0\}^{\mathrm{T}}, (0, w_{k+1}^{-1}K_{22}^{\mathrm{T}}I_2^{-}(\theta_2)K_{22})^{\mathrm{T}}]$, which gives, for the criterion (9), $\phi_p(\xi) = (t^{-1}\{(1 - w_{k+1})^p t_1\}\{\tilde{\phi}_p(\tilde{\xi})\}^p + w_{k+1}^p t^{-1}\text{tr}[\{K_{22}^{\mathrm{T}}I_2^{-}(\theta_2)K_{22}\}^{-p}])^{1/p}$ when $p \neq 0, -\infty$, where $t = t_1 + t_2$ and the function $\tilde{\phi}_p$ is as defined in (13). It is easy to see that the function $\phi_p$ is an increasing function of $\tilde{\phi}_p(\tilde{\xi})$. Consequently, the locally $\phi_p$-optimal design problem for the dose-response model with an active control can be solved by determining a design $\tilde{\xi}^*_p$ which maximizes the criterion (13) in a first step. If $\phi^* = \tilde{\phi}_p(\tilde{\xi}^*_p) = \max_{\tilde{\xi}} \tilde{\phi}_p(\tilde{\xi})$ denotes the optimal value for this criterion, it remains to maximize the function $\phi_p$ with respect to the weight $w_{k+1}$ assigned to the active control, which gives the expression (15) and proves the assertion for the case where $p \neq 0, -\infty$. The remaining cases of $p = 0$ and $p = -\infty$ are proved similarly. □

*Proof of Theorem 3.* The proof of (i) was given in § 3. For the remaining cases we restrict ourselves to the Poisson distribution, for which the Fisher information matrix in model (1) is $I_1(d,\theta_1) = d\{\vartheta_1(\vartheta_2+d)\}^{-1}(\{1,-\vartheta_1(\vartheta_2+d)^{-1}\}^{\mathrm{T}},[-\vartheta_1(\vartheta_2+d)^{-1},\vartheta_1^2\{(\vartheta_2+d)^2\}^{-1}]^{\mathrm{T}})$. All other cases are treated similarly. By Corollary 1, the D-optimal design can be obtained from the D-optimal design $\tilde{\xi}^*$ in a regression model with Fisher information matrix $I_1(d,\theta_1)$. If $M_1(\tilde{\xi},\theta_1) = \int_{\mathcal{D}} I_1(d,\theta_1)\,\mathrm{d}\tilde{\xi}(d)$ denotes an information matrix of a design $\tilde{\xi}$ in this model, then $\tilde{\xi}^*$ is D-optimal if and only if the inequality $\mathrm{tr}\{I_1(d,\theta_1)M_1^{-1}(\tilde{\xi}^*,\theta_1)\} \leqslant 2$ holds for all $d \in \mathcal{D}$; see Lemma 1. Moreover, there must be equality at the support points of the design $\tilde{\xi}^*$. This inequality is equivalent to an inequality of the form $P_3(d) \leqslant 0$ where $P_3$ is a polynomial of degree 3 with $P(0) < 0$. A straightforward argument now shows that $\tilde{\xi}^*$ has exactly two support points $d_1^* > 0$ and $d_2^* = R$. Consequently, the D-optimal design $\tilde{\xi}_1^*$ for the regression model with information matrix $I_1(d,\theta_1)$ has equal masses at the points $d_1^*$ and $R$, where $d_1^*$ maximizes the function $f(d) = R(R-d)^2 d\{4(R+\vartheta_2)^3(\vartheta_2+d)^3\}^{-1}$ in the interval $[L, R]$, that is, $d_1^* = L \vee \vartheta_2 R(3\vartheta_2+2R)^{-1}$. The assertion now follows from an application of Corollary 1, upon observing that $t_1 = 2$ and $t_2 = 1$ in the case under consideration. □

## REFERENCES


BORNKAMP, B., BRETZ, F., DETTE, H. & PINHEIRO, J. (2011). Response-adaptive dose-finding under model uncertainty. *Ann. Appl. Statist.* **5**, 1611–31.

BRETZ, F., HSU, J. C., PINHEIRO, J. C. & LIU, Y. (2008). Dose finding – a challenge in statistics. *Biomet. J.* **50**, 480–504.

BROWN, L. D. (1986). *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory*. Hayward, California: Institute of Mathematical Statistics.

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

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

CLERC, M. (2006). *Particle Swarm Optimization*. London: Iste Publishing Company.

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

DETTE, H., KISS, C. & WONG, W. K. (2010). A web-based tool for finding optimal designs for the Michaelis–Menten model and an overview. *Statist. Biopharm. Res.* **2**, 383–93.

DETTE, H., KISS, C., BENDA, N. & BRETZ, F. (2014). Optimal designs for dose finding studies with an active control. *J. R. Statist. Soc.* B **76**, 265–95.

DRAGALIN, V., HSUAN, F. & PADMANABHAN, S. K. (2007). Adaptive designs for dose-finding studies based on sigmoid $E_{\max}$ model. *J. Biopharm. Statist.* **17**, 1051–70.

ELFVING, G. (1952). Optimal allocation in linear regression theory. *Ann. Statist.* **23**, 255–62.

HELMS, H.-J., BENDA, N. & FRIEDE, T. (2015). Point and interval estimators of the target dose in clinical dose-finding studies with active control. *J. Biopharm. Statist.* **25**, 939–57.

KIEFER, J. (1974). General equivalence theory for optimum designs (approximate theory). *Ann. Statist.* **2**, 849–79.

MILLER, F., GUILBAUD, O. & DETTE, H. (2007). Optimal designs for estimating the interesting part of a dose-effect curve. *J. Biopharm. Statist.* **17**, 1097–115.

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

TING, N. (2006). *Dose Finding in Drug Development*. New York: Springer.


[*Received October* 2014. *Revised June* 2015]