Optimal designs for dose-response models with restricted design spaces

Stefanie Biedermann
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany
email: stefanie.biedermann@ruhr-uni-bochum.de

Holger Dette
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany
email: holger.dette@ruhr-uni-bochum.de

Wei Zhu
Department of Applied Mathematics and Statistics
State University of New York
Stony Brook, NY 11794-3600, USA
email: zhu@ams.sunysb.edu

July 7, 2004

Abstract

In dose response studies, the dose range is often restricted due to concerns over drug toxicity and/or efficacy. We present restricted and unrestricted interval locally optimal designs with respect to a very general class of optimality criteria for estimating the underlying dose response curve. The underlying curve belongs to a diversified set of link functions suitable for the dose response studies and having a common canonical form. These include the fundamental binary response models – the logit and the probit as well as the skewed versions of these models. The results are illustrated through the re-design of a dose ranging trial conducted at the Merck Research Laboratories (Zeng and Zhu, 1997). This work is a generalization of the results of Dai and Zhu (2002) in terms of the design interval, the underlying dose response curve and the optimality criterion.

AMS Classification: 62K05, 62J12
Keywords and Phrases: Binary response model; Dose ranging; Dose response; Link function; General Equivalence Theorem; Locally compound optimal design; Information function; Matrix mean.

1 Introduction

The motivation for preparing this article arose from a commonly observed design problem in dose-response experiments. Dose-response experiment is routinely conducted in preclinical and Phase I and II clinical trials to study the relationship between the dose level of a drug and the probability of a response, be it "cured" or "poisoned". For decades, statisticians have been searching and advocating the adoption of optimal designs in clinical trials. However, computer algorithms for generating theoretical optimal designs usually assume we have an unrestricted dosage range; see for example, Chaloner and Larntz (1989), Zhu and Wong (1998, 2000, 2001). The issue of restricted dose-range is especially pertinent to studies done with human subjects. In clinical trials it was often noted (Gart et al., 1986) that "for many agents, the response rate at the high dose in the optimal design may exceed that found at the maximum tolerated dose". We recently encountered such an example. Prior to a dose-ranging trial on a new rheumatoid arthritis drug at the Merck Research Laboratories, a pilot study was done where 120 patients were equally randomized into a placebo (dose 0) and a high dose (dose 50) group for a 6-week trial. The response rates were 35% at the placebo and 65% at the high dose, and furthermore the dose-response curve was logit with the response rate at dose level x being $\pi(x) = 1/[1 + \exp(-\beta(x - \alpha))]$ (Zeng and Zhu, 1997). Based on these estimates, the locally optimal design for estimating the shape (slope) of the underlying dose-response curve (logit) would allocate half of the subjects to dose -71 and the other half to dose 121. It is impossible to implement such a design because the higher dose (121) exceeds the safety limit, and the lower dose (-71) has less drug content than the placebo. It is also hard to justify the role of this optimal design as a 'gold standard' in gauging other competing designs.

Traditionally, most optimal designs for dose response studies are derived assuming that the dose interval is as large as necessary, that is, $(-\infty, \infty)$. Let ED(100π) denote the dose level with response probability π, the unrestricted slope optimal design for the logit model is equally supported at ED8.3 and ED91.7, symmetrical around ED50 which is simply the location parameter α. From the afore mentioned Merck pilot study, we have the estimates of ED35 $= 0$ and ED65 $= 50$, and therefore it is not surprising to see that the slope optimal design is equally supported at a negative dose of ED8.3 = -71 and a high toxic dose of ED91.7 = 121.

Little work has been done on the construction of optimal designs for dose response studies with a restricted dose interval. Extensive literature search yielded two related papers one by Mats, Rosenberger and Flournoy (1998) where they derived the locally c- and D-optimal design for estimating the maximum tolerated dose in a Phase I clinical trial on a restricted design space and one by Haines, Perevozskaya and Rosenberger (2003) where they extend the latter approach to Bayesian c- and D-optimal designs. For the Merck dose-ranging trial, the goal was indeed to estimate the shape of the logit curve (β) as precisely as possible subject to the constraint that the median effective dose (α), which is often regarded as the key index of a dose-response study, will be estimated with a certain precision. Such an optimal design is called a constrained optimal design (Lee, 1987). In other words, we are seeking a design that would minimize $\text{Var}(\widehat{\beta})$ subject to the constraint that $\text{Var}(\widehat{a}) \leq c$, where $\widehat{\alpha}$ and $\widehat{\beta}$ are the maximum likelihood estimators of the corresponding model parameters and c is a user defined constant. The constrained optimal designs are often hard to derive and little progress was made until Cook and Wong (1994) showed that there is a 1-1 correspondence between the constrained and the compound optimal designs. The compound optimal design would minimize a convex combination of the individual design criteria (in our example the two variances) and is in general easier to solve for than its constrained counterpart. The implication is that we can now construct the entire class of compound optimal designs first and then search among which for the desired constrained optimal design using such straightforward tool as the efficiency plot proposed in Cook and Wong (1994).

For the Merck dose ranging trial, the locally compound optimal designs with unrestricted design interval carry the same undesirable feature as the slope optimal design. Therefore in this paper, we will focus on the derivation of restricted interval locally compound optimal designs for estimating the location and slope parameters of a binary response model, where the probability of a response at dose level x is given in the canonical form $\pi(x) = H(\beta(x-\alpha))$ with a distribution function $H : \mathbb{R} \to [0,1]$; see Wu (1988) and Mats, Rosenberger and Flournoy (1998) among many others. It includes for example two most common dose response models — the logit model with $H(z) = 1/[1+\exp(-z)]$ and the probit model with $H(z) = \int_{-\infty}^{z} \varphi(u)\,du$, where $\varphi(u)$ is the unit normal density. Previously Dai and Zhu (2002) had derived the unrestricted locally compound optimal designs for these two models. Each unrestricted compound optimal design was found to be equally supported at two points symmetrical to the location parameter α.

We put this design problem into a much more general framework. In section 2 of this article, we generalize the result from Dai and Zhu (2002) to any information function in the sense of Pukelsheim (1993) used as the criterion under consideration. We also extend these findings to a broader class of possible link functions, which are commonly applied in practice. Section 3 contains the main results of this paper and deals with the problem of restricted design spaces, which includes restrictions with respect to one as well as the two boundaries of the design interval. Again, we derive locally optimal designs with respect to arbitrary optimality criteria from the family of information functions. We also utilize some results of Pukelsheim and Torsney (1991) to derive formulas for the weights of the optimal designs with respect to a broad class of criteria, the so-called matrix means, among which we find the most commonly applied optimality criteria such as the D-, A-, E- and the T-criterion. To show the practical relevance of this approach, we apply our results in Section 4 towards the re-designing of the dose ranging trial conducted at the Merck Research Laboratories (Zeng and Zhu, 1997). We determine the above-mentioned locally compound optimal designs with respect to different restrictions on the design interval as special cases of the results of sections 2-3. The proofs of our results, finally, are deferred to an appendix. In this article, we are taking the Frequentists' approach (Chernoff, 1953) and thus our designs are termed "locally optimal". In the ensuing sections, we will omit the word "locally" for simplicity. Furthermore, we will derive the designs analytically which would yield the universally optimal designs (optimal among designs of any support points).

2 Optimal designs on an unrestricted design space

In a dose-response experiment, suppose we have n subjects and a proportion of ω_i subjects are allocated to dose $x_i,\ \ i = 1, 2, \ldots, k$. The corresponding design is usually denoted by

$$\xi = \left\{ \begin{array}{cccc} x_1 & x_2 & \cdots & x_k \\ \omega_1 & \omega_2 & \cdots & \omega_k \end{array} \right\}.$$

Let $\pi(x)$ represent the probability of success at a given dose level x, the Fisher information matrix under the model $\pi(x) = H(\beta(x-\alpha))$ is

$$M(\xi) = \sum_{i=1}^{k} n\omega_i h^2(\beta(x_i-\alpha)) \begin{pmatrix} \beta^2 & -\beta(x_i-\alpha) \\ -\beta(x_i-\alpha) & (x_i-\alpha)^2 \end{pmatrix}$$

where H is a given distribution function (assumed to be differentiable) and

$$h^2(z) = \frac{(H')^2}{H(1-H)}(z). \tag{1}$$

Note that the matrix $M(\xi)$ depends on the parameters in the model but this is not reflected in our notation for the sake of simplicity. For moderate sample sizes, the covariance of the maximum likelihood estimator for the parameter $K^T\theta = K^T(\alpha,\beta)^T$ is approximately proportional to the matrix $K^T M^-(\xi) K$. We therefore consider design criteria of the form $\Phi(C(\xi))$ with the information matrix $C(\xi)$ being given by

$$C(\xi) = (K^T M^-(\xi) K)^{-1}, \tag{2}$$

where $K \in \mathbb{R}^{2\times s}$ is a given matrix of rank $s \le 2$, M^- denotes a generalized inverse of the matrix M and it is assumed that the linear combinations $K^T\theta$ are estimable by the design, i.e. range$(K) \subset$ range$(M(\xi))$. In this article, we will concentrate on a special 2×2 matrix K defined by

$$K = \begin{pmatrix} \sqrt{\lambda} & 0 \\ 0 & \sqrt{1-\lambda} \end{pmatrix} \tag{3}$$

for some value $\lambda \in [0,1]$, reflecting different emphasis on the precision of the estimation of the respective parameters α and β. An optimal design maximizes an appropriate function of the information matrix $C(\xi)$, say Φ, which is called optimality criterion and there are numerous optimality criteria to compare competing designs; see, e.g., Silvey (1980). In the present article, unless not specified otherwise, the optimality criterion Φ is an information function in the sense of Pukelsheim (1993), which maps the non-negative definite 2×2 information matrices onto the non-negative real axis to make their different contents of information comparable. More precisely, an information function features the properties of being positively homogeneous, concave, non-negative, non-constant and upper semicontinuous. The most widely used criteria are the well known A-, D- and E-optimality criteria defined by

$$\Phi_{-1}(\xi) = \Big(\frac{1}{s}\mathrm{tr}\ C^{-1}\Big)^{-1}, \quad \Phi_0(\xi) = (\det C)^{1/s}, \quad \Phi_{-\infty}(\xi) = \lambda_{\min}(C).$$

A design ξ^* is called Φ-optimal for estimating the vector of weighted parameters $K^T\theta = (\sqrt{\lambda}\alpha, \sqrt{1-\lambda}\beta)$ if ξ^* maximizes the function $\Phi(C(\xi))$ for the matrix K defined in (3). If the parameter λ takes the value $1/2$ the weight matrix K is proportional to the identity matrix and we use the notation Φ-optimal design for simplicity. This notation is in line with the usual one since for $\lambda = 1/2$ we obtain the familiar Φ-optimal design problem for estimating the parameter vector $(\alpha,\beta)^T$ with respect to the Fisher information $M(\xi)$. The implementation of the information matrix $C(\xi)$ instead of $M(\xi)$ is therefore a generalization of the usual approach. We further note that the matrix $C(\xi)$ is also proportional to the Fisher information matrix for the parameter $K^T(\tau_1,\tau_2)^T$ in the linear regression model

$$y = \phi_1(z)\,\tau_1 + \phi_2(z)\,\tau_2 + \eta, \tag{4}$$

where $z = \beta(x-\alpha)$, $\phi_1(z) = -\beta h(z)$, $\phi_2(z) = z/\beta \cdot h(z)$, τ_1 and τ_2 are model parameters and η is a normal distributed error with mean 0 and variance σ^2. Thus, the Φ-optimal design problem for estimating the weighted parameter vector $(\sqrt{\lambda}\ \alpha, \sqrt{1-\lambda}\ \beta)$ in the binary response model coincides with a Φ-optimal design problem for the linear model (4).
To fix ideas, we will give an example, which is of particular significance due to a further important property. Let Φ be the average-variance or A-criterion, i.e. $\Phi(C) = (\frac{1}{2}\mathrm{tr}\ C^{-1})^{-1}$, then it is easy to see that a design ξ^* maximizing $\Phi(C(\xi))$ minimizes the weighted average

$$\lambda \mathrm{Var}(\widehat{\alpha}) + (1-\lambda)\mathrm{Var}(\widehat{\beta}) \tag{5}$$

of the variances of the maximum likelihood estimators $\widehat{\alpha}$, $\widehat{\beta}$ for the corresponding model parameters. Thus, the design ξ^* maximizing $\Phi(C(\xi))$ is a compound optimal design with respect to the two objectives of precise estimation of both model parameters. From the preceding discussion, we obtain that ξ^* is at the same time a constrained optimal design in the sense that it minimizes $\mathrm{Var}(\widehat{\beta})$ subject to the constraint that $\mathrm{Var}(\widehat{\alpha}) \leq c$, where c is some user-selected constant.
Since the variances of $\widehat{\alpha}$ and $\widehat{\beta}$ can be of very different scale the impact of one of these estimators on the optimality criterion might outbalance the influence of the second estimator so that the resulting optimal design is in fact almost the optimal design with respect to the estimation of the more influential parameter. Several authors (e.g. Dette (1997)) therefore recommend the use of standardized optimality criteria. For a criterion of the form (5) the corresponding standardized version is given by

$$\tilde{\lambda}\,\frac{\mathrm{Var}(\widehat{\alpha})}{v_\alpha(\xi_\alpha^*)} + (1-\tilde{\lambda})\,\frac{\mathrm{Var}(\widehat{\beta}\,)}{v_\beta(\xi_\beta^*)} \tag{6}$$

where $v_\alpha(\xi_\alpha^*)$ denotes the asymptotic variance $(1,0)M^-(\xi_\alpha^*)(1,0)^T$ of $\widehat{\alpha}$ evaluated at the optimal design ξ_α^* for the estimation of α, $v_\beta(\xi_\beta^*)$ denotes the analogous expression with respect to β, and $\tilde{\lambda} \in [0,1]$ is a preliminary weight chosen by the experimenter. Obviously, minimizing (6) with respect to the design ξ is equivalent to minimizing (5) where a particular choice of the weight λ in (5) is used, that is

$$\lambda = \frac{\frac{\tilde{\lambda}}{v_\alpha(\xi_\alpha^*)}}{\frac{\tilde{\lambda}}{v_\alpha(\xi_\alpha^*)} + \frac{1-\tilde{\lambda}}{v_\beta(\xi_\beta^*)}} = \frac{\tilde{\lambda} v_\beta(\xi_\beta^*)}{\tilde{\lambda} v_\beta(\xi_\beta^*) + (1-\tilde{\lambda})v_\alpha(\xi_\alpha^*)}, \quad 1-\lambda = \frac{(1-\tilde{\lambda})v_\alpha(\xi_\alpha^*)}{\tilde{\lambda} v_\beta(\xi_\beta^*) + (1-\tilde{\lambda})v_\alpha(\xi_\alpha^*)}. \tag{7}$$

For optimality criteria with respect to an information matrix from the class defined by (2) and (3), a standardized version can also be obtained by using this choice (7) of the weight λ.
The optimal designs ξ_α^* and ξ_β^* are already established in design literature for many different choices of link functions and can for example be found in Ford, Torsney and Wu (1991) so that the values of $v_\alpha(\xi_\alpha^*)$ and $v_\beta(\xi_\beta^*)$ can easily be calculated by standard software such as Mathematica or Matlab. It turns out that $v_\alpha(\xi_\alpha^*) = c_h/\beta^2$ and $v_\beta(\xi_\beta^*) = d_h\,\beta^2$ where the constants c_h and d_h depend on the function h and thus on the underlying link function H. For either very large or very small values of β we can therefore expect a large difference in scale of the variances of $\widehat{\alpha}$ and $\widehat{\beta}$ so that the use of the standardized criteria, i.e. the implementation of a weight λ of the form (7), is strongly recommended.
For a general criterion, we can utilize the well-known general equivalence theorem for information functions Φ to find conditions, which characterize Φ-optimal designs. For this purpose, define

$$\phi(z) = (\phi_1(z), \phi_2(z))^T = (-\beta\, h(z), z/\,\beta \cdot h(z)\,)^T$$

as the vector of regression functions in the linear model (4) where $z \in \mathcal{Z}$ and $\mathcal{Z} \subset \mathbb{R}$ is the appropriately chosen design interval. We also need the definition of the polar function of an information function which in the present context is defined for a non-negative definite 2×2 matrix as

$$\Phi^\infty(D) \;=\; \inf_A \{\mathrm{tr}\ (AD)\}$$

where the infimum is taken over all 2×2 positive definite matrices; see Pukelsheim (1993). The proof of the following theorem, which does not necessarily require the special form of the matrix K in (3) can also be found in Pukelsheim (1993).

Theorem 1 *Let ξ^* denote a design with Fisher information matrix $M = M(\xi^*)$, for which $K^T\theta$ is estimable with information matrix $C = C(\xi^*) = (K^T M^-(\xi^*)\, K)^{-1}$. Then ξ^* is Φ-optimal for $K^T\theta$ if and only if there exists a non-negative definite matrix $D \in \mathbb{R}^{2\times 2}$ that solves the polarity equation*

$$\Phi(C)\Phi^\infty(D) = trace\, CD = 1 \tag{8}$$

and there exists a generalized inverse G of M such that the normality inequality

$$\phi^T(z) G K C D C K^T G^T \phi(z) \le 1 \tag{9}$$

is satisfied for all $z \in \mathcal{Z}$. Moreover, there is equality in (9) for any support point of the design ξ^*.

For the problem under consideration the above result simplifies substantially. The matrix K is given by (3) and thus non-singular if $\lambda \in (0, 1)$, which will be assumed throughout the following discussion. Therefore $K^T\theta$ is estimable by the design ξ^* if and only if the matrix $M(\xi^*)$ is non-singular. This corresponds to intuition, because we are interested in the estimation of both parameters of the binary response model. We therefore admit only designs with at least two different support points, thus dealing with only positive definite (invertible) matrices $M(\xi^*)$ and $C(\xi^*)$. By these considerations, the normality inequality (9) from Theorem 1 reduces to

$$\phi^T(z)K^{-1}DK^{-1}\phi(z) \leq 1 \quad \forall\, z \in \mathcal{Z}. \tag{10}$$

For a given information function Φ the solution D of the polar equation (8) is usually unique and can be determined explicitly; see Pukelsheim (1993). Moreover, the structure of the equivalence theorem is the same for all optimality criteria under consideration and this observation will be used for analyzing the properties of Φ-optimal designs for estimating the vector of weighted parameters $K^T\theta = (\sqrt{\lambda}\alpha, \sqrt{1-\lambda}\beta)$ in the binary response model. For this purpose, note that the matrix $K^{-1}DK^{-1}$ does not depend on z and therefore (10) yields the inequality

$$az^2 + bz + c \leq h^{-2}(z) \quad \forall\, z \in \mathcal{Z} \tag{11}$$

for some coefficients a, b, c depending on λ, β and the entries of the matrix $K^{-1}DK^{-1}$. Moreover, there is equality in (11) for any support point of a Φ-optimal design ξ^*. In other words, inequality (11) gives a necessary and sufficient condition for the Φ-optimality of the design ξ^* for estimating the vector of weighted parameters $K^T\theta = (\sqrt{\lambda}\alpha, \sqrt{1-\lambda}\beta)$.

In order to derive bounds on the number of support points of the optimal designs in the binary response model the following condition on the function $h(z)$ and thus the link function is assumed to be satisfied throughout this article:

Condition (I): Let $g(z) = h^{-2}(z)$. Suppose that the function $g(z)$ is twice differentiable on the entire real axis $\mathbb{R}$ and that the equation $g''(z) = c$ has at most two solutions for any real constant c.

We are now ready to state the main result of this section, i.e. the features of the Φ-optimal design ξ^* for estimating the vector of weighted parameters $K^T\theta = (\sqrt{\lambda}\alpha, \sqrt{1-\lambda}\beta)$ on the unrestricted design interval $\mathcal{Z} = \mathbb{R}$. The proof is given in the appendix.

Theorem 2 *Assume that condition (I) is satisfied for the binary response model under consideration.*

(a) Any Φ-optimal design for estimating the vector of weighted model parameters $K^T\theta = (\sqrt{\lambda}\alpha, \sqrt{1-\lambda}\beta)$ is supported at two points, which are uniquely determined.

(b) If, additionally, the function $h(z)$ is symmetric and the information function Φ defined on the set of non-negative definite 2×2 matrices satisfies

$$\Phi\left(\begin{pmatrix} a & b \\ b & c \end{pmatrix}\right) = \Phi\left(\begin{pmatrix} a & -b \\ -b & c \end{pmatrix}\right) \tag{12}$$

for all non-negative definite matrices

$$\begin{pmatrix} a & b \\ b & c \end{pmatrix} \in \mathbb{R}^{2\times 2},$$

then there exists a symmetric Φ-optimal design on two points for estimating the vector of weighted parameters $K^T\theta = (\sqrt{\lambda}\alpha, \sqrt{1-\lambda}\beta)$. If the information function Φ is even strictly concave, the Φ-optimal design is unique.

If the assumptions of part *(b)* in Theorem 2 are complied with, the maximization problem for the determination of the Φ-optimal designs contains only one variable, say z_0^*, and the solution can be easily obtained. The condition on the information function Φ is satisfied for most of the common optimality criteria such as the well-known matrix means or Φ_p-criteria,

$$\Phi_p(C) = (\frac{1}{2}\mathrm{tr}\, C^p)^{1/p} \text{ with } p \in (-\infty, 1], \quad \Phi_{-\infty}(\xi) = \lambda_{\min}(C) \text{ with } p = -\infty \tag{13}$$

defined as the p-mean of the eigenvalues of the positive definite matrix C; see Pukelsheim (1993). The class of Φ_p-criteria includes for example the well-known D-, A-, E- and T-criterion functions, which correspond to the values $p = 0,\ p = -1,\ p = -\infty$ and $p = 1$, respectively. The additional condition in *(b)* on the link function is met by the logit and the probit link. The uniqueness of the Φ-optimal design applies whenever the information function Φ is a strictly concave function on the set of non-negative definite matrices.

Condition (I) is satisfied by most of the commonly applied link functions, such as the logit and the probit link. A more detailed overview on the behavior of link functions with respect to condition (I) is displayed in Table 1. Note that condition (I) is not complied with by the double exponential and the double reciprocal link functions, since the function $g(z) = h^{-2}(z)$ is not differentiable in the origin $z = 0$. This coincides with the result of Ford, Torsney and Wu (1992), who showed that the D-optimal designs for these two links are supported on three points, whereas Φ- (and in particular D-) optimal designs for link functions satisfying condition (I) have exactly two support points as shown in Theorem 2.

Table 1: *Behavior of the function $h(z)$ for several common link functions $H(z)$ with respect to condition (I) ($s(z) = sign(z)$).*

link function	$H(z)$	$h^2(z)$	condition (I)
Double Exponential	$\frac{1+s(z)}{2} - \frac{s(z)}{2}e^{-\lvert z\rvert}$	$\frac{1}{2e^{\lvert z\rvert}-1}$	not met
Double Reciprocal	$\frac{1+s(z)}{2} - \frac{s(z)}{2}\left(\frac{1}{1+\lvert z\rvert}\right)$	$\frac{1}{(1+\lvert z\rvert)^2(2\lvert z\rvert+1)}$	not met
Complementary Log-Log	$1 - e^{-e^z}$	$\frac{e^{2z}}{-1+e^{e^z}}$	met
Logit	$\frac{1}{(1+e^{-z})}$	$\frac{e^z}{(1+e^z)^2}$	met
Probit	$\Phi(z)$	$\frac{\phi^2(z)}{\Phi(z)(1-\Phi(z))}$	met
Skewed Logit ($m > 0$)	$\frac{1}{(1+e^{-z})^m}$	$\frac{m^2}{(1+e^z)^2(-1+(1+e^{-z})^m)}$	met

The results of this paper apply to all four models satisfying Condition (I). The probit and logit models are the most fundamental models in dose response studies. The logit model closely resembles the probit model and both are symmetrical around the ED50. One can easily envisage situations where the researcher would not want to impose the symmetry feature of the logit/probit link functions on their data (Stukel, 1988). The skewed logit model, also called power logit model, was first proposed by Prentice in 1976. It generalizes the logit model by adding an additional skew parameter in the form of the power of the logit function and has found applications in the biomedical field as well as other scientific research areas (e.g. Gaudard et al., 1993; Nagler, 1994; Hedayat et al., 1997; Wang and Hung,1997; Leuraud and Benichou 2001). The complementary log-log model is another asymmetrical extension of the logit/probit model especially pertinent to the toxicity studies (Kuk, 2004) and design in the time domain (Throne et al., 1995).

3 Optimal designs on restricted design spaces

To guarantee a certain level of drug efficacy due to the increasing ethics concerns, one must impose a lower bound on the design interval. By the same coin, to avoid a severe side-effect or drug toxicity, one would have to impose an upper bound on the design interval. We will first present the left-restricted Φ-optimal designs for estimating the vector of weighted parameters $K^T\theta = (\sqrt{\lambda}\alpha, \sqrt{1-\lambda}\beta)$ in the binary response model with the normalized design interval in terms of $z = \beta(x-\alpha)$ being $[A, \infty)$, where $0 > A > z_L^*$, and z_L^* denotes the lower design point of the Φ-optimal design with the same λ on the unrestricted design interval $\mathcal{Z} = \mathbb{R}$. The derivation of optimal designs in the right-restricted case where the design interval is $(-\infty, B]$ with $B < z_U^*$, z_U^* denoting the upper design point, would follow the same lines and is thus omitted (see also the discussion in Lemma 2). The results on one-side restricted Φ-optimal designs will lead us to the more realistic scenario where both sides of the design interval are bounded as is true with the Merck dose-ranging trial.

3.1 One-side restricted intervals

We show that the left-restricted Φ-optimal design on $[A, \infty)$ is a two-point design, and furthermore the lower design support point is always the left boundary point of the design interval. Let $g(z) = h^{-2}(z)$ as defined in (1), the following results hold as long as two conditions are satisfied: Condition (I) which was already defined in the last section as well as

Condition (II): $z \cdot h(z) \to 0$ as $z \to \pm\infty$.

It is easy to show that all link functions from Table 1 satisfy condition (II). Again, we can use inequality (11) from the equivalence theorem to prove the following result on the number of support points of a Φ-optimal design ξ_A^* for estimating the vector of weighted parameters $K^T\theta = (\sqrt{\lambda}\alpha, \sqrt{1-\lambda}\beta)$ on the design space $[A, \infty)$. Throughout this section, we assume that condition (I) is satisfied and denote by z_L^* and z_U^* the lower and upper support point of the unrestricted Φ-optimal design for estimating the vector of weighted parameters $K^T\theta = (\sqrt{\lambda}\alpha, \sqrt{1-\lambda}\beta)$. Note that in the case $z_L^* \geq A$ the Φ-optimal design on the unrestricted design space is obviously also Φ-optimal on the restricted design space $[A, \infty)$. The following two results consider the remaining case $z_L^* < A$.

Lemma 1 *Let Φ be an arbitrary information function and assume that condition (I) is fulfilled. If $z_L^* < A$ for a given boundary value A, then a Φ-optimal design ξ^* for estimating the vector of weighted parameters $K^T\theta = (\sqrt{\lambda}\alpha, \sqrt{1-\lambda}\beta)$ on the interval $[A, \infty)$ is supported on exactly two different points.*

The following theorem shows that in the case $z_L^* < A$ the point A is always a support point of the Φ-optimal design for estimating the vector of weighted parameters $K^T\theta = (\sqrt{\lambda}\alpha, \sqrt{1-\lambda}\beta)$.

Theorem 3 *For a given $\lambda \in (0,1)$, if the smaller support point of the Φ-optimal design for estimating the vector $K^T\theta = (\sqrt{\lambda}\alpha, \sqrt{1-\lambda}\beta)$ on $\mathcal{Z} = \mathbb{R}$ is not included in the interval $[A, \infty)$, assume that condition (II) is satisfied. A Φ-optimal design for estimating the vector of weighted parameters $K^T\theta = (\sqrt{\lambda}\alpha, \sqrt{1-\lambda}\beta)$ on the interval $[A, \infty)$ has two unique support points, one of which is the boundary point A.*

Theorem 3 shows that a Φ-optimal design for estimating the vector of weighted parameters $K^T\theta = (\sqrt{\lambda}\alpha, \sqrt{1-\lambda}\beta)$ on the design space $[A, \infty)$ is always of the form

$$\xi_A^* = \left\{ \begin{array}{cc} A & z \\ \omega_1 & 1-\omega_1 \end{array} \right\},$$

where z denotes the upper support point of the design and ω_1 is the weight corresponding to the point A. A formula for the weight ω_1 with respect to a broad class of design criteria Φ is derived in paragraph 3.3, leaving only a univariate optimization problem in the variable z, which can be solved by standard numerical methods. The proofs for Lemma 1 and Theorem 3 are given in the Appendix.

3.2 Two-side restricted intervals

The more realistic situation in dose response experiments is when there exist restrictions on both the upper and the lower bound of the design interval. Furthermore, the restricted interval is not necessarily symmetrical around the location parameter α, or 0 in terms of the normalized dose level $z = \beta(x - \alpha)$. In the following, we assume the normalized design interval to be $[A, B]$. For a given $\lambda \in (0,1)$, we assume that for the upper support point for the corresponding left-restricted Φ-optimal design on the design space $[A, \infty)$, say $z_{U,A}^*$, and the smaller support point for the right-restricted Φ-optimal design on the design $(-\infty, B]$, say $z_{L,B}^*$, are not contained in the design interval $[A, B]$. If one of these points is in the interval $[A, B]$ the two-side restricted Φ-optimal design for estimating the vector of weighted parameters $K^T\theta = (\sqrt{\lambda}\alpha, \sqrt{1-\lambda}\beta)$ coincides with the corresponding Φ-optimal design for the one-side restricted design space. By the equivalence theorem we have the following result. The proof is given in the Appendix.

Theorem 4 *Assume that conditions (I) and (II) are satisfied and $z_{L,B}^* < A < B < z_{U,A}^*$. The two-side restricted Φ-optimal design $\xi_{A,B}^*$ for estimating the vector of weighted parameters*

$K^T\theta = (\sqrt{\lambda}\alpha, \sqrt{1-\lambda}\beta)$ on the design interval $[A, B]$ has two support points, which are given by the boundary points A and B. In particular, when the design interval is symmetrical, i.e. $A = -B$, and the assumptions in part (b) of Theorem 2 are met, the design which allocates equal weight to both boundary points is Φ-optimal for estimating the vector of weighted parameters $K^T\theta = (\sqrt{\lambda}\alpha, \sqrt{1-\lambda}\beta)$.

3.3 A formula for the weights of optimal designs

In the cases considered in the two previous paragraphs, the optimal design problem reduces to the determination of one support point with corresponding weight in the case of a one-side restricted design space and to the determination of the weight at one point in the case of a two-side restricted design space. In this section, we will consider the problem of determining the optimal weight. For simplicity, we restrict ourselves to the important family of the Φ_p-criteria defined by (13) with $p \in [-\infty, 1]$. As we have pointed out previously, this class includes the well-known D-, A-, E- and T-criterion corresponding to the values $p = 0,\ p = -1,\ p = -\infty$ and $p = 1$, respectively, which are most important from a practical point of view. In the preceding discussion, we have provided some tools for the determination of the support points of a Φ-optimal design for estimating the vector of weighted parameters $K^T\theta = (\sqrt{\lambda}\alpha, \sqrt{1-\lambda}\beta)$ on restricted and unrestricted design spaces. We will now derive formulas for the weights of the optimal designs using the results of Pukelsheim and Torsney (1991). These authors gave an explicit method for constructing the weights maximizing $\Phi(C(\xi))$ if the support points of the design ξ are given. For simplicity, we restrict ourselves to the case of a two-side restricted interval $[A, B]$. The analogous results apply in the cases of unrestricted or one-side restricted design spaces if A and B are replaced by the Φ_p-optimal support points in the formulas below. In the case $p \in (-\infty, 1]$, it follows from the results of Pukelsheim and Torsney (1991) that the weight vector $\omega = (\omega_1, \omega_2)$ of a Φ_p-optimal design is given by

$$\omega_1 = \frac{\sqrt{L_{11}}}{\sum_{i=1}^2 \sqrt{L_{ii}}} \qquad \text{and} \qquad \omega_2 = 1 - \omega_1 \tag{14}$$

where L_{ii}, $i = 1, 2$ are the diagonal elements of the non-negative definite 2×2 matrix

$$L = VC^{p+1}V^T.$$

The matrix V is defined by $V = (XX^T)^{-1}XK$ with $X^T = (\phi(A), \phi(B)) \in \mathbb{R}^{2\times 2}$. For the given matrix K defined by (3) the matrix V for a two-side restricted Φ_p-optimal design on the design interval $[A, B]$ supported at A and B has the form

$$V = \begin{pmatrix} \frac{B\sqrt{\lambda}}{(B-A)\beta h(A)} & \frac{\beta\sqrt{1-\lambda}}{(B-A)h(A)} \\ \frac{A\sqrt{\lambda}}{(A-B)\beta h(B)} & \frac{\beta\sqrt{1-\lambda}}{(A-B)h(B)} \end{pmatrix},$$

which implies for the information matrix $C = C(\xi) = (K^T M^{-1}(\xi) K)^{-1}$ of the design ξ with masses ω_1 and $1 - \omega_1$ at the points A and B

$$C = \begin{pmatrix} \frac{B^2\lambda}{\omega_1(B-A)^2\beta^2h^2(A)} + \frac{A^2\lambda}{(1-\omega_1)(A-B)^2\beta^2h^2(B)} & \frac{B\sqrt{1-\lambda}\sqrt{\lambda}}{\omega_1(B-A)^2h^2(A)} + \frac{A\sqrt{1-\lambda}\sqrt{\lambda}}{(1-\omega_1)(A-B)^2h^2(B)} \\ \frac{B\sqrt{1-\lambda}\sqrt{\lambda}}{\omega_1(B-A)^2h^2(A)} + \frac{A\sqrt{1-\lambda}\sqrt{\lambda}}{(1-\omega_1)(A-B)^2h^2(B)} & \frac{\beta^2(1-\lambda)}{\omega_1(B-A)^2h^2(A)} + \frac{\beta^2(1-\lambda)}{(1-\omega_1)(A-B)^2h^2(B)} \end{pmatrix}^{-1}.$$

The matrix C^{p+1} for any criterion $p \in (-\infty, 1]$ results from the eigenvalues λ_i, $i = 1, 2$, and the eigenvectors x_i, $i = 1, 2$, with $\| x_i \| = 1$ of the matrix $V^T \Delta_\omega^{-1} V$ as follows:

$$C^{p+1} = \sum_{i=1}^{2} \lambda_i^{-p-1} x_i x_i^T.$$

The weight ω_1 can now be determined explicitly solving the nonlinear equation (14). The direct weight formula for arbitrary values of $p \in (-\infty, 1]$ that can be derived from equation (14) is somewhat lengthy and therefore not displayed in this article. However, it can be easily implemented in standard software such as Mathematica or Matlab. For $p = -\infty$, i.e. the E-criterion, the $\Phi_{-\infty}$-optimal weight ω_1 corresponding to the support point A can be determined by the formula

$$\omega_1 = \sqrt{J_{11}}, \tag{15}$$

where J_{11} denotes the first diagonal element of the matrix J. Similarly to the definition of the matrix L from the case $p \in (-\infty, 1]$, J is given by

$$J = VCDCV^T,$$

where D is a solution of the polarity equation (8). From Pukelsheim (1993), we obtain that for $p = -\infty$ the polar matrix D is given by

$$D = \begin{cases} \dfrac{xx^T}{\lambda_{min}(C)} & \lambda_{min}(C) \text{ is of multiplicity 1} \\[2ex] \dfrac{\gamma x_1 x_1^T + (1-\gamma) x_2 x_2^T}{\lambda_{min}(C)} & \lambda_{min}(C) \text{ is of multiplicity 2.} \end{cases}$$

The expressions x, x_1, x_2 denote the norm 1 eigenvectors of the information matrix C corresponding to its smallest eigenvalue $\lambda_{min}(C)$, and γ is a constant from the open unit interval $(0, 1)$. Due to the implicit case differentiation with respect to the polar matrix D, the implementation of formula (15) in standard software is somewhat more complex than for the case $p \in (-\infty, 1]$ but still feasible.
Moreover, the formulas above can be used to derive the following symmetry property of a design with Φ_p-optimal weights at two given points, say z_1 and z_2.

Lemma 2 *If the function h is symmetric and the design ξ has Φ_p-optimal weights $\omega_1, 1 - \omega_1$ at the points z_1, z_2, then the Φ_p-optimal weights of a design supported on $-z_2, -z_1$ are given by $1 - \omega_1$ and ω_1, respectively.*

This result reduces the design problem for the binary response model substantially because optimal designs on a given design space can be obtained from the optimal designs on the reflection of the design space at the origin. This observation is summarized in the following corollary corresponding to the situations considered in Theorems 3 and 4.

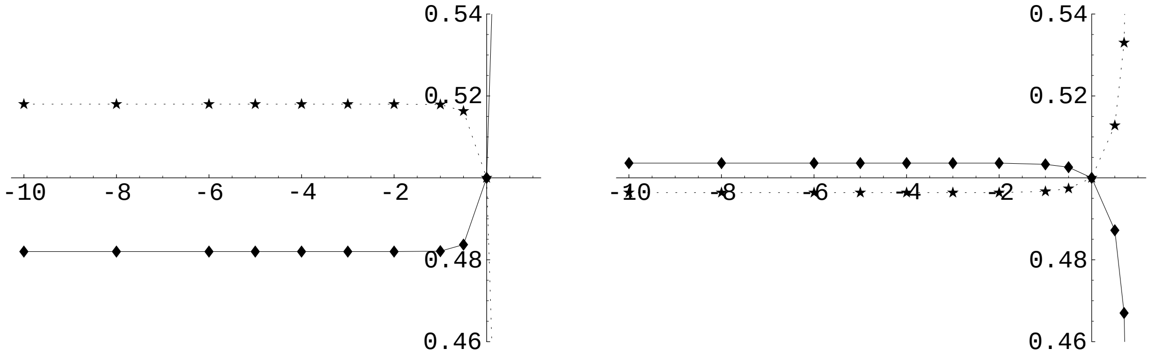
Corollary 1 *Let $p \in [-\infty, 1]$ and assume that the function h is symmetric.*

(a) If the assumptions of Theorem 3 are satisfied and let $\xi_A^* = \{A, z; \omega_1, 1-\omega_1\}$ denote a Φ_p-optimal design for estimating the vector of weighted parameters $K^T\theta = (\sqrt{\lambda}\alpha, \sqrt{1-\lambda}\beta)$ on the one-side restricted interval $[A, \infty)$, then the design $\xi_{-A}^* = \{-z, -A; 1-\omega_1, \omega_1\}$ is Φ_p-optimal for estimating the vector of weighted parameters $K^T\theta = (\sqrt{\lambda}\alpha, \sqrt{1-\lambda}\beta)$ on the one-side restricted interval $(-\infty, -A]$.

(b) If the assumptions of Theorem 4 are satisfied and let $\xi_{A,B}^* = \{A, B; \omega_1, 1-\omega_1\}$ denote a Φ_p-optimal design for estimating the vector of weighted parameters $K^T\theta = (\sqrt{\lambda}\alpha, \sqrt{1-\lambda}\beta)$ on the two-side restricted interval $[A, B]$, then the design $\xi_{-B,-A}^* = \{-B, -A; 1-\omega_1, \omega_1\}$ is Φ_p-optimal for estimating the vector of weighted parameters $K^T\theta = (\sqrt{\lambda}\alpha, \sqrt{1-\lambda}\beta)$ on the two-side restricted interval $[-B, -A]$.

Corollary 1 is illustrated in Figure 1 where we display the behavior of the weight ω_1 corresponding to the smaller support point A as a function of the value of $p \in (-\infty, 1]$ in the optimality criterion for the logit link function and various two-side restricted design spaces $[A, B]$ and parameter values β and λ. Due to the factorization of the weights for the determinant criterion, we always obtain the value $\omega_1 = 0.5$ for $p = 0$.

Figure 1: The Φ_p-optimal weight ω_1 corresponding to the smaller support point A as a function of the parameter p in the Φ_p-optimality criterion. Left panel: $A = -0.4$, $B = 0.9$, $\beta = 2$, $\lambda = 0.25$ (solid line); $A = -0.9$, $B = 0.4$, $\beta = 2$, $\lambda = 0.25$ (dotted line). Right panel: $A = -0.8$, $B = 0.5$, $\beta = 2$, $\lambda = 0.75$ (solid line); $A = -0.5$, $B = 0.8$, $\beta = 2$, $\lambda = 0.75$ (dotted line).



In the remaining part of this section, we discuss the important example of (weighted) A-optimality (5), which was considered by Dai and Zhu (2002) in the case of an unrestricted design space and corresponds to the particular choice $p = -1$ in the Φ_p-optimality criterion. For the Φ_{-1}-criterion, i.e. the average-variance or A-criterion, we obtain the following results. The one-side restricted Φ_{-1}-optimal design ξ_A^* on $[A, \infty)$ is given by

$$\xi_A^* = \left\{ \begin{array}{cc} A & z \\ \omega_1 & 1-\omega_1 \end{array} \right\}.$$

The A-optimal design problem is now to find the pair of normalized design support point z and the allocation proportion $\omega_1 \in [0,1]$ such that $\mathrm{tr}[C^{-1}(\xi)]$ is minimized. Solving equation (14) with respect to ω_1 yields for the weight corresponding to the support point A

$$\omega_1 = \left[1 + \frac{h(A)}{h(z)}\sqrt{\frac{(1-\lambda)\beta^2 + \lambda A^2/\beta^2}{(1-\lambda)\beta^2 + \lambda z^2/\beta^2}}\right]^{-1}. \tag{16}$$

Plugging (16) in the criterion function $\mathrm{tr}[C^{-1}(\xi)]$, this equation becomes univariate and standard numerical methods can be applied to solve for the support point z.
In the two-side restricted case, we obtain for ω_1 the corresponding expression with the unknown support point z replaced by the upper boundary B.

$$\omega_1 = \left[1 + \frac{h(A)}{h(B)}\sqrt{\frac{(1-\lambda)\beta^2 + \lambda A^2/\beta^2}{(1-\lambda)\beta^2 + \lambda B^2/\beta^2}}\right]^{-1}$$

In particular, when the design interval is symmetrical, that is $A = -B$, and the function $h(z)$ is also symmetrical, we have $\omega_1 = 0.5$, confirming the results with the symmetrically restricted optimal designs.
If the standardized A-criterion defined in (6) is used for the determination of optimal designs the formula for the weight ω_1 can be expressed independently of the value of β, i.e.

$$\omega_1 = \left[1 + \frac{h(A)}{h(B)}\sqrt{\frac{(1-\tilde{\lambda})/d_h + A^2\tilde{\lambda}/c_h}{(1-\tilde{\lambda})/d_h + B^2\tilde{\lambda}/c_h}}\right]^{-1}$$

where we used (7), the preliminary weights $\tilde{\lambda}$ and $1-\tilde{\lambda}$ as well as the formulas $v_\alpha(\xi^*_\alpha) = c_h/\beta^2$ and $v_\beta(\xi^*_\beta) = d_h\,\beta^2$.

4 Merck Dose Ranging Trial Revisited

In this section, we reanalyze a data example and demonstrate the practical relevance of the designs derived in this work. As a reasonable optimality criterion, we choose the average-variance or A-criterion, which belongs to the class of Φ_p-criteria with $p = -1$, i.e. we maximize the function $\Phi_{-1}(C) = (\frac{1}{2}\,\mathrm{tr}\,C^{-1})^{-1}$ over the positive definite information matrices $C = C(\xi) = (K^T M^{-1}(\xi)\,K)^{-1}$, where the matrix K is defined by (3). As pointed out previously, by this choice of optimality criterion, we obtain designs that would minimize the weighted average variances $\lambda\mathrm{Var}(\widehat{\alpha}) + (1-\lambda)\mathrm{Var}(\widehat{\beta})$ of the maximum likelihood estimators for the unknown model parameters α and β. For simplicity, the optimal designs with respect to this criterion will be denoted by compound optimal designs in the following. As described in the introduction, the compound optimal designs are at the same time constrained optimal with respect to some constraint on one of the particular variances above. (See Cook and Wong (1994) for a more detailed discussion on the equivalence of compound and constrained optimal designs).
From the 6-week pilot study done before the Merck dose-ranging trial, the response rate at the placebo was 35% and the response rate at the high dose (dosage 50) was 65%. The logit

model was found to be a suitable model and the maximum likelihood estimates of the model parameters are $\widehat{\alpha} \approx 25$ and $\widehat{\beta} \approx 0.025$. Since the value of $\widehat{\beta}$, which will be assumed the true value of β in the derivation of the optimal designs, is very small, i.e. β^2 differs from $1/\beta^2$ considerably, we felt the necessity to use the standardized criterion. Based on the above estimates, we found the locally slope optimal design to be a two point symmetrical design with the two dose levels being -71 and 121. We derived unrestricted compound optimal designs for estimating the two model parameters α and β under the logit model. For each choice of preliminary weight $\tilde{\lambda}$, the compound optimal design is equally supported at two dose levels symmetrical to α. Selected unrestricted compound optimal designs are presented in Table 2. Using the dose-percentile relationship $x = \alpha + \beta^{-1} \ln [\pi/(1-\pi)]$ and plugging in the above maximum likelihood estimates, we found that the same negative lower dose and large higher dose pattern persists in the compound optimal designs for the Merck dose-ranging (Table 3).
To avoid the negative dose levels, we restricted the design interval to $[0, \infty)$ in terms of the original dosages. This translates to a normalized dose range of $[-0.625, \infty)$. The smaller support points of the corresponding unrestricted compound optimal designs (Table 3) are not included in this interval. Selected left-restricted compound optimal designs, in terms of the normalized as well as the original design support points and the corresponding allocation proportions are given in Table 4.
As the right support points from the left-restricted compound optimal designs appear to be high or very high (Table 4), we felt it is necessary to restrict the design interval at both ends, and thus the interval $[0, 60]$ in the original dose scale. The ensuing designs are supported on the two ending points with the corresponding design allocation proportions shown in Table 5.
Merck started their dose ranging trial on that new rheumatoid arthritis drug using a safe but sub-optimal design scheme before we could finish our derivation of the restricted interval compound optimal designs. We were assured, however, of future considerations, at least as the 'gold standard', for the upcoming dose ranging trials. In addition to estimating the model parameters of the underlying dose response curve (Kalish, 1990; Meier et al., 1993; Zeng and Zhu, 1997; Zhu and Wong, 2000), there is also a great need to estimate other percentiles besides the median effective dose (α) (Flournoy, 1993; Rosenberger and Grill, 1997; Mats et al, 1997; Zhu and Wong, 2001). For example, the low percentiles are of interest in toxicity studies such as in virtually safe dose extrapolation problems, and the high percentiles are of interest in efficacy studies. We are currently deriving the locally compound optimal designs for estimating a set of percentiles in a dose response experiment for the same class of underlying dose response curves on restricted and unrestricted design intervals.

Acknowledgements: The support of the Deutsche Forschungsgemeinschaft (SFB 475, Komplexitätsreduktion in multivariaten Datenstrukturen, Teilprojekt A2) and the U.S. National Institute of Health (NIH 3 M01 RR010710-06S1 and NIH 2 P30 AG08051-11) are gratefully acknowledged. The authors are also grateful to Z. Dai, M. Marchlewski, S. Titoff and W.K. Wong for helpful computational assistance and helpful comments on an earlier version of this paper.

5 Appendix: Proofs

5.1 Proof of Theorem 2

The proof of Theorem 2 is divided into two lemmata, from which the assertion of the theorem becomes obvious. The first lemma shows that a Φ-optimal design ξ^* for estimating the vector of weighted parameters in a binary response model will always be a two point design. The second lemma deals with the symmetry of a Φ-optimal design with respect to the origin $z = 0$.

Lemma 3 *Assume condition (I) is satisfied. Then any Φ-optimal design for estimating the vector of weighted parameters $K^T\theta = (\sqrt{\lambda}\alpha, \sqrt{1-\lambda}\beta)$ is supported on exactly two points.*

Proof of Lemma 3: Suppose that the Φ-optimal design has at least three support points $z_1 < z_2 < z_3$, i.e. we have $f(z_i) = g(z_i)$, $i = 1, 2, 3$, where $f(z)$ denotes the left hand side of the inequality (11). By the mean value theorem we obtain that there exist z_1', z_3' such that $z_1 < z_1' < z_2 < z_3' < z_3$ and $f'(z_i') = g'(z_i')$, $i = 1, 3$. Since $f(z) \leq g(z)$ holds for all $z \in \mathbb{R}$, the points z_i, $i = 1, 2, 3$, are all tangent points, i.e. $f'(z_i) = g'(z_i)$. Applying the mean value theorem again to the functions $f'(z), g'(z)$, we receive points z_i'', $i = 1, \ldots, 4$ such that $z_1 < z_1'' < z_1' < z_2'' < z_2 < z_3'' < z_3' < z_4'' < z_3$ and $f''(z_i'') = g''(z_i'')$, $i = 1, \ldots, 4$. Since $f''(z) = 2a$ for all $z \in \mathbb{R}$, we have $g''(z_i'') = 2a$, $i = 1, \ldots, 4$, which contradicts with condition (I). $\Box$

Lemma 4 *Assume that the function h is symmetric and that Φ is an information function defined on the set of non-negative 2×2 matrices which satisfies (12). If there exists a Φ-optimal design for estimating the vector of weighted parameters $K^T\theta = (\sqrt{\lambda}\alpha, \sqrt{1-\lambda}\beta)$ on $\mathcal{Z} = \mathbb{R}$, then there also exists a symmetric Φ-optimal design for this problem.*

Proof of Lemma 4: Assume

$$\xi = \left\{ \begin{matrix} z_1 & z_2 & \cdots & z_k \\ \omega_1 & \omega_2 & \cdots & \omega_k \end{matrix} \right\}$$

is a Φ-optimal design for estimating the vector of weighted parameters $K^T\theta = (\sqrt{\lambda}\alpha, \sqrt{1-\lambda}\beta)$, define ξ^- as the reflection of ξ at the origin and consider

$$\frac{1}{2}\,\xi + \frac{1}{2}\,\xi^- = \left\{ \begin{matrix} -z_k & \cdots & -z_1 & z_1 & \cdots & z_k \\ \frac{\omega_k}{2} & \cdots & \frac{\omega_1}{2} & \frac{\omega_1}{2} & \cdots & \frac{\omega_k}{2} \end{matrix} \right\}.$$

From the assumptions of the lemma we can conclude that $\Phi(C(\xi)) = \Phi(C(\xi^-))$. It then follows from the concavity of the information function Φ that

$$\begin{aligned} \Phi(C(\xi)) &= \frac{1}{2}\,\Phi(C(\xi)) + \frac{1}{2}\,\Phi(C(\xi^-)) \\ &\leq \Phi(\frac{1}{2}\,C(\xi) + \frac{1}{2}\,C(\xi^-)) \;=\; \Phi(C(\frac{1}{2}\,\xi + \frac{1}{2}\,\xi^-)) \end{aligned}$$

Thus the design $\frac{1}{2}\,\xi + \frac{1}{2}\,\xi^-$ is at least as good as the original design ξ and is therefore a symmetric Φ-optimal design for estimating the vector of weighted parameters $K^T\theta = (\sqrt{\lambda}\alpha, \sqrt{1-\lambda}\beta)$. $\Box$

Proof of Theorem 2: By Lemma 3 and 4 it remains to prove the uniqueness of the support of the Φ-optimal design.

(a) Let ξ_1, ξ_2 be two Φ-optimal designs for estimating the vector of weighted parameters $K^T\theta = (\sqrt{\lambda}\alpha, \sqrt{1-\lambda}\beta)$. From Lemma 3, we conclude that they are both supported on two points. The concavity of the criterion function Φ implies that the design $\xi_3 = \frac{1}{2}\xi_1 + \frac{1}{2}\xi_2$ is also Φ-optimal for estimating the vector of weighted parameters $K^T\theta = (\sqrt{\lambda}\alpha, \sqrt{1-\lambda}\beta)$. If the support of the design ξ_1 does not coincide with the support of ξ_2, the design ξ_3 is supported on more than two points, which contradicts the assertion of Lemma 3.

(b) Let ξ be a Φ-optimal design for estimating the vector of weighted parameters $K^T\theta = (\sqrt{\lambda}\alpha, \sqrt{1-\lambda}\beta)$. The assumptions regarding the information function Φ imply that ξ^- is also Φ-optimal for estimating the vector of weighted parameters $K^T\theta = (\sqrt{\lambda}\alpha, \sqrt{1-\lambda}\beta)$. From the concavity of the criterion function we derive that $\xi_s = \frac{1}{2}\xi + \frac{1}{2}\xi^-$ is also Φ-optimal. If the support points of ξ are not equal to the support of ξ^-, i.e. the support points of ξ are not symmetric about the origin, this is a contradiction to the assertion of Lemma 3. Furthermore, if the criterion function Φ is strictly concave, it follows that $\xi_s = \xi$ and thus the uniqueness of the Φ-optimal design. $\square$

5.2 Proofs of Theorem 3 and Lemma 1

Proof of Lemma 1: Suppose that the Φ-optimal design ξ^* for estimating the vector of weighted parameters $K^T\theta = (\sqrt{\lambda}\alpha, \sqrt{1-\lambda}\beta)$ has at least three support points, i.e. there are points $A \le z_1 < z_2 < z_3 < \infty$ such that $f(z_i) = g(z_i)$, $i = 1, 2, 3$, where $f(z)$ denotes the polynomial $az^2 + bz + c$ on the left hand side of inequality (11). The mean value theorem implies that there exist points z_1', z_3' such that $z_1 < z_1' < z_2 < z_3' < z_3$ and $f'(z_i') = g'(z_i')$, $i = 1, 3$. Moreover, since $f(z) \le g(z)$ for $z \in [A, \infty)$, z_2 and z_3 are both tangent points, i.e. $f'(z_i) = g'(z_i)$, $i = 2, 3$. Applying the mean value theorem again, there exists z_i'' such that $z_1' < z_1'' < z_2 < z_2'' < z_3' < z_3'' < z_3$ and $f''(z_i'') = g''(z_i'')$, $i = 1, 2, 3$. Since $f''(z) = 2a$, we have $g''(z_i'') = 2a$, $i = 1, 2, 3$, which contradicts with condition (I). Therefore the Φ-optimal design ξ^* for estimating the vector of weighted parameters $K^T\theta = (\sqrt{\lambda}\alpha, \sqrt{1-\lambda}\beta)$ has at most two support points. From the discussion below Theorem 3, we obtain that ξ^* has at least two support points, hence ξ^* is supported on exactly two points. $\square$

Proof of Theorem 3 : For each $\lambda \in (0, 1)$, first we notice that the Φ-optimal design ξ_A^* for estimating the vector of weighted parameters $K^T\theta = (\sqrt{\lambda}\alpha, \sqrt{1-\lambda}\beta)$ on the restricted design space $[A, \infty)$ satisfies condition (11) on the interval $[A, \infty)$ but not on the whole real axis. Otherwise, ξ_A^* would also be optimal in the unrestricted situation. Since we assume that the smaller support point for the unrestricted Φ-optimal design is not included in $[A, \infty)$, we have two designs with different support points which are both optimal in the unrestricted sense. This contradicts the uniqueness of the support of the unrestricted Φ-optimal design. The above reasoning implies that there exists some $z' < A$ such that $f(z')h^2(z') > 1$. Moreover, from condition (II), it follows that there exists a point $z'' < z'$ such that for $z \le z''$ the inequality $f(z)h^2(z) \le 1$ holds again. Therefore we will encounter two intersection points z_0, z_1 between the functions $f(z)$ and $h^{-2}(z)$ on the interval $(-\infty, A]$. Assume that $z_0 < z_1 < A$ and denote by z_2, z_3 the support points of the design ξ_A^*, where $A < z_2 < z_3$, i.e. $f(z_i) = g(z_i)$ for $i = 0, 1, 2, 3$. By applying the mean value theorem to f and g and bearing in mind that z_2 and z_3 are tangent points, we obtain that there exist points z_1', $z_0 < z_1' < z_1 < A$ and z_2', $z_2 < z_2' < z_3$ with $f'(z_i') = g'(z_i')$, $i = 1, 2$ and $f'(z_i) = g'(z_i)$, $i = 2, 3$. A further application of the mean value theorem yields that there exist points z_i'', $i = 1, 2, 3$, $z_1' < z_1'' < z_2 < z_2'' < z_2' < z_3'' < z_3$ where $f''(z_i'') = g''(z_i'')$, $i = 1, 2, 3$, which leads to a contradiction with the fact that the equation $g''(z) = 2a$ can have at most two different solutions. Therefore, the smaller support point for the one-side restricted Φ-optimal design must be the left boundary point of the design interval. $\square$

5.3 Proofs of Theorem 4 and Lemma 2

Proof of Theorem 4 : By the same line of argument as in the proof of Theorem 3, we can show that there exist points z', z'', $z' > B$, $z'' < A$ such that $f'(z') = g'(z')$ and $f'(z'') = g'(z'')$. Next, we will show that a Φ-optimal design $\xi_{A,B}^*$ for estimating the vector of weighted parameters $K^T\theta = (\sqrt{\lambda}\alpha, \sqrt{1-\lambda}\beta)$ on the two-side restricted interval $[A, B]$ has only two support points by contradiction. Assume that $z_1 < z_2 < z_3$ are support points for $\xi_{A,B}^*$. The mean value theorem implies that there exist points z_1' and z_2' such that $z_1 < z_1' < z_2 < z_2' < z_3$ and $f'(z_i') = g'(z_i')$, $i = 1, 2$. Applying the mean value theorem to f' and g' again, we found three different values z_1'', z_2'', z_3'' such that $f''(z_i'') = g''(z_i'')$, $i = 1, 2, 3$. This leads to a contradiction to the fact that the equation $g''(z) = 2a$ can have at most two different solutions. Thus the Φ-optimal design $\xi_{A,B}^*$ has only two support points, which are given by the two boundary points A and B of the design interval $[A, B]$. If $A = -B$ and the conditions of part *(b)* of Theorem 2 are fulfilled, the optimality of the equally weighted design on $-B$, B follows along the same lines as in the proof of part *(b)* of Theorem 2. $\square$

Proof of Lemma 2 : The two designs under consideration are given by

$$\xi_{A,B}^* = \left\{ \begin{array}{cc} A & B \\ \omega_1 & 1-\omega_1 \end{array} \right\} \quad \text{and} \quad \xi_{-B,-A}^* = \left\{ \begin{array}{cc} -B & -A \\ 1-\omega_1 & \omega_1 \end{array} \right\}.$$

In order to determine the optimal weights at these points, we have to solve the system of equations in (14) for both designs. Now a straightforward but tedious calculation shows that both designs yield the same system of equations in (14). $\square$

References

Chaloner, K. and Larntz, K. (1989). Optimal Bayesian Design Applied to Logistic Regression Experiments. *Journal of Statistical Planning and Inference* **21**, 191-208.

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

Cook, R.D. and Wong, W.K. (1994). On the equivalence of constrained and compound optimal designs. *Journal of the American Statistical Association* **89**, 687-692.

Dai, Z. and Zhu, W. (2002). Locally compound optimal designs. *Department of Applied Mathematics and Statistics, State University of New York at Stony Brook, Technical Report* **18**.

Dette, H. (1997). Designing experiments with respect to standardized optimality criteria. *Journal of the Royal Statistical Society, Ser. B*, **59 (1)**, 97-110.

Flournoy, N. (1993). A clinical experiment in bone marrow transplantation: Estimating a percentage point of a quantal response curve. In: Case Studies in Bayesian Statistics, Editors: C. Gatsonis, J. S. Hodges, R. E. Kass, and N. D. Singpurwalla. Springer-Verlag, 324-335.

Ford, I., Torsney, B. and Wu, C.F.J. (1992). The use of a canonical form in the construction of locally optimal designs for non-linear problems. *Journal of the Royal Statistical Society* **54**, 569-583.

Gart, J.J., Krewski, D, Lee, P.N., Tarone, R.E. and Wahrendorf, J. (1986). Statistical methods in cancer research. Volume III–The design and analysis of long-term animal experiments. IARC Sci Publ. **79**, 1-219.

Gaudard, M.A., Karson, M.J., Linder, E. and Tse, S.K. (1993). Efficient designs for estimation in the power logistic quantal response model. *Statistica Sinica* **3**, 233-243.

Haines, L., Perevozskaya, I. and Rosenberger, W. (2003). Bayesian optimal designs for phase I clinical trials. *Biometrics* **59**, 591-600.

Hedayat, A. S., Yan, B. and Pezzuto, J. M. (1997). Modeling and identifying optimum designs for fitting dose-resonse curves based on raw optical density data. *Journal of the American Statistician Association* **92**, 1132-1140.

Kalish, L.A. (1990). Efficient design for estimation of median lethal dose and quantal dose-response curves. *Biometrics* **46**, 737-748.

Kuk, A.Y.C. (2004). A litter-based approach to risk assessment in developmental toxicity studies via a power family of completely monotone functions. *Journal of the Royal Statistical Society Series C* **53(2)**, 369 -402.

Lee, C.M.S. (1987). Constrained optimal designs for regression models. *Comm. Statist. Part A - Theory and Methods* **16**, 765-783.

Leuraud, K. and Benichou, J. (2001). A comparison of several methods to test for the existence of a monotonic dose-response relationship in clinical and epidemiological studies. *Statistics in Medicine* **20(22)**, 3335-3351.

Mats, V.A., Rosenberger, W.F. and Flournoy, N. (1997). Multiple-Objective Designs in Dose-Response Experiments. Institute of Mathematical Statistics Lecture Notes –Monograph Series: New Developments and Applications in Experimental Designs, 50-61.

Meier, K.L., Bailer, A.J. and Portier, C.J.(1993). A measure of tumorigenic potency incorporating dose-response shape. *Biometrics* **49**, 917-926.

Nagler, J. (1994). Scobit: an alternative estimator to logit and probit. *American Journal of Political Science* **38**, 230-255.

Prentice, R.L. (1976). A generalization of the probit and logit methods for dose response curves. *Biometrics* **32**, 761-768.

Pukelsheim, F. (1993). Optimal Design of Experiments. Wiley, New York.

Pukelsheim, F. and Torsney, B. (1991). Optimal weights for experimental designs on linearly independent support points. *Annals of Statistics* **19 (3)**, 1614-1625.

Rosenberger, W.F. and Grill, S.E. (1997). A sequential design for psychophysical experiments: An application to estimating timing of sensory events. *Statistics in Medicine* **16**, 2245-2260.

Silvey, S.D. (1980). Optimal Design. Chapman and Hall, London.

Stukel, T. A. (1988) Generalized logistic models. *Journal of the American Statistical Association* **83**, 426-431.

Throne, J.E., Weaver, D.K., Chew, V. and Baker, J.E. (1995). Probit analysis of correlated data: multiple observations over time at one concentration. *Journal of Economic Entomology* **88(5)**, 1510-1512.

Wang, S.J. and Hung, H.M. (1997). Large sample tests for binary outcomes in fixed-dose combination drug studies. *Biometrics* **53(2)**, 498-503.

Wu, C.F.J. (1988). Optimal design for percentile estimation of a quantal response curve. In: Optimal Design and Analysis of Experiments, Editors: J. Dodge, V.V. Fedorov and H.P. Wynn. North Hollander, Amsterdam, 213-233.

Zeng, Q. and Zhu, W. (1997). Optimal designs with multiple objectives for rheumatoid arthritis dose-rangeing study using ACR responder index. *Merck Technical Report* **54**.

Zhu, W. and Wong, W.K. (1998). Multiple-Objective Designs in Dose-Response Experiments. Institute of Mathematical Statistics Lecture Notes –Monograph Series: New Developments and Applications in Experimental Designs, 73-82.

Zhu, W. and Wong, W.K. (2000). Multiple-Objective Designs in Dose-Response Experiments. *Journal of Biopharmaceutical Statistics* **10**, 1-14.

Zhu, W. and Wong, W.K. (2001). Bayesian Optimal Designs for Estimating a Set of Symmetrical Quantiles. *Statistics in Medicine* **20**, 123-137.

Table 2: Selected unrestricted compound optimal designs for the Merck dose ranging trial in terms of percentiles

$\tilde{\lambda}$	.1	.2	.3	.4	.5	.6	.7	.8	.9
x_1	ED11.4	ED13.6	ED15.5	ED17.2	ED18.9	ED20.7	ED22.7	ED25.2	ED28.8
x_2	ED88.6	ED86.4	ED84.5	ED82.8	ED81.1	ED79.3	ED77.3	ED74.8	ED71.2

Table 3: Selected unrestricted compound optimal designs for the Merck dose ranging trial in terms of support points

$\tilde{\lambda}$	.1	.2	.3	.4	.5	.6	.7	.8	.9
x_1	-56.94	-48.94	-42.96	-37.90	-33.26	-28.73	-23.99	-18.56	-11.23
x_2	106.94	98.94	92.96	87.90	83.26	78.73	73.99	68.56	61.23

Table 4: Selected left-restricted compound optimal designs in terms of support points x_1, x_2 and normalized support points z_1, z_2

$\tilde{\lambda}$	z_1	z_2	x_1	x_2	ω_1	ω_2
.1	-0.625	2.606	0	129.25	0.385	0.615
.2	-0.625	2.478	0	124.10	0.428	0.572
.3	-0.625	2.360	0	119.39	0.464	0.536
.4	-0.625	2.242	0	114.70	0.496	0.504
.5	-0.625	2.118	0	109.71	0.526	0.474
.6	-0.625	1.977	0	104.07	0.553	0.447
.7	-0.625	1.805	0	97.21	0.578	0.422
.8	-0.625	1.575	0	88.02	0.597	0.403
.9	-0.625	1.213	0	73.51	0.598	0.402

Table 5: Selected 2-side unsymmetrically-restricted compound optimal designs (supported at the boundary values $x_1 = 0$, $x_2 = 60$ ($z_1 = -0.625$, $z_2 = 0.875$))

$\tilde{\lambda}$	.1	.2	.3	.4	.5	.6	.7	.8	.9
ω_1	0.492	0.495	0.499	0.503	0.509	0.516	0.524	0.535	0.550
ω_2	0.508	0.505	0.501	0.497	0.491	0.484	0.476	0.465	0.450