

# Optimal designs for discriminating dose response models in toxicology studies

Holger Dette

Ruhr-Universität Bochum

Fakultät für Mathematik

Universitätsstrasse 150

44780 Bochum, Germany

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

Andrey Pepelyshev

St. Petersburg State University

Department of Mathematics

Universitetskij pr. 28

St. Petersburg, 198504, Russia

email: andrey@ap7236.spb.edu

Piter Shpilev

St. Petersburg State University

Department of Mathematics

Universitetskij pr. 28

St. Petersburg, 198504, Russia

email: pitsh@front.ru

Weng Kee Wong\*

University of California at Los Angeles

Department of Biostatistics

10833 Le Conte Ave.

Los Angeles, CA 90095, USA

email: wkwong@ucla.edu

6 July 2009

---

\*Correspondence author

## Abstract

We consider design issues for toxicology studies when we have a continuous response but the true mean response is only known to be a member in a class of nested models. This class of models were proposed by toxicologists who were concerned with only estimation problems. We develop robust and efficient designs for model discrimination and optimal designs for estimating parameters in the selected model at the same time. In particular, we propose designs that maximize the minimum of  $D$ - or  $D_1$ -efficiencies over all models in the given class. We show that these optimal designs are efficient for determining an appropriate model from the postulated class, quite efficient for estimating model parameters in the identified model and also robust with respect to model mis-specification. To facilitate use of these designs in practice, we have also constructed a web site to enable practitioners to generate optimal designs for their problems.

Keyword and phrases: continuous design, local optimal design, maximin optimal design, model discrimination, robust design.

# 1 Introduction

This paper addresses design issues for toxicology studies when the primary outcome is continuous and it is not known a priori which model is an appropriate one to use. Under this situation, one may consider a class of plausible models within which we believe lies an adequate model for fitting the data at hand. The issues of interest are how to design the study to ascertain the most appropriate model from within the postulated class of models and at the same time able to estimate the parameters of the selected model efficiently. Our design decisions include how to select the number of dose levels to observe the continuous outcome, where these levels are and how many repeated observations to take at each of these levels. Ideally, we want the design to be able to identify the correct model within the postulated class of models and also provide efficient estimates for the parameters in the identified model. In this paper, it is further assumed for simplicity that there is only one independent variable, the dose level, and all design issues have to be decided in advance of the study. Sequential designs are not considered.

Addressing design issues invariably require model assumptions that specify how the mean outcome relates to the independent variable. Usually a specific functional form is assumed either from experts' opinions or from the science of the problem, see Gaylor and Chen (1993), Catalano et al. (1993), Slob and Pieters (1998), Oscar (2004), Moerbeek et al. (2004), among several others. When it is problematic to specify a single model to describe the functional relationship between the mean outcome and the dose level, a strategy is to work with a class of plausible nested models assumed to include the 'true' model. This class of models is usually arrived at after consultation with experts in the area. The key research question is how to design the study to identify an appropriate model from the postulated class of models and at the same time able to estimate the model parameters efficiently. Such design problems are important and arise frequently in practice across disciplines. To our knowledge, only a couple of papers have tried to address such design questions. A recent example is Dette et al. (2005) where they wanted to discriminate between two popular nonlinear mod-

els and at the same time wanted to estimate model parameters, regardless which one of the two is the more appropriate model. The two models are the Michaelis-Menten model and the Emax model which are popular among scientists in enzyme-kinetic studies. These are nested models because when one of the parameters in the Emax model is set to unity, the Michaelis-Menten obtains. A main reason for lack of research in addressing design issues in such a setup is that there are serious technical difficulties involved, especially for nonlinear models.

The motivation for this work comes from repeated proposals recently to use a class of models to study a continuous outcome in toxicological studies [Moerbeek et al. (2004), Piersma et al. (2002), Woutersen et al. (2001), Slob (2002)]. In all these papers, the interest was only in estimation problems and so they did not consider design issues. As is typical in such publications, the rationale for the design or designs employed is either lacking or not properly explained. Our purpose here is to develop optimal designs for identifying an appropriate model within the class of models and also at the same time provide reliable estimates in the selected model. Design issues are always difficult to address and we begin first by considering local optimal designs because they are the easiest to construct for nonlinear models [Chernoff, (1953)]. However it is well known that such designs are sensitive to nominal values and more so on the mean function in the model specification. To overcome the risk of selecting an inappropriate model, we propose maximin optimal designs that appear to be robust to mis-specification of the model in other settings. These maximin optimal design maximizes the minimum efficiency regardless which model in the class of models is the appropriate model. As such, these optimal designs provide some global protection against picking the wrong model from the postulated class of models. As we will show, they are also quite robust to mis-specification in the mean function and seem to provide good efficiencies for estimating parameters in the selected model.

In section 2, we present background and the proposed class of models. We describe relationships among models in the class and provide local optimal designs for discriminating

between plausible models. We also show how optimal designs constructed for a set of design parameters can be used to deduce the optimal design under another set of design parameters. In section 3, we construct maximin optimal designs for various subclasses of plausible models and show in section 4 that maximin optimal designs are robust to mis-specification of models in the postulated class. We offer a conclusion in section 5 and all technical justifications are given in the appendix.

## 2 A class of dose response models

In a series of papers published in the toxicology literature, Woutersen et al. (2001), Piersma et al. (2002), Slob (2002), Moerbeek et al. (2004), among others, proposed a class of dose response models for modeling the effect of a drug in toxicological studies. On the user-selected interval  $[0, T]$ , the nonlinear regression model is given by

$$Y = \eta(t, \theta) + \varepsilon, \quad (2.1)$$

where  $\varepsilon$  is a normally distributed error term with mean 0 and constant variance  $\sigma^2 > 0$ . We assume the unknown parameter  $\theta$  is  $m$ -dimensional and all observations are independent. The expectation of  $Y$  under experimental condition  $t$  is  $E[Y|t] = \eta(t, \theta)$ , and  $\eta(t, \theta)$  is one of the functions defined below:

$$\eta(t, \theta) = a ; m = 1, \theta = a > 0 \quad (2.2)$$

$$\eta(t, \theta) = ae^{-bt} ; m = 2, \theta = (a, b)^T, a > 0, b > 0 \quad (2.3)$$

$$\eta(t, \theta) = ae^{-bt^d} ; m = 3, \theta = (a, b, d)^T, a, b > 0, d \geq 1 \quad (2.4)$$

$$\eta(t, \theta) = a(c - (c - 1)e^{-bt}) ; m = 3, \theta = (a, b, c)^T, a, b > 0, c \in [0, 1] \quad (2.5)$$

$$\eta(t, \theta) = a(c - (c - 1)e^{-bt^d}) ; m = 4, \theta = (a, b, c, d)^T, a, b > 0, c \in [0, 1], d \geq 1. \quad (2.6)$$

The authors gave arguments why this is a reasonable class to use for modeling continuous (toxicological) endpoints in dose response relationships that cannot be derived from biological mechanism. Note that the different models are nested, in the sense that the models with a lower number of parameters can be obtained from an extended model by special choices

of the parameters. For instance, model (2.6) is an extension of the models (2.5) and (2.4), model (2.4) is an extension of model (2.3), model (2.5) is an extension of the models (2.3) and (2.2). The hierarchy of the different models is illustrated in the following diagram.

$$\begin{array}{ccccc}
(2.6) & \xrightarrow{d=1} & (2.5) & \xrightarrow{c=1} & (2.2) \\
& & \Downarrow c=0 & & \Downarrow c=0 \\
(2.4) & \xrightarrow{d=1} & (2.3) & & 
\end{array}$$

We note that when  $b = 0$  all models (2.3) - (2.6) reduce to the constant model (2.2), and this relation is not shown in the diagram.

Following Kiefer (1974), we consider only continuous designs. A continuous design is simply a probability measure  $\xi$  with a finite number of support points, say  $t_1, \dots, t_n \in [0, T]$  and corresponding weights  $\omega_1, \dots, \omega_n$  with  $\omega_i > 0$ ,  $\sum_{i=1}^n \omega_i = 1$ . If we fix the number of observations  $N$  in advance, either by cost or time considerations, then roughly  $n_i = N\omega_i$  observations are taken at point  $t_i$ , with  $\sum_{i=1}^n n_i = N$ . The reason for working with continuous designs is that they can be described analytically for many problems and so they are easier to study.

Jennrich (1969) showed that under regularity assumptions, the asymptotic covariance matrix of the standardized least squares estimator  $\sqrt{N/\sigma^2} \hat{\theta}$  for the parameter  $\theta$  in the model (2.1) is given by the matrix  $M^{-1}(\xi, \theta)$ , where

$$M(\xi, \theta) = \int_0^T f(t, \theta) f^T(t, \theta) d\xi(t)$$

is the information matrix using design  $\xi$  for the model (2.1) and

$$f(t, \theta) = \frac{\partial \eta}{\partial \theta}(t, \theta) = (f_1(t, \theta), \dots, f_m(t, \theta))^T \tag{2.7}$$

is the vector of partial derivatives of the conditional expectation  $\eta(t, \theta)$  with respect to the parameter  $\theta$ . Additionally, we consider only designs with a non-singular information matrix. A sufficient condition for this property to hold is that the design has  $k$  support points and

$k$  exceeds the number of parameters in the model.

A local optimal design maximizes an appropriate function of the information matrix  $M(\xi, \theta)$  using nominal values of  $\theta$  (Chernoff, 1953). There are several commonly used optimality criteria for estimating purposes and for discriminating among competing designs [see Silvey (1980) or Pukelsheim (1993)]. Our interests here are to find efficient designs for model selection among models defined by (2.2) – (2.6) and also provide good and robust estimates for the parameters in the selected model. For this purpose we construct an optimal experimental design for the following pairs of competing models that fulfills at least two requirements.

- (1) The design should allow to test the hypothesis corresponding to the problem of discriminating between two rival models. For example the hypothesis for discriminating between the model (2.5) and (2.3) is given by

$$H_0 : c = 0 \quad \text{vs} \quad H_1 : c \in (0, 1]$$

- (2) The design should be efficient for the estimation of the parameters in the corresponding pair of the regression models and for all models which are sub models of the model with the larger number of parameters. For example, for the model (2.5) the corresponding sub models are given by (2.3) and (2.4).
- (3) The design should also be efficient for discriminating between the different sub models of the model with the larger number of parameters (which may also be nested). For example, the optimal design for discriminating between the models (2.3) and (2.5) should also be efficient for discriminating between the models (2.2)/(2.5) and (2.2)/(2.3).

In the rest of this section we concentrate on local optimal designs, and restrict ourselves to the  $e_m$ -optimality criterion, where  $e_m = (0, \dots, 0, 1)^T$  denotes the  $m$ th unit vector (and  $m$  is the larger of the number of parameters in the two regression models under consideration). For a fixed  $\theta$  a local  $e_m$ -optimal design minimizes

$$e_m^T M^{-1}(\xi, \theta) e_m = \frac{\det \widetilde{M}(\xi, \theta)}{\det M(\xi, \theta)}, \quad (2.8)$$

where the matrix  $M(\xi, \theta)$  is the information matrix in the model with the larger number of parameters, and the matrix  $\widetilde{M}(\xi, \theta)$  is obtained from  $M(\xi, \theta)$  by deleting the  $m$ -th row and the  $m$ -th column. We note that the expression (2.8) is proportional to the asymptotic variance of the least squares estimate for the parameter  $e_m^T \theta$ , which is relevant for model discrimination. In other words, minimizing the asymptotic variance (2.8) provides a design with maximal power for testing a simple hypothesis for the parameter  $e_m^T \theta$ . For example, in the problem of discriminating between model (2.5) and (2.3) we have  $m = 3$ ,  $e_m^T \theta = (0, 0, 1)(a, b, c)^T = c$  and the cases  $c \neq 0$  and  $c = 0$  give the two rival models (2.5) and (2.3), respectively. Consequently, a design that minimizes expression (2.8), which is proportional to the variance of the least squares estimate for the parameter  $c$ , is optimal for discriminating between the two models.

In the following subsections we construct local  $e_m$ -optimal designs for the pairs of the models (2.4) – (2.6). We give a detailed description for the model (2.4) and briefly summarize the corresponding results for the other models.

## 2.1 Optimal discriminating designs for the models (2.3) and (2.4)

For the model (2.4) we have  $\theta = (a, b, d)^T$  and a straightforward calculation yields that the vector of partial derivatives defined by (2.7) is given by

$$f(t, \theta) = f(t, a, b, d) = \left( e^{-bt^d}, -at^d e^{-bt^d}, -abt^d \ln(t) e^{-bt^d} \right)^T. \quad (2.9)$$

Our first result establishes basic properties of local  $e_3$ -optimal designs for model (2.4), and facilitates their numerical calculation.

**Lemma 2.1** *The local  $e_3$ -optimal design in the regression model (2.4) does not depend on the parameter  $a$ . Moreover, if  $t_i(b, d, T)$  is a support point of a local  $e_3$ -optimal design on the interval  $[0, T]$  with corresponding weight  $\omega_i(b, d, T)$ , then for any  $r > 0$  and  $d > 0$ ,*

$$\begin{aligned} t_i(b, d, T^{\frac{1}{d}}) &= t_i(b, 1, T)^{\frac{1}{d}}, & \omega_i(b, d, T^{\frac{1}{d}}) &= \omega_i(b, 1, T), \\ t_i(rb, 1, T) &= \frac{1}{r} t_i(b, 1, rT), & \omega_i(rb, 1, T) &= \omega_i(b, 1, rT). \end{aligned} \quad (2.10)$$



Note that we are interested in efficient designs for discriminating between models (2.3) and (2.4). Because this problem is most difficult if the parameter  $d$  is close to unity we assume that an initial parameter value of  $d$  is set equal to unity throughout this section. The lemma also implies that it is enough to calculate local  $e_3$ -optimal designs on a fixed design space for various values of  $b$  after the remaining parameters  $a, d$  are fixed. Local optimal designs on a different design space or have other values of the parameters can then be calculated using the relationships in the lemma.

Our next result characterizes the local  $e_3$ -optimal designs explicitly. We recall that a set of  $k$  functions  $h_1, \dots, h_k : I \rightarrow \mathbb{R}$  is a Chebyshev system (on the interval  $I$ ) if there exists an  $\varepsilon \in \{-1, 1\}$  such that the inequality

$$\varepsilon \cdot \begin{vmatrix} h_1(t_1) & \dots & h_1(t_k) \\ \vdots & \ddots & \vdots \\ h_k(t_1) & \dots & h_k(t_k) \end{vmatrix} > 0$$

holds for all  $t_1, \dots, t_k \in I$  with  $t_1 < t_2 < \dots < t_k$ . From Karlin and Studden (1966, Theorem II 10.2), we know that if  $\{h_1, \dots, h_k\}$  is a Chebyshev system, there exists a unique function, say  $\sum_{i=1}^k c_i^* h_i(t) = c^{*T} h(t)$ , ( $h = (h_1, \dots, h_k)^T$ ) with the following properties

- (i)  $|c^{*T} h(t)| \leq 1 \quad \forall t \in I$
- (ii) there exist  $k$  points  $t_1^* < \dots < t_k^*$  such that  $c^{*T} h(t_i^*) = (-1)^i$ ,  $i = 1, \dots, k$ .

The function  $c^{*T} h(t)$  alternates at the points  $t_1^*, \dots, t_k^*$  and is called the Chebyshev polynomial. The points  $t_1^*, \dots, t_k^*$  are called Chebyshev points and they need not to be unique. However, they are unique in many cases, in particular if  $1 \in \text{span}\{h_1, \dots, h_k\}$ ,  $k \geq 1$  and  $I$  is a bounded and closed interval. In this case it follows  $t_1^* = \min_{t \in I} t$ ,  $t_k^* = \max_{t \in I} t$ . The following result characterizes the local  $e_3$ -optimal design.

**Theorem 2.1** *The components of the vector defined by (2.9) form a Chebyshev system on the interval  $[0, T]$ . The local optimal  $e_3$ -optimal design for model (2.4) is unique and is supported at the uniquely determined three Chebyshev points, say  $t_1^* < t_2^* < t_3^*$ . The corre-*

sponding weights  $\omega_1^*, \omega_2^*, \omega_3^*$  can be obtained explicitly as

$$\omega^* = (\omega_1^*, \omega_2^*, \omega_3^*)^T = \frac{JF^{-1}e_3}{\mathbf{1}_3 JF^{-1}e_3}, \quad (2.11)$$

where the matrices  $F$  and  $J$  are defined by  $F = (f(t_1^*, \theta), f(t_2^*, \theta), f(t_3^*, \theta))$ ,  $J = \text{diag}(1, -1, 1)$ , respectively, and  $\mathbf{1}_3 = (1, 1, 1)^T$ .

Local  $e_3$ -optimal designs for model (2.4) on the design space  $[0, T]$  with  $T = 1$  are given in Table 1 for various values of the parameter  $b$ . We observe that the boundary points of the design space are support points of the local optimal designs if  $T$  is enough small, i.e.  $t_1^* = 0$  and  $t_3^* = 1$ .

Table 1: Local  $e_3$ -optimal designs for models (2.4) and (2.5) on the design space  $[0, 1]$  for various values of the parameter  $b$ .

	model (2.4)						model (2.5)					
$b$	$t_1$	$t_2$	$t_3$	$\omega_1$	$\omega_2$	$\omega_3$	$t_1$	$t_2$	$t_3$	$\omega_1$	$\omega_2$	$\omega_3$
0.1	0	0.355	1	0.311	0.500	0.189	0	0.492	1	0.242	0.500	0.259
0.5	0	0.305	1	0.294	0.493	0.213	0	0.458	1	0.212	0.492	0.296
1.0	0	0.251	1	0.276	0.473	0.251	0	0.418	1	0.180	0.469	0.351
2.0	0	0.167	1	0.241	0.403	0.356	0	0.343	1	0.127	0.384	0.490
3.0	0	0.112	0.751	0.232	0.381	0.387	0	0.281	1	0.083	0.267	0.650

## 2.2 Optimal discriminating designs for the models (2.4) and (2.5) and (2.2) and (2.5)

We now briefly mention the corresponding results for regression model (2.5). In this case we have  $\theta = (a, b, c)^T$  and when  $c = 0$  or  $c = 1$ , model (2.5) reduces to model (2.4) or (2.2), respectively. Therefore the  $e_3$ -optimal design is optimal for two purposes, namely, discriminating between (2.4) and (2.5) and discriminating between (2.2) and (2.5). A direct

calculation shows the vector of partial derivatives in (2.7) is

$$f(t, \theta) = f(t, a, b, c) = (c - (c - 1)e^{-bt}, a(c - 1)te^{-bt}, a(1 - e^{-bt}))^T.$$

It can be shown that the components of this vector form a Chebyshev system on the interval  $[0, T]$  and that the local  $e_3$ -optimal design does not depend on the parameter  $a$ . Moreover, for other positive values of  $b$  and  $T$ , the support points  $t_i(b, T)$  and corresponding weights  $\omega_i(b, T)$  of the optimal design are found using the relationships  $t_i(rb, T) = \frac{1}{r}t_i(b, rT)$  and  $\omega_i(rb, T) = \omega_i(b, rT)$ .

**Theorem 2.2** *Let  $0 \leq c < 1$ . The  $e_3$ -optimal design for model (2.5) is unique and it has three points, say  $t_1^* < t_2^* < t_3^*$ . Moreover,  $t_1^* = 0$ ,  $t_3^* = T$  and*

$$t_2^* = \frac{1}{b} + \frac{t_1^*e^{-bt_1^*} - t_3^*e^{-bt_3^*}}{e^{-bt_1^*} - e^{-bt_3^*}},$$

*and the corresponding weights  $\omega_1^*, \omega_2^*, \omega_3^*$  can be obtained explicitly by formula (2.11).*

### 2.3 Optimal discrimination designs for the models (2.5) and (2.6), (2.2) and (2.6), (2.4) and (2.6)

For model (2.6) we have  $\theta = (a, b, c, d)^T$ . This model reduces to model (2.4), (2.2) and (2.5), when  $c = 0$ ,  $c = 1$  and  $d = 1$  respectively. Therefore, we are interested in  $e_3$ -optimal designs corresponding to hypothesis about the parameter  $c$  and  $e_4$ -optimal designs referring to the parameter  $d$ . A direct calculation shows the vector of partial derivatives of  $\eta$  for model (2.6) is

$$f(t, \theta) = (c - (c - 1)e^{-bt^d}, a(c - 1)t^d e^{-bt^d}, a(c - 1)t^d \ln(t)be^{-bt^d}, a(1 - e^{-bt^d}))^T \quad (2.12)$$

and forms a Chebyshev system on the interval  $[0, T]$ . Similar arguments as given in the proof of Lemma 2.1 show that the support points  $t_i(b, d, T)$  and weights  $\omega_i(b, d, T)$  of a local  $e_3$ - or  $e_4$ -optimal design on the interval  $[0, T]$  satisfy relations (2.10). Moreover, the optimal designs do not depend on the parameter  $a$ .

The following result describe the  $e_3$ - and  $e_4$ -optimal design for the model (2.6). The proof is similar to the proof of Theorem 2.1 and therefore omitted. Some numerical results are listed in Table 2.

**Theorem 2.3** *The  $e_3$ - and  $e_4$ -optimal designs for model (2.6) are unique and supported at the unique determined four Chebyshev points, say  $t_1^* < t_2^* < t_3^* < t_4^*$  corresponding to the Chebyshev system defined by the components of the vector in (2.12). The corresponding weights  $\omega_1^*, \dots, \omega_4^*$  are explicitly given by*

$$\omega^* = (\omega_1^*, \dots, \omega_4^*)^T = \frac{JF^{-1}e_k}{\mathbf{1}_4 JF^{-1}e_k}, \quad k = 3, 4,$$

where the matrices  $F$  and  $J$  are defined by  $F = (f(t_1^*, \theta), f(t_2^*, \theta), f(t_3^*, \theta), f(t_4^*, \theta))$ ,  $J = \text{diag}(1, -1, 1, -1)$ , respectively,  $\mathbf{1}_4 = (1, 1, 1, 1)^T$  and  $f(t, \theta)$  is given in (2.12).

Table 2: Local  $e_3$ - and  $e_4$ -optimal designs for model (2.6) on the design space  $[0, 1]$  for various values of the parameter  $b$ .

$b$					$e_3$ -optimal				$e_4$ -optimal			
	$t_1$	$t_2$	$t_3$	$t_4$	$\omega_1$	$\omega_2$	$\omega_3$	$\omega_4$	$\omega_1$	$\omega_2$	$\omega_3$	$\omega_4$
0.1	0	0.131	0.648	1	0.286	0.416	0.214	0.084	0.174	0.328	0.326	0.172
0.5	0	0.123	0.626	1	0.277	0.410	0.223	0.090	0.156	0.302	0.342	0.200
1.0	0	0.113	0.596	1	0.267	0.403	0.233	0.097	0.137	0.272	0.352	0.239
2.0	0	0.094	0.530	1	0.253	0.392	0.246	0.108	0.106	0.215	0.341	0.338
3.0	0	0.079	0.463	1	0.244	0.382	0.256	0.118	0.080	0.163	0.289	0.468

### 3 Maximin optimal discriminating designs

In general, when a class of models is available, an optimal discriminating design should be efficient for testing several hypotheses among the competing models. For this reason we now consider the problem of constructing an experimental design, that is efficient for testing several hypotheses corresponding to the discrimination problem between the models (2.4) and (2.3), (2.5) and (2.3), (2.6) and (2.4), and, (2.6) and (2.5). In section 4, we investigate the efficiencies of these optimal designs for estimating parameters in the various models,

including the identified model.

To fix ideas, consider the problem of finding an optimal design to discriminate between two models (2.*i*) and (2.*j*). Let

$$\text{eff}^{(2.i)-(2.j)}(\xi, \theta)$$

be the efficiency of the design  $\xi$  for discriminating between the two models. To calculate this quantity, consider an illustrative case for finding the local optimal design for discriminating between the models (2.4) and (2.3). This optimal design minimizes  $e_3^T M_{(2.4)}^{-1}(\xi, \theta) e_3$  in the class of all designs for which the matrix is regular (see Theorem 2.1). Here the matrix  $M_{(2.4)}(\xi, \theta)$  is the information matrix relative to model (2.4). If  $\xi_3^*(\theta)$  is the local optimal design for discriminating between models (2.4) and (2.3), the efficiency of a design  $\xi$  for discriminating between the models (2.4)-(2.3) is defined by

$$\text{eff}^{(2.4)-(2.3)}(\xi, \theta) = \frac{e_3^T M_{(2.4)}^{-1}(\xi_3^*(\theta), \theta) e_3}{e_3^T M_{(2.4)}^{-1}(\xi, \theta) e_3}.$$

This ratio is always between 0 and 1; if the value is 0.5, this means that twice as many observations are required from the design  $\xi$  than the optimal design, to discriminate the two models with the same level of precision. The efficiencies of  $\xi$  for discriminating other pairs of models are similarly defined and denoted by  $\text{eff}^{(2.5)-(2.3)}(\xi, \theta)$ ,  $\text{eff}^{(2.6)-(2.4)}(\xi, \theta)$  and  $\text{eff}^{(2.6)-(2.5)}(\xi, \theta)$ . Here and elsewhere in our work, we remind readers that we assume  $\theta$  is fixed throughout and so all optimal designs are only locally optimal.

We next apply the maximin efficient approach proposed by Dette (1995), Müller (1995) to find efficient designs for all four discrimination problems [see also Müller and Pázman (1998)]. More precisely, for fixed  $\theta$ , we call a design maximin optimal discriminating design for the models (2.2) – (2.6), if it maximizes

$$\min \left\{ \text{eff}^{(2.4)-(2.3)}(\xi, \theta), \text{eff}^{(2.5)-(2.3)}(\xi, \theta), \text{eff}^{(2.6)-(2.4)}(\xi, \theta), \text{eff}^{(2.6)-(2.5)}(\xi, \theta) \right\}. \quad (2.13)$$

In general maximin optimal discriminating designs have to be found numerically in all cases of practical interest. Table 3 shows maximin optimal discriminating designs for the param-

eter  $\theta = (a, b, d, c)^T = (1, b, 1, 0)^T$  for different values  $b$  and their efficiencies. We observe that the maximin optimal discriminating design yields are between 68% – 85% efficient for discriminating between different pairs of rival models from the postulated class.

Table 3: *Maximin optimal discriminating designs for the optimality criterion (2.13) on the design space  $[0, 1]$  and their efficiencies.*

$b$	$t_1$	$t_2$	$t_3$	$t_4$	$\omega_1$	$\omega_2$	$\omega_3$	$\omega_4$	(2.4)-(2.3)	(2.5)-(2.3)	(2.6)-(2.4)	(2.6)-(2.5)
0.1	0	0.175	0.552	1	0.236	0.255	0.322	0.187	0.724	0.724	0.724	0.786
0.5	0	0.170	0.531	1	0.220	0.260	0.308	0.212	0.719	0.719	0.719	0.787
1.0	0	0.160	0.507	1	0.200	0.265	0.287	0.249	0.714	0.714	0.714	0.793
2.0	0	0.130	0.468	1	0.161	0.250	0.249	0.340	0.705	0.702	0.702	0.848
3.0	0	0.105	0.440	1	0.141	0.233	0.199	0.427	0.705	0.682	0.682	0.871

## 4 Efficiencies of maximin optimal designs for estimating model parameters under model uncertainty

In this section we investigate performance of maximin discrimination designs for estimating parameters in the different models. We first present results for estimating each parameter in the model and D-efficiencies of the maximin discrimination design for estimating all parameters in the model. We recall that D-efficiencies are computed relative to the D-optimal design for the specific model and D-optimal designs are found by maximizing the determinant of the expected information matrix over all designs on the design space. D-optimal designs are appealing because they minimize the generalized variance and so provide the smallest possible confidence ellipsoid for all parameters in the mean function.

In Table 4, we display efficiencies of selected maximin optimal discriminating design for

estimating the individual parameters in models (2.3) and (2.4). We observe that the efficiencies for estimating the parameter  $a$  are consistently low, but with improving efficiencies for estimating  $b$  and  $d$ . The average efficiency for estimating  $a, b$  in model (2.3) are 0.364 and 0.511 respectively, and the average efficiency for estimating  $a, b$  and  $d$  in model (2.4) respectively are 0.235, 0.494 and 0.725. For model (2.5), mean efficiencies are 0.271, 0.727 and 0.722 for estimating  $a, b$  and  $c$ , and for model (2.6), the mean efficiencies are 0.218, 0.767, 0.722 and 0.785 for estimating  $a, b, c$  and  $d$ . It is not surprising to observe that the efficiencies are highest for estimating the particular parameter that sets the two models apart.

Table 4: *Efficiencies of the maximin optimal discriminating designs in Table 3 for estimating individual coefficients in model (2.3), (2.4), (2.5) and (2.6). The first two columns are efficiencies for estimating  $a$  and  $b$  in model (2.3), then the three columns are for estimating  $a, b, d$  in model (2.4), then the three columns are efficiencies for estimating  $a, b$  and  $c$  in model (2.5) and the last four columns are for estimating  $a, b, c$  and  $d$  in model (2.6).*

	model (2.3)		model (2.4)			model (2.5)			model (2.6)			
$b$	eff <sub>1</sub>	eff <sub>2</sub>	eff <sub>1</sub>	eff <sub>2</sub>	eff <sub>3</sub>	eff <sub>1</sub>	eff <sub>2</sub>	eff <sub>3</sub>	eff <sub>1</sub>	eff <sub>2</sub>	eff <sub>3</sub>	eff <sub>4</sub>
0.1	0.42	0.495	0.26	0.495	0.724	0.30	0.726	0.724	0.24	0.784	0.724	0.786
0.5	0.37	0.501	0.25	0.490	0.719	0.27	0.725	0.719	0.22	0.773	0.719	0.787
1.0	0.32	0.545	0.22	0.495	0.714	0.25	0.716	0.714	0.20	0.760	0.714	0.793
2.0	0.24	0.609	0.17	0.325	0.705	0.21	0.661	0.702	0.16	0.759	0.702	0.849
3.0	0.20	0.429	0.15	0.514	0.705	0.18	0.560	0.682	0.14	0.708	0.682	0.871

Table 5 shows D-efficiencies of the maximin optimal discriminating designs in Table 3. These are efficiencies relative to each of the local D-optimal designs found for each model in the class. D-optimal designs are suitable for estimating all the parameters in the model and are probably the most popular and over-used optimal designs in the literature. For the values of  $b$  in Table 5, all efficiencies are very high. Recall that the optimal discriminating designs were constructed for discriminating between models (2.3) and (2.4) and we observe

that these efficiencies are highest for the most complicated model (2.6) averaging 96% while the efficiencies are about 67% for the least complicated model (2.3). This implies that the maximin optimal designs are quite robust to mis-specification of models in the class of models and also quite insensitive to small changes to the nominal values of the parameter  $b$  common to all the models. In models (2.3) and (2.4), we note that the D-efficiencies can vary by roughly 15% or 10% when the nominal value of  $b$  is increased from 2 to 3.

Table 5: *D-efficiencies for the maximin designs from Table 3 under different model assumptions.*

$b$	$\text{eff}_D^{(2.3)}$	$\text{eff}_D^{(2.4)}$	$\text{eff}_D^{(2.5)}$	$\text{eff}_D^{(2.6)}$
0.1	0.710	0.851	0.851	0.963
0.5	0.737	0.862	0.861	0.968
1.0	0.786	0.873	0.869	0.972
2.0	0.703	0.864	0.860	0.959
3.0	0.525	0.716	0.820	0.917

## 5 Conclusions

Our work is motivated from toxicologists' recent interest in working with a class of nonlinear nested models in studies where the outcomes are continuous. The toxicologists were primarily interested in estimating model parameters or function of model parameters. The designs employed in their studies lacked justifications. Our work is the first to address design issues for such a problem, where there is model uncertainty and all candidate models are nonlinear models nested among one another. The optimal designs proposed here are efficient for model discrimination and parameter estimation at the same time. Previous work on an explicit construction of optimal designs for discriminating between nonlinear models usually focuses on two possible models; our work represents among the first to propose efficient designs for discriminating among several nonlinear models. In addition, our work provides closed for-



mulae for describing local optimal designs for discriminating between pairs of models within the stipulated class.

Our proposed optimal designs was constructed using large sample theory. The variances of the parameters were obtained via the asymptotic covariance matrix and our optimal designs minimize the asymptotic variances. It is reasonable to ask if the asymptotic variance is a good approximation to the actual variance of the parameters in practice with realistic sample size. To this end, we briefly investigate this issue using a small and limited simulation study.

For example, in Piersma et al. (2002), rats were prenatally exposed to diethylstilbestrol and the design  $\xi_u$  had 6 animals in each of the 10 dose groups at 0, 1.0, 1.7, 2.8, 4.7, 7.8, 13, 22, 36 and 60 mg/kg body weight per day. Thus we have 60 observations on the design space  $[0, 60]$ . The maximin design  $\xi_{mm}$  for  $b = 0.1$ ,  $d = 1$ ,  $c = 0$  is given by

$$\{0, 3.6, 24, 60; 7/60, 12/60, 13/60, 28/60\}.$$

We simulated data with  $a = 1$  and  $\sigma = 0.05$  and several values of parameters  $b$ ,  $d$  and  $c$ . A total of 1000 repetitions were used in each simulation. In Table 6 we give simulated normalized variances of estimated parameters that are most important for discrimination. We see that in all the cases we investigated, the variances using the maximin optimal design  $\xi_{mm}$  are smaller than the variances obtained from using the design  $\xi_u$  and in many cases by a huge margin. This simple illustration shows the benefits of incorporating optimal design ideas in the design of a toxicology study. The toxicologists' design is not theory based and so can result in poorer estimates. This means that in addition to extra labor, material and time cost, more animals are unnecessarily sacrificed without a corresponding gain in precision for the estimates when compared to the optimal design. Additional simulation results not shown here also confirm that the asymptotic variances are close to the simulated variances.

Current work is underway to implement computer algorithms to generate these optimal designs on a web site housed at <http://www.optimal-design.org/> so that practitioners can

Table 6: *Simulated normalized variances of some parameters in models (2.4)-(2.6) for several true values of parameters (left three columns).*

			maximin design $\xi_{mm}$				design $\xi_u$			
			(2.4)	(2.5)	(2.6)	(2.6)	(2.4)	(2.5)	(2.6)	(2.6)
$b$	$d$	$c$	$\text{var}(\hat{d})$	$\text{var}(\hat{c})$	$\text{var}(\hat{d})$	$\text{var}(\hat{c})$	$\text{var}(\hat{d})$	$\text{var}(\hat{c})$	$\text{var}(\hat{d})$	$\text{var}(\hat{c})$
0.10	1.0	0.0	58.85	2.02	71.07	2.48	62.73	5.53	88.42	7.81
0.10	0.8	0.0	30.38	5.39	83.93	28.91	34.93	26.72	86.34	58.80
0.10	1.0	0.2	11.86	1.96	103.40	2.79	18.43	4.83	138.97	7.77
0.10	0.8	0.2	19.39	4.21	135.00	35.91	23.57	10.88	148.29	81.56
0.06	1.0	0.0	61.67	3.91	103.78	7.17	61.62	11.35	115.70	23.33
0.08	1.0	0.1	22.58	2.47	92.44	3.68	36.83	6.48	113.39	10.80

freely find tailor-made optimal designs for their specific problems. This web site is funded by the National Institutes of General Medical Sciences and already has an array of algorithms for generating a variety of optimal designs for a broad range of models. We hope that this site will help inform practitioners of design issues and enable them to incorporate optimal design ideas in their work.

## 6 Appendix: Proofs of Lemma 2.1 and Theorem 2.1

### 6.1 Proof of Lemma 2.1

Let  $I(t, a, b, d) = f(t, a, b, d)f^T(t, a, b, d)$ , where  $f(t, a, b, d)$  is given in (2.9). Lemma 2.1 follows from the identities

$$\det \int_0^T I(t, a, b, d)d\xi(t) = \gamma \det \int_0^{T^d} I(t^d, 1, b, 1)d\xi(t) = \gamma \det \int_0^T I(t, 1, b, 1)d\xi(t^{1/d})$$

and

$$\det \int_0^T I(t, a, rb, 1)d\xi(t) = \gamma' \det \int_0^T I(rt, 1, b, 1)d\xi(t) = \gamma' \det \int_0^T I(t, 1, b, 1)d\xi(t/r)$$

where  $\gamma$  and  $\gamma'$  denote appropriate constants. □

## 6.2 Proof of Theorem 2.1.

Let  $g(t) = p^T f(t)$  be an arbitrary linear combination of the functions  $e^{-bt}$ ,  $-te^{-bt}$ ,  $-t \ln(t)e^{-bt}$ . It is easy to see that the function  $(g(t)e^{bt})'' = c/t$  does not have any roots in the interval  $[0, T]$ . Consequently, the function  $g(t)$  has at most two roots, which proves the Chebyshev property for system of functions  $e^{-bt}$ ,  $-te^{-bt}$ ,  $-t \ln(t)e^{-bt}$ . Moreover, this argument also shows that there exist precisely three Chebyshev points.

The proof of the remaining part now follows by a standard argument from classical optimal design theory. Using arguments similar to above, one shows the functions  $e^{-bt}$ ,  $-te^{-bt}$  form a Chebyshev system on the interval  $[0, T]$ . Consequently,

$$\begin{vmatrix} e^{-bt_1} & e^{-bt_2} & 0 \\ -t_1 e^{-bt_1} & -t_2 e^{-bt_2} & 0 \\ -t_1 \ln(t_1) e^{-bt_1} & -t_2 \ln(t_2) e^{-bt_2} & 1 \end{vmatrix} = \begin{vmatrix} e^{-bt_1} & e^{-bt_2} \\ -t_1 e^{-bt_1} & -t_2 e^{-bt_2} \end{vmatrix} \neq 0$$

for all  $0 \leq t_1 < t_2 \leq T$ , and we obtain from Theorem 7.7 (Chap X) in Karlin and Studden (1966) that the local  $e_3$ -optimal design is supported at the Chebyshev points. The assertion on the weights of the local  $e_3$ -optimal design follows from Pukelsheim and Torsney (1991). □

**Acknowledgements.** The authors are grateful to Martina Stein who typed parts of this paper with considerable technical expertise. This work has been supported in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823) of the German Research Foundation (DFG) and in part by a NIH grant award IR01GM072876 and the BMBF-grant SKAVOE. The work of Weng Kee Wong was partially by NIH grant awards R01GM072876, R24MD001762, P01 CA109091 and P30 CA16042-33.

## References

- P. J. Catalano, D. O. Scharfstein, L. M. Ryan, C. A. Kimmel and G. L. Kimmel (1993). Statistical model for fetal death, fetal weight, and malformation in developmental toxicity studies. *Teratology* **47**, 281-290.
- H. Chernoff (1953). Locally optimal designs for estimating parameters. *Ann. Math. Statist.* **24**, 586-602.
- H. Dette (1995). Optimal designs for identifying the degree of a polynomial regression. *Ann. Statist.* **23**, 1248-1266.
- H. Dette, V.B. Melas & W.K. Wong (2005). Optimal design for goodness-of-fit of the Michaelis-Menten enzyme kinetic function. *Journal of the American Statistician Association*, **100**, 1370-1381.
- D. W. Gaylor & J. J. Chen (1993). Dose-response models for developmental malformations. *Teratology* **47**, 291-297.
- R.J. Jennrich (1969). Asymptotic properties of nonlinear least squares estimators. *Ann. Math. Statist.* **40**, 633-643.
- S. Karlin & W. Studden (1966). *Tchebysheff Systems: with Application in Analysis and Statistics*, New York: Wiley.
- J.C. Kiefer (1974). General equivalence theory for optimum designs (approximate theory). *Ann. Statist.* **2**, 849-879.
- M. Moerbeek, A. H. Piersma & W. Slob. (2004). A comparison of three methods for calculating confidence intervals for the benchmark dose. *Risk Analysis* **24**, 31-40.
- C.H. Müller (1995). Maximin efficient designs for estimating non-linear aspects in linear models. *J. Statist. Plann. Inference* **44**, No.1, 117-132.
- C.H. Müller & A. Pázman (1998). Applications of necessary and sufficient conditions for maximin efficient designs. *Metrika* **48**, 1-19.
- T. Oscar (2004). Dose-response model for 13 strains of Salmonella. *Risk Analysis* **24**, 41-49.

- F. Pukelsheim (1993). *Optimal design of experiments*. Wiley, N.Y.
- F. Pukelsheim & B. Torsney (1991). Optimal weights for experimental designs on linearly independent support points. *Ann. Statist.* **19**, 1614-1625.
- S.D. Silvey (1980). *Optimal design*. Chapman and Hall, London.
- W. Slob (2002). Dose-response modeling of continuous endpoints. *Toxicological Sciences* **66**, 298-312.
- A.H. Piersma, A. Verhoef, C.G.J. Sweep, W.H. de Jong & H. van Loveren (2002). Developmental toxicity but no immunotoxicity in the rat after prenatal exposure to diethylstilbestrol. *Toxicology* **174**, 173-181.
- W. Slob & M. N. Pieters (1998). A probabilistic approach for deriving acceptable human intake limits and human health risks from toxicological studies: general framework. *Risk Analysis* **18**, 787-798.
- R.A. Woutersen, D. Jonker, H. Stevenson, J.D. te Biesebeek & W. Slob (2001). The benchmark approach applied to a 28-day toxicity study with Rhodorsil Silane in rats: the impact of increasing the number of dose groups. *Food and Chemical Toxicology* **39**, 697-707.