Journal of Statistical Planning and Inference 144 (2014) 188–197

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Adaptive optimal designs for dose-finding studies based on sigmoid E_{max} models

Tianhua Wang [a], Min Yang [b,*]

[a] Department of Statistics, University of Missouri, USA
[b] Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, USA

ARTICLE INFO

Available online 16 September 2013

Keywords:
Dose-finding
Adaptive designs
Loewner order
Optimal designs

ABSTRACT

The adaptive design strategy is a feasible approach for optimal designs in dose-finding studies, where the information matrices usually depend on the unknown parameters. Focusing on three pharmacodynamics sigmoid E_{max} models, we derive the corresponding simple formats of the adaptive optimal designs regardless of the optimality criteria or parameters of interest. An algorithm for deriving a specific adaptive optimal design is developed. A simulation study comparing the adaptive optimal designs and the uniform designs is also performed.

1. Introduction

Dose-finding studies are an indispensable part of the drug discovery processes. Poor estimation of drug effective levels may have a direct impact on drug development. The accuracy of estimation depends on how the data is collected, i.e., the design of experiments. An optimal/efficient design can save time, money, and energy, so it is important to develop the optimal/efficient designs for a variety of possible situations.

Although there exists a vast literature on optimal designs for linear models, less is known about optimal designs for nonlinear models, especially some nonlinear modeled dose-finding studies. Perhaps the main reason is the nonlinearity. The dose-response models are nonlinear in general, which makes the optimal designs rather complicated. While linear models are all of the form $E(y) = X\beta$, there is no simple canonical form for nonlinear models and the mathematics tends to become more difficult for nonlinear models. In addition, optimal designs depend on the parameters of interest as well as optimality criteria. Each different combination may need different approaches of deriving optimal designs. On the other hand, the Fisher information matrices for nonlinear models depend on the unknown parameters. Thus, the challenge in designing an experiment for such a model is that one is looking for the best design with the aim of estimating the unknown parameters, and yet one has to know the parameters to find the best design. One way to solve this problem is the "locally" optimal design, which is based on the "best guess" of the unknown parameters (Chernoff, 1953). Hereafter, the word "locally" is omitted for simplicity.

How can we get the reliable "best guess"? The practical way is adaptive designs. An initial experiment is used to gather some information about the unknown parameters. The preliminary estimates can then serve as the "best guess" on which the next stage design can be based on. In the second stage, the question then becomes how to add more design points in an optimal fashion, such that the combination of the existing design and the new added design is optimal/efficient according to

* Corresponding author. Tel.: +1 1312 9968612.
E-mail addresses: tianhuawang2009@gmail.com (T. Wang), minyang.stat@gmail.com, myang2@uic.edu (M. Yang).

0378-3758/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jspi.2013.09.003

some optimality criteria. The observations from the initial and second stage designs are used to obtain new estimation of parameters. If a third stage design is needed, the new estimates obtained from previous two stages' designs can serve as the "best guess". The process can be carried on to the next stage in the similar fashion.

Adaptive optimal designs have been proved to be efficient. In a simulation study, Maloney et al. (2007) showed that adaptive designs perform extremely well both when prior information is accurate and inaccurate. Zamuner et al. (2010) showed that an adaptive optimal design can provide an efficient experimental design for receptor occupancy studies using positron emission tomography. While adaptive optimal designs continue to gain popularity in biopharmaceutical applications, there is lack of theoretical and systematic work in this field. Dragalin et al. (2007) studied three sigmoid E_{max} models. Focusing on D-optimality, they proposed a procedure of deriving adaptive designs. Their procedure selected the dose levels from the set of all available dose levels which determined in the early phase of drug development. The dose levels are selected by optimizing the sensitivity function. Leonov and Miller (2009) studied a slightly simplified E_{max} model. Once new dose levels are selected, the next dose level is to optimize a new sensitivity function, which is adjusted by adding the previous selected dose levels to the existing design. Also, there are Minimax and Bayesian approaches, which were well developed in the design literature through years. The Minimax approach constructs optimal designs with respect to the minimize the partial maximum variance criterion introduced by Elfving (1959), it has been discussed by many authors (see e.g. Murty, 1971; Torsney and López-Fidalgo, 1995; Krafft and Schaefer, 1995 or Dette and Studden, 1994). The Bayesian design approaches specify probability-measure on the parameters and select any design maximizing the expected value of some function of the information matrix. Chaloner and Verdinelli (1995) gave an review paper on Bayesian experimental design, the vast literature on Bayesian design approaches has been introduced in that review paper. Like the sensitivity function approaches, Minimax and Bayesian design approaches may not be the best for a finite number of steps.

As we mentioned before, optimal designs depend on the function of parameters that we are interested in and the chosen optimality criterion. There are many combinations of parameters of interest and optimality criteria. Although we may be able to derive a strategy for a specific combination, this may not be feasible for the practical use of adaptive optimal designs. In this paper, focusing on three different sigmoid E_{max} models studied in Dragalin et al. (2007), we are targeting on a general strategy of deriving adaptive optimal designs regardless of parameters of interest or optimality criteria. Specifically, we derive simple formats of adaptive optimal designs regardless of the optimality criteria or parameters of interest. In addition, we propose an algorithm of deriving a specific optimal design. Utilizing this algorithm, we perform a simulation study, which compares the adaptive optimal designs with uniform designs. It demonstrates that the adaptive optimal designs are highly efficient.

For the lay-out of the remainder of the paper, three pharmacodynamics sigmoid E_{max} models as well as their corresponding information matrices are introduced in Section 2. The formats of the adaptive optimal designs are presented in Section 3. In Section 4, the algorithm is introduced followed by the simulation study. Section 5 is a short discussion.

2. The models and information matrices

The sigmoid E_{max} model in pharmacodynamics is to characterize the concentration-response curve (see e.g., Holford and Sheiner, 1981; MacDougall, 2006). This is a logistic model with the mean of the response at a given dose x is

$$\eta(x,\theta)=\theta_1+(\theta_2-\theta_1)\frac{x^{\theta_4}}{x^{\theta_4}+\theta_3^{\theta_4}} \tag{2.1}$$

Here, θ_1 is the minimum mean response; θ_2 is the maximum mean response; θ_3 is the ED_{50}, it is the dose at which the 50% of the maximum mean effect is attained; θ_4 is the slope parameter, which describes the steepness of the curve. Here $\theta_2>\theta_1$ and $\theta_3>0$. If the θ_4 is larger, the curve is steeper.

We study the same three models as those in Dragalin et al. (2007)

Model 1. $Y\sim N(\eta(x,\theta),\sigma^2)$ with σ^2 constant.

Model 2. $Y\sim N(\eta(x,\theta),\sigma^2(x,\theta))$ with $\sigma^2(x,\theta)=\lambda^2[\eta(x,\theta)]^2$. The coefficient of variation is constant and equals to λ.

Model 3. Y has a Gamma distribution with the shape parameter α and scale parameter β, where $\alpha=1/\lambda^2$ and $\beta=\lambda^2\eta(x,\theta)$. With this parametrization, $E(Y)=\eta(x,\theta)$, $V(Y)=\lambda^2\eta^2(x,\theta)$, and the coefficient of variation is constant and equals to λ.

The corresponding Fisher information matrices have been derived by Dragalin et al. (2007). We reproduce the information matrices here using a slightly different format. An exact design can be presented as $\{(x_i,n_i),i=1,\ldots,k\}$, where n_i is the number of subjects assigned to design point x_i. With n denoting the total number of subjects, we have that $\sum_i n_i=n$. Since finding an exact optimal design is often a difficult problem, the corresponding approximate design, in which n_i/n is replaced by ω_i, is considered. Thus a design can be denoted by $\xi=\{(x_i,\omega_i),i=1,\ldots,k\}$, where $\omega_i>0$ and $\sum_i\omega_i=1$. By standard methods, the information matrix for θ under each of the three models has the same format: $I_\xi(\theta)=\sum_i^k\omega_i I(x_i,\theta)$, where $I(x_i,\theta)$ is the information matrix at the single point x_i. The expression of $I(x,\theta)$ depends on the model we choose.

We first compute the expression of $\partial\eta(x,\theta)/\partial\theta$. Let $c=(x/\theta_3)^{\theta_4}$ then

$$\frac{\partial\eta(x,\theta)}{\partial\theta}=\left[\frac{1}{1+c},\frac{c}{1+c},\frac{\theta_4(\theta_1-\theta_2)c}{\theta_3(1+c)^2},-\frac{(\theta_1-\theta_2)c\ln(c)}{\theta_4(1+c)^2}\right]^T. \tag{2.2}$$

For Model 1, when variance, σ^2, is known (not a design parameter), $I(x,\theta)$ can be written as

$$I(x,\theta)=\frac{1}{\sigma^2}\left[\frac{\partial\eta(x,\theta)}{\partial\theta}\frac{\partial\eta(x,\theta)}{\partial\theta^T}\right]. \tag{2.3}$$

When σ^2 is unknown, $I(x,\theta)$ can be written as

$$I(x,\theta,\sigma^2)=\frac{1}{\sigma^2}\begin{pmatrix}\frac{\partial\eta(x,\theta)}{\partial\theta}\frac{\partial\eta(x,\theta)}{\partial}\theta^T & 0\\ 0 & \frac{1}{2\sigma^2}\end{pmatrix}. \tag{2.4}$$

For Model 2, when the coefficient of variation, λ, is known (not a design parameter), $I(x,\theta)$ can be written as

$$I(x,\theta)=\frac{2\lambda^2+1}{\lambda^2}\frac{1}{\eta^2(x,\theta)}\left[\frac{\partial\eta(x,\theta)}{\partial\theta}\frac{\partial\eta(x,\theta)}{\partial\theta^T}\right]. \tag{2.5}$$

When λ is unknown, $I(x,\theta)$ can be written as

$$I(x,\theta,\lambda)=\begin{pmatrix}\frac{2\lambda^2+1}{\lambda^2}\frac{1}{\eta^2(x,\theta)}\left\{\frac{\partial\eta(x,\theta)}{\partial\theta}\frac{\partial\eta(x,\theta)}{\partial\theta^T}\right\} & \frac{2}{\lambda\eta(x,\theta)}\frac{\partial\eta(x,\theta)}{\partial\theta}\\ \frac{2}{\lambda\eta(x,\theta)}\frac{\partial\eta(x,\theta)}{\partial\theta^T} & \frac{2}{\lambda^2}\end{pmatrix} \tag{2.6}$$

For Model 3, when the coefficient of variation, λ, is known (not a design parameter), $I(x,\theta)$ can be written as

$$I(x,\theta)=\frac{1}{\lambda^2}\frac{1}{\eta^2(x,\theta)}\left[\frac{\partial\eta(x,\theta)}{\partial\theta}\frac{\partial\eta(x,\theta)}{\partial\theta^T}\right]. \tag{2.7}$$

When$\lambda$is unknown, the Fisher information matrix for a single point x is

$$I(x,\theta,\lambda)=\begin{pmatrix}\frac{1}{\lambda^2}\frac{1}{\eta^2(x,\theta)}\left\{\frac{\partial\eta(x,\theta)}{\partial\theta}\frac{\partial\eta(x,\theta)}{\partial\theta^T}\right\} & 0\\ 0 & -\frac{4}{\lambda^2}\left(\frac{1}{\lambda^2}-\frac{1}{\lambda^4}\psi\left(\frac{1}{\lambda^2}\right)\right)\end{pmatrix}. \tag{2.8}$$

where ψ is the trigamma function.

3. Structure of optimal designs

As we discuss earlier, optimal designs depend on the optimality criteria and the parameters of interest. It is unlikely that we could find a design which is best among all criteria and any parameters of interest. However, it is well-known that, for two designs ξ_1 and ξ_2, if the information matrix of ξ_1 ($I(\xi_1)$) dominates that of ξ_2, i.e., $I(\xi_1)\geq I(\xi_2)$ under the Loewner ordering, then ξ_1 is not inferior to ξ_2, regardless of the choices of criteria or parameters of interest. If we can identify a subclass of designs with a simple format, so that for any given design ξ outside of the subclass, there exists a design ξ^* in that subclass with $I_{\xi^*}\geq I_\xi$, then we can focus on the designs in that subclass. This will greatly simplify the searching of optimal designs both analytically and numerically.

This approach can be directly applied to adaptive experiments. In an adaptive design, the next stage design depends on the estimation of parameters from the previous stages, which are random variables. Thus the new added design points are also random variables. They are conditionally independent (given specific values of parameters estimation) rather than independent. It is in general very complicated, perhaps even impossible, to derive the explicit form of the true information matrix. Silvey (1980) suggested to use the working information matrix computed by ignoring the dependence instead of the true information matrix. Such strategy is commonly used in the adaptive design literature. If ξ_1 denotes the design used so far and ξ_2 is the design to be used at the next stage, then the total information matrix is $I_{\xi_1}+I_{\xi_2}$. Since the first matrix is fixed, if for any given design ξ_2, there exists a design ξ_2^* in that subclass with $I_{\xi_2^*}\geq I_{\xi_2}$, then we have $I_{\xi_1}+I_{\xi_2^*}\geq I_{\xi_1}+I_{\xi_2}$. Therefore, we can again focus on the designs in that subclass. We can implement this strategy by considering the structure of the information matrix.

For a nonlinear regression model, suppose the information matrix of a design $\xi=\{(\omega_i,x_i),\ i=1,\ldots,n\}$ can be written as

$$I_\xi(\theta)=P(\theta)\left(\sum_{i=1}^{n}\omega_i C(\theta,c_i)\right)P(\theta)^T. \tag{3.1}$$

Here $P(\theta)$ is a matrix that depends on θ only and

$$C(\theta,c_i)=\begin{pmatrix}\Psi_{11}(c_i) & \Psi_{12}(c_i) & \dots & \Psi_{1p}(c_i)\\ \Psi_{12}(c_i) & \Psi_{22}(c_i) & \dots & \Psi_{2p}(c_i)\\ \vdots & \vdots & \ddots & \vdots\\ \Psi_{1p}(c_i) & \Psi_{2p}(c_i) & \dots & \Psi_{pp}(c_i)\end{pmatrix}, \tag{3.2}$$

where c_i is a function x_i and θ and $c_i\in[A,B]$. Rename all linear independent Ψ_{lt}, $1\le l\le t\le p$ to $\Psi_1,\dots,\Psi_k$ such that (i) Ψ_k is one of Ψ_{ll}, $1\le l\le p$ and (ii) there is no $\Psi_{lt}=\Psi_k$ for $l<t$. If there exists a design $\tilde{\xi}=\{(\tilde{\omega}_j,\tilde{x}_j),\ j=1,\dots,\tilde{n}\}$, such that

$$\sum_i^n \omega_i\Psi_l(c_i)=\sum_j^{\tilde{n}}\tilde{\omega}_j\Psi_l(\tilde{c}_j),\quad l=1,\dots,k-1 \tag{3.3}$$

and

$$\sum_i^n \omega_i\Psi_k(c_i)<\sum_j^{\tilde{n}}\tilde{\omega}_j\Psi_k(\tilde{c}_j), \tag{3.4}$$

then clearly $I_{\tilde{\xi}}(\theta)\le I_{\tilde{\xi}}(\theta)$.

For any design ξ, how can we find $\tilde{\xi}$ with a simple format? Yang (2010) gives a sufficient condition for deriving such a subclass of designs. Assume that (i) $\Psi_1,\dots,\Psi_k$ are infinitely differentiable; (ii) $f_{l,l}$ has no zero value on [A,B]. Here, $f_{l,t}$, $1\le t\le k$; $t\le l\le k$ are defined as follows:

$$f_{l,t}(c)=\begin{cases}\Psi_l'(c), & t=1,\ \ l=1,\dots,k\\ \left(\dfrac{f_{l,t-1}(c)}{f_{t-1,t-1}(c)}\right)', & 2\le t\le k,\ \ t\le l\le k.\end{cases} \tag{3.5}$$

The computations of $f_{l,t}$ can be viewed as the following lower triangular matrix.

$$\begin{pmatrix} f_{1,1}=\Psi_1' & & & & \\ f_{2,1}=\Psi_2' & f_{2,2}=\left(\dfrac{f_{2,1}}{f_{1,1}}\right)' & & & \\ f_{3,1}=\Psi_3' & f_{3,2}=\left(\dfrac{f_{3,1}}{f_{1,1}}\right)' & f_{3,3}=\left(\dfrac{f_{3,2}}{f_{2,2}}\right)' & & \\ f_{4,1}=\Psi_4' & f_{4,2}=\left(\dfrac{f_{4,1}}{f_{1,1}}\right)' & f_{4,3}=\left(\dfrac{f_{4,2}}{f_{2,2}}\right)' & f_{4,4}=\left(\dfrac{f_{4,3}}{f_{3,3}}\right)' & \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

The $(t+1)$th column is obtained from the tth column. The lth $(l\ge t+1)$ element of the $(t+1)$th column is the derivative of the ratio between the lth and the tth element of the tth column.

Let $F(c)=\prod_{l=1}^k f_{l,l}(c)$, $c\in[A,B]$. Yang (2010) showed that for any given design ξ, there exists a design $\tilde{\xi}$, such that $I_\xi\le I_{\tilde{\xi}}$. Here, the number of support points of $\tilde{\xi}$ depends on different situations. We refer readers to Theorem 2 of Yang (2010) for details.

With this approach, we are ready to present our first result.

Theorem 1. *Under Model*1, *i.e., normal linkage models with constant variance* σ^2, *in the continuous design space* [*A*, *B*], *for an arbitrary design* ξ, *there exists a design* $\tilde{\xi}$ *with at most four support points such that* $I_{\tilde{\xi}}(\theta)\le I_{\tilde{\xi}}(\theta)$ *regardless of* σ^2 *being known or unknown.*

Proof. Because of the structures of information matrices (2.3) and (2.4), it is clear that if the conclusion holds when variance is known, then the same conclusion also holds when variance is unknown. So we only consider the case when variance is known. By some routine algebra, the Fisher information matrix, (2.3), can be written as (3.1) with

$$P(\theta)=\frac{1}{\sigma}\begin{pmatrix}0 & 1 & 0 & 0\\ 1 & -1 & 0 & 0\\ 0 & \dfrac{\theta_4(\theta_1-\theta_2)}{\theta_3} & -\dfrac{\theta_4(\theta_1-\theta_2)}{\theta_3} & 0\\ 0 & 0 & 0 & -\dfrac{\theta_1-\theta_2}{\theta_4}\end{pmatrix} \tag{3.6}$$

and

$$C(\theta, c_i) = \begin{pmatrix} 1 & \frac{1}{1+c_i} & \frac{1}{(1+c_i)^2} & \frac{c_i \log(c_i)}{(1+c_i)^2} \\ \frac{1}{1+c_i} & \frac{1}{(1+c_i)^2} & \frac{1}{(1+c_i)^3} & \frac{c_i \log(c_i)}{(1+c_i)^3} \\ \frac{1}{(1+c_i)^2} & \frac{1}{(1+c_i)^3} & \frac{1}{(1+c_i)^4} & \frac{c_i \log(c_i)}{(1+c_i)^4} \\ \frac{c_i \log(c_i)}{(1+c_i)^2} & \frac{c_i \log(c_i)}{(1+c_i)^3} & \frac{c_i \log(c_i)}{(1+c_i)^4} & \frac{c_i^2 \log^2(c_i)}{(1+c_i)^4} \end{pmatrix}. \tag{3.7}$$

Here, $c_i = (x_i/\theta_3)^{\theta_4}$. Let $\Psi_1(c) = 1/(1+c)^4$, $\Psi_2(c) = 1/(1+c)^3$, $\Psi_3(c) = c\log(c)/(1+c)^4$, $\Psi_4(c) = 1/(1+c)^2$, $\Psi_5(c) = c\log(c)/(1+c)^3$, $\Psi_6(c) = 1/1+c$, $\Psi_7(c) = c\log(c)/(1+c)^2$, and $\Psi_8(c) = c^2\log^2(c)/(1+c)^4$. We can verify that the corresponding $f_{1,1} = -4/(1+c)^5$, $f_{2,2} = \frac{3}{4}$, $f_{3,3} = 3c+1/3c^2$, $f_{4,4} = 4c(3c+2)/(3c+1)^2$, $f_{5,5} = (9c^3+15c^2+7c+1)/c^2(3c+2)^2$, $f_{6,6} = 9c(3c+2)/(9c^2+6c+1)$, $f_{7,7} = (3c+1)/3c^2$, and $f_{8,8} = 2/3c^2$. Notice that $c > 0$, which implies that $F(c) < 0$. Thus the conclusion follows by Theorem 2 of Yang (2010). □

Although the information matrices of Model 2, (2.5) and (2.6), are similar to their counterparts of Model 1, the additional term, $1/\eta^2(x,\theta)$, makes it difficult to study the properties of f_{ll} and $F(c)$ due to their cumbersome expression. Dette and Melas (2011) generalized Yang (2010)'s result and their result could address this issue. They showed that if one of the following conditions holds

$$\{1, \Psi_1, \ldots, \Psi_{k-1}\} \quad \text{and} \quad \{1, \Psi_1, \ldots, \Psi_{k-1}, \Psi_k\} \quad \text{form Chebyshev systems,} \tag{3.8}$$

or

$$\{1, \Psi_1, \ldots, \Psi_{k-1}\} \text{ and } \{1, \Psi_1, \ldots, \Psi_{k-1}, -\Psi_k\} \text{ form Chebyshev systems,} \tag{3.9}$$

then for any arbitrary design ξ, there exists $\tilde{\xi}$, such that (3.3) and (3.4) hold. Consequently, $I_\xi \le I_{\tilde{\xi}}$. The format of $\tilde{\xi}$ depends on the situation and has the same format as that in the preceding approach. In fact, $F(c) > 0$ for all $c \in [A,B]$ is a sufficient condition that (3.8) holds while $F(c) < 0$ for all $c \in [A,B]$ is a sufficient condition that (3.9) holds.

Following Karlin and Studden (1966), $\{1, \Psi_1, \ldots, \Psi_{k-1}, \Psi_k\}$ is a Chebyshev system if

$$\begin{vmatrix} 1 & 1 & \ldots & 1 \\ \Psi_1(z_0) & \Psi_1(z_1) & \ldots & \Psi_1(z_k) \\ \vdots & \vdots & \ddots & \vdots \\ \Psi_k(z_0) & \Psi_k(z_1) & \ldots & \Psi_k(z_k) \end{vmatrix} \tag{3.10}$$

is strictly positive whenever $A \le z_0 < z_1 < \cdots < z_k \le B$.

Clearly, the sign of (3.10) remains fixed if we multiply each function of $\{1, \Psi_1, \ldots, \Psi_{k-1}, \Psi_k\}$ by a positive function, say G. For some problems, it is more convenient to study the Chebyshev system property of $\{G, G\Psi_1, \ldots, G\Psi_{k-1}, G\Psi_k\}$ instead of that of $\{1, \Psi_1, \ldots, \Psi_{k-1}, \Psi_k\}$.

Theorem 2. *Under Model 2, i.e., normal linkage models with constant coefficient of variation λ, in the continuous design space [A,B], for an arbitrary design ξ, there exists a design $\tilde{\xi}$ with at most four support points such that $I_{\tilde{\xi}}(\theta) \le I_{\xi}(\theta)$ regardless λ being known or unknown.*

Proof. We first consider the case when λ is known. By some routine algebra, the Fisher information matrix, (2.5), can be written as (3.1) with

$$P(\theta) = \sqrt{\frac{2\lambda^2+1}{\lambda^2}} \begin{pmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{\theta_2} & -\frac{\theta_1}{\theta_2} & 0 & 0 \\ 0 & \frac{\theta_4(\theta_1-\theta_2)}{\theta_3} & -\frac{\theta_4(\theta_1-\theta_2)}{\theta_3} & 0 \\ 0 & 0 & 0 & -\frac{\theta_1-\theta_2}{\theta_4} \end{pmatrix} \tag{3.11}$$

and

$$C(\theta,c_i)=\begin{pmatrix} 1 & \frac{1}{\theta_1+\theta_2c_i} & \frac{1}{(1+c_i)(\theta_1+\theta_2c_i)} & \frac{c_i\log(c_i)}{(1+c_i)(\theta_1+\theta_2c_i)} \\ \frac{1}{\theta_1+\theta_2c_i} & \frac{1}{(\theta_1+\theta_2c_i)^2} & \frac{1}{(1+c_i)(\theta_1+\theta_2c_i)^2} & \frac{c_i\log(c_i)}{(1+c_i)(\theta_1+\theta_2c_i)^2} \\ \frac{1}{(1+c_i)(\theta_1+\theta_2c_i)} & \frac{1}{(1+c_i)(\theta_1+\theta_2c_i)^2} & \frac{1}{(1+c_i)^2(\theta_1+\theta_2c_i)^2} & \frac{c_i\log(c_i)}{(1+c_i)^2(\theta_1+\theta_2c_i)^2} \\ \frac{c_i\log(c_i)}{(1+c_i)(\theta_1+\theta_2c_i)} & \frac{c_i\log(c_i)}{(1+c_i)(\theta_1+\theta_2c_i)^2} & \frac{c_i\log(c_i)}{(1+c_i)^2(\theta_1+\theta_2c_i)^2} & \frac{c_i^2\log^2(c_i)}{(1+c_i)^2(\theta_1+\theta_2c_i)^2} \end{pmatrix}. \tag{3.12}$$

Here, $c_i=(x_i/\theta_3)^{\theta_4}$. It can be shown that $1/(1+c)(\theta_1+\theta_2c)$ is a linear combination of $1/(\theta_1+\theta_2c)^2$ and $1/(1+c)(\theta_1+\theta_2c)^2$. Let $\Psi_1(c)=1/(1+c)^2(\theta_1+\theta_2c)^2$, $\Psi_2(c)=1/(1+c)(\theta_1+\theta_2c)^2$, $\Psi_3(c)=c\log(c)/(1+c)^2(\theta_1+\theta_2c)^2$, $\Psi_4(c)=1/(\theta_1+\theta_2c)^2$, $\Psi_5(c)=c\log(c)/(1+c)(\theta_1+\theta_2c)^2$, $\Psi_6(c)=1/(\theta_1+\theta_2c)$, $\Psi_7(c)=c\log(c)/(1+c)(\theta_1+\theta_2c)$, and $\Psi_8(c)=c^2\log^2(c)/(1+c)^2(\theta_1+\theta_2c)^2$. Direct computation shows that the corresponding f_{ll} is cumbersome and difficult to handle.

Let $G=(1+c)^2(\theta_1+\theta_2c)^2$. As we discuss earlier, it is sufficient to study the Chebyshev system property of the following system

$$\{(1+c)^2(\theta_1+\theta_2c)^2,1,1+c,c\log(c),(1+c)^2,c(1+c)\log(c),(1+c)^2(\theta_1+\theta_2c),c(1+c)(\theta_1+\theta_2c)\log(c),c^2\log^2(c)\}. \tag{3.13}$$

By using row reduction of a matrix (i.e. multiplying a row by a constant, switching two rows or adding constant times a row to another row) and the fact both θ_1 and θ_2 are positive, the Chebyshev system property of (3.13) is equivalent to that of the following system

$$\{1,c,c\log(c),c^2,c^2\log(c),c^3,c^3\log(c),-c^4,c^2\log^2(c)\}. \tag{3.14}$$

Now we can compute the corresponding f_{ll} for the new $\tilde{\Psi}_1(c)=c$, $\tilde{\Psi}_2(c)=c\log(c)$, $\tilde{\Psi}_3(c)=c^2$, $\tilde{\Psi}_4(c)=c^2\log(c)$, $\tilde{\Psi}_5(c)=c^3$, $\tilde{\Psi}_6(c)=c^3\log(c)$, $\tilde{\Psi}_7(c)=-c^4$, and $\tilde{\Psi}_8(c)=c^2\log^2(c)$. We can verify that the corresponding $f_{1,1}=1$, $f_{2,2}=1/c$, $f_{3,3}=2$, $f_{4,4}=1/x$, $f_{5,5}=6$, $f_{6,6}=1/c$, $f_{7,7}=-12$, and $f_{8,8}=1/18c^3$. Since $c>0$, we have $F(c)<0$. This implies that (3.9) holds. Our conclusion follows by Theorem 3.1 of Dette and Melas (2011).

Now, we consider the case when λ is unknown. For an arbitrary design $\xi=\{(\omega_i,x_i),\ i=1,\ldots,n\}$, there exists a design $\tilde{\xi}=\{(\tilde{\omega}_j,\tilde{x}_j),\ j=1,\ldots,4\}$ such that $I_{\tilde{\xi}}(\theta)\le I_{\xi}(\theta)$ for information matrix (2.5). From (2.6), the information matrix when λ is unknown, it is sufficient to show that

$$\sum_{i=1}^{n}\omega_i\frac{1}{\eta(x_i,\theta)}\frac{\partial\eta(x_i,\theta)}{\partial\theta^T}=\sum_{j=1}^{4}\tilde{\omega}_j\frac{1}{\eta(\tilde{x}_j,\theta)}\frac{\partial\eta(\tilde{x}_j,\theta)}{\partial\theta^T}. \tag{3.15}$$

Direct computation shows that

$$\frac{1}{\eta(x,\theta)}\frac{\partial\eta(x,\theta)}{\partial\theta}=\left[\frac{1}{\theta_1+\theta_2c},\frac{c}{\theta_1+\theta_2c},\frac{\theta_4(\theta_1-\theta_2)c}{\theta_3(1+c)(\theta_1+\theta_2c)},-\frac{(\theta_1-\theta_2)c\ln(c)}{\theta_4(1+c)(\theta_1+\theta_2c)}\right]^T \tag{3.16}$$

with the same transformation on $c=(x/\theta_3)^{\theta_4}$. The first component of (3.16) is $\Psi_6(c)$, the second component is linear transformation of $\Psi_6(c)$, the third component is a linear combination of $\Psi_2(c)$ and $\Psi_4(c)$, and the last component is proportional to $\Psi_7(c)$. Notice that (3.3) holds for $\Psi_1(c),\ldots,\Psi_7(c)$. This implies that (3.15) holds. Thus, the conclusion follows. □

Although Model 3 involves a distribution other than normal distribution, its information matrices, (2.7) and (2.8), are similar to the information matrix (2.5) of Model 2 when λ is known. A design determines the information matrices through the expression $1/\eta^2(x,\theta)[(\partial\eta(x,\theta)/\partial\theta)(\partial\eta(x,\theta)/\partial\theta^T)]$, the same way as that of Model 2 when λ is known. Thus, the conclusion of Theorem 2 can be immediately extended to Model 3.

Theorem 3. *Under Model*3, *i.e., Gamma linkage models with constant coefficient of variation* λ, *in the continuous design space* [*A*,*B*], *for an arbitrary design* ξ, *there exists a design* $\tilde{\xi}$ *with at most four support points such that* $I_{\xi}(\theta)\le I_{\tilde{\xi}}(\theta)$ *regardless* λ *being known or unknown.*

4. Simulation studies

With Theorems 1, 2, and 3 we can focus on a subclass of designs with a simple format for any optimal design. But we still need to determine the four design points and their corresponding weights. How to derive a specific optimal design? With the complicated information structure, it could be rather difficult to derive an analytical solution, if not impossible. Although numerical solution is in general feasible, with total seven variables to be determined, the computation is rather complicated.

Since deriving a numerical optimal design is reduced to maximization of a function of a few arguments, some standard methods from Matlab or R usually have some limitations. They often stuck at a local minimum/maximization and failed to find a global minimum/maximum solution, especially when the optimized functions are complicated. The local solutions are

not usually optimal designs. Given that our log-likelihood functions in our examples are complicated functions with 7 arguments (four design points and three weights), it might be rather difficult to find the true optimal designs with standard methods from Matlab or Stufken and Yang (2012) provided an algorithm of deriving optimal designs. Although their approach is for one-stage design, their approach can be extended for adaptive designs. The main idea of their approach is using adaptive grid search, which starts with a coarse grid that is made increasing finer in later stages. At each stage we identify the best design based on the grid at that stage. For the next stage, a finer grid is restricted to neighborhoods of the best support points found at the current stage. The search continues until a specified accuracy for the design points is reached. For a given design support, the corresponding optimal weights are determined through solving a few nonlinear equations. We refer readers to Theorems 5 and 6 of Stufken and Yang (2012) for details.

Although this approach cannot guarantee the resulting design is an optimal design, the following general equivalence theorem for adaptive optimal designs can be used to verify the optimality. The proof is standard and therefore omitted (c.f., Pukelsheim, 2006). Let ξ_0 be the design we have used so far, n_0 be the associated sample size, ξ_1 be the next stage design, n_1 be the associated sample size, $F(\theta)$ be a vector of parameter functions of interest. The covariance matrix of the maximum likelihood estimator of $F(\theta)$, Σ_F, can be written as

$$\Sigma_F = \frac{\partial F(\theta)}{\partial \theta^T}(n_0 I_{\xi_0} + n_1 I_{\xi_1})^{-1}\frac{\partial F(\theta)}{\partial \theta}. \tag{4.1}$$

Here, we assume that the matrix $n_0 I_{\xi_0} + n_1 I_{\xi_1}$ is non-singular. Such assumption is true in general since our initial design ξ_0 is to estimate all parameters. We consider two optimality criteria here: *A*- and *D*-optimality. A similar conclusion can be generalized to more general optimality criteria, such as Φ_p-optimality (Biedermann et al., 2006). Under *D*-optimality, we are looking for ξ, such that $|\Sigma_F^{-1}|$ is maximized. This optimality criterion results in minimizing the generalized variance of the parameter estimates. Under *A*-optimality, we are looking for ξ, such that $\mathrm{Tr}(\Sigma_F)$ is minimized. This criterion results in minimizing the average variance of the estimates.

Theorem 4. *Let ξ_0 be the design we have used so far, ξ_1 is the next stage design, $F(\theta)$ be a vector of parameter functions of interest, and Σ_F be the covariance matrix of $F(\theta)$.*

Then a design ξ_1^* is D-optimal for $F(\theta)$ if and only if, for all design points x,

$$\begin{aligned}&\mathrm{Tr}\left(\Sigma_F^{-1}\frac{\partial F(\theta)}{\partial \theta^T}(n_0 I_{\xi_0} + n_1 I_{\xi_1^*})^{-1} I(x,\theta)(n_0 I_{\xi_0} + n_1 I_{\xi_1^*})^{-1}\frac{\partial F(\theta)}{\partial \theta}\right)\\ &\quad \le \mathrm{Tr}\left(\Sigma_F^{-1}\frac{\partial F(\theta)}{\partial \theta^T}(n_0 I_{\xi_0} + n_1 I_{\xi_1^*})^{-1} I_{\xi_1^*}(n_0 I_{\xi_0} + n_1 I_{\xi_1^*})^{-1}\frac{\partial F(\theta)}{\partial \theta}\right),\end{aligned} \tag{4.2}$$

where equality holds when x is any support point of a D-optimal design for $F(\theta)$.

Similarly, a design ξ_1^* is A-optimal for $F(\theta)$ if and only if, for all design points x,

$$\begin{aligned}&\mathrm{Tr}\left(\frac{\partial F(\theta)}{\partial \theta^T}(n_0 I_{\xi_0} + n_1 I_{\xi_1^*})^{-1} I(x,\theta)(n_0 I_{\xi_0} + n_1 I_{\xi_1^*})^{-1}\frac{\partial F(\theta)}{\partial \theta}\right)\\ &\quad \le \mathrm{Tr}\left(\frac{\partial F(\theta)}{\partial \theta^T}(n_0 I_{\xi_0} + n_1 I_{\xi_1^*})^{-1} I_{\xi_1^*}(n_0 I_{\xi_0} + n_1 I_{\xi_1^*})^{-1}\frac{\partial F(\theta)}{\partial \theta}\right),\end{aligned} \tag{4.3}$$

where equality holds when x is any support point of an A-optimal design for $F(\theta)$.

Next, we will study some numerical examples about Models 1, 2, and 3. For adaptive optimal designs, we need an initial design. Since we do not have much information about the parameters, it is better to choose some robust design as the initial design. Here, we adopt the same set up as that of Dragalin et al. (2007). The four parameters are $\theta = (3, 15, \theta_3, 4)$. There are four possible values of θ_3: 200, 300, 400, and 500. The dose range is from 0 to 1000, the finest grid unit is 0.1 when searching adaptive optimal designs using the algorithms from Stufken and Yang (2012). The initial design is an uniform design, where the first 120 subjects are equally allocated to the doses $\{0, 200, 400, 600, 800, 1000\}$. The design question is how to allocate the next 240 subjects after the first stage design.

The set up of simulation studies is as follows: we first generated data based on the uniform design. Then the estimation of parameters was analyzed using NLMINB function of R. An adaptive optimal design was derived based on the parameter estimation. We considered four types of optimal designs: (1) *D*-optimal designs for all parameters; (2) *A*-optimal designs for all parameters; (3) *D*-optimal designs for part of the parameters; and (4) *A*-optimal designs for part of the parameters. The corresponding efficiencies, which are defined below, are evaluated with respect to the locally optimal design based on the true parameter values.

Let ξ^* be the locally optimal design based on the true parameter values; ξ^M be the adaptive optimal design based on the estimated parameter values; ξ_0 be the initial uniform design (all designs are in the form of approximate designs). Suppose $F(\theta)$ is the vector of parameter functions of interest. The covariance matrix of the MLE of $F(\theta)$ under ξ_0 and ξ^M can be written as

$$\Sigma_F^M = \frac{\partial F(\theta)}{\partial \theta^T}(120 I_{\xi_0} + 240 I_{\xi^M})^{-1}\frac{\partial F(\theta)}{\partial \theta}. \tag{4.4}$$

The covariance matrix of the MLE of $F(\theta)$ under ξ^* can be written as

$$\Sigma_F^* = \frac{\partial F(\theta)}{\partial \theta^T}(360 I_{\xi^*})^{-1}\frac{\partial F(\theta)}{\partial \theta}. \tag{4.5}$$

The D-efficiency is defined as

$$eff_{\xi^M}^D = \left(\frac{|\Sigma_F^*|}{|\Sigma_F^M|}\right)^{1/s}, \tag{4.6}$$

where s is the dimension of matrix Σ_F^M. The A-efficiency is defined as

$$eff_{\xi^M}^A = \frac{\mathrm{Tr}(\Sigma_F^*)}{\mathrm{Tr}(\Sigma_F^M)}. \tag{4.7}$$

Notice that both (4.6) and (4.7) are evaluated under the true parameter values. The D- and A-efficiencies of the uniform design, denoted by $eff_{\xi^U}^D$ and $eff_{\xi^U}^A$, respectively, are computed in the similar fashion. Here, the uniform design means that $(120+240)$ subjects are equally allocated to the doses $\{0, 200, 400, 600, 800, 1000\}$. The relative efficiency between the adaptive optimal design and the uniform design is also computed.

In the following six tables, we present the simulation results for some selected parameters for Models 1, 2 and 3. The simulation was repeated 1000 times for each combination of the model and parameter values. For each estimated parameter values, the four types of optimal designs were derived. For each derived adaptive optimal design, its optimality has been verified through the modified general equivalence theorem (Theorem 4). The mean values of efficiencies are reported. The corresponding standard deviations are also reported in brackets.

Example 1. Model 1 with known variance. Tables 1 and 2 summarize the simulation results for $(\theta_1, \theta_2, \theta_3, \theta_4)$ and $(\theta_1, \theta_2, \theta_3)$, respectively. Notice that some parameter values are much larger than the others, thus the corresponding variances likely dominate those of the others. Since A-optimal design is to minimize the summation of the variances of the corresponding MLE, A-optimal design could be mainly influenced by those parameters with larger values. To avoid this, we consider the summation of the "normalized" variances, i.e., A-optimal design here is to minimize $\sum \mathrm{Var}(\hat{\theta}_i)/\hat{\theta}_i$. The same procedure is applied to A-optimal designs in other examples.

From Tables 1 and 2, we can see that D-efficiencies of the adaptive optimal designs are good in general, range from 0.87 to 0.92 for $(\theta_1, \theta_2, \theta_3, \theta_4)$, and range from 0.81 to 0.86 for $(\theta_1, \theta_2, \theta_3)$. On the other hand, the A-efficiencies are relatively small, especially for Table 2, where they range from 0.56 to 0.67. However, the adaptive optimal designs perform significantly better than the uniform designs. For one case (A-optimality in Table 2 when $\theta_3 = 300$), the relative efficiency is 1.91, which means we can expect that adaptive optimal designs save 48% sample size compared to the uniform designs. The smallest relative efficiency is 1.06 (D-optimality in Table 1 when $\theta_3 = 500$). In fact, the adaptive design has high efficiency in this case (0.92). The relative efficiency is small due to the relatively high efficiency of uniform design (0.86).

Example 2. Model 2 with unknown coefficient of variance $\lambda = 0.33$.

Tables 3 and 4 show that D-efficiencies of the adaptive optimal designs are good for most cases (above 0.84) except for the two cases when $\theta_3 = 500$, which are 0.75 and 0.69, respectively. A-efficiencies are again relatively small, ranging from 0.65 to 0.78. An interesting observation is that A-efficiencies are almost identical for both tables with very small difference. Just like Example 1, the adaptive optimal designs perform better than the uniform designs. For one case (A-optimality in Table 4 when $\theta_3 = 200$), the relative efficiency of the adaptive optimal design and the uniform design is 2.91, which means we can expect that the adaptive optimal design saves 64% of the sample size compared to the uniform design. Interestingly, the adaptive optimal design is not highly efficient (0.65) at this case. But the uniform design is also very inefficient (0.22). The smallest relative efficiency is 1.13 (D-optimality in Table 3 when $\theta_3 = 400$). This is due to the fact that uniform design already has 84% efficiency, which means not much gain can be expected.

Example 3. Model 3 with known coefficient of variance $\lambda = 0.33$.

Table 1
D- and A-efficiencies for $(\theta_1, \theta_2, \theta_3, \theta_4)$.

θ_3	200	300	400	500
$eff_{\xi^M}^D$	0.87(0.10)	0.92(0.05)	0.90(0.05)	0.92(0.04)
$eff_{\xi^U}^D$	0.58	0.80	0.78	0.86
$eff_{\xi^M}^D / eff_{\xi^U}^D$	1.51(0.17)	1.15(0.06)	1.16(0.07)	1.06(0.05)
$eff_{\xi^M}^A$	0.77(0.20)	0.82(0.14)	0.75(0.16)	0.75(0.13)
$eff_{\xi^U}^A$	0.45	0.47	0.64	0.60
$eff_{\xi^M}^A / eff_{\xi^U}^A$	1.72(0.46)	1.73(0.30)	1.18(0.26)	1.25(0.22)

Table 2
D- and *A*-efficiencies for $(\theta_1, \theta_2, \theta_3)$.

θ_3	200	300	400	500
$eff^D_{\xi^M}$	0.81(0.09)	0.84(0.07)	0.84(0.06)	0.86(0.06)
$eff^D_{\xi^U}$	0.64	0.62	0.68	0.77
$eff^D_{\xi^M}/eff^D_{\xi^U}$	1.27	1.34(0.11)	1.24(0.09)	1.11(0.08)
$eff^A_{\xi^M}$	0.56(0.23)	0.67(0.18)	0.65(0.20)	0.65(0.18)
$eff^A_{\xi^U}$	0.53	0.35	0.56	0.52
$eff^A_{\xi^M}/eff^A_{\xi^U}$	1.07(0.43)	1.91(0.52)	1.17(0.35)	1.26(0.35)

Table 3
D- and *A*-efficiencies for $(\theta_1, \theta_2, \theta_3, \theta_4, \lambda)$.

θ_3	200	300	400	500
$eff^D_{\xi^M}$	0.87(0.13)	0.95(0.05)	0.95(0.03)	0.75(0.15)
$eff^D_{\xi^U}$	0.47	0.81	0.84	0.45
$eff^D_{\xi^M}/eff^D_{\xi^U}$	1.86(0.29)	1.17(0.06)	1.13(0.04)	1.66(0.33)
$eff^A_{\xi^M}$	0.65(0.28)	0.82(0.13)	0.81(0.12)	0.78(0.11)
$eff^A_{\xi^U}$	0.22	0.36	0.56	0.45
$eff^A_{\xi^M}/eff^A_{\xi^U}$	2.90(1.26)	2.25(0.37)	1.46(0.22)	1.76(0.24)

Table 4
D- and *A*-efficiencies for $(\theta_1, \theta_2, \theta_3, \theta_4)$.

θ_3	200	300	400	500
$eff^D_{\xi^M}$	0.84(0.16)	0.94(0.05)	0.94(0.04)	0.69(0.13)
$eff^D_{\xi^U}$	0.39	0.77	0.81	0.38
$eff^D_{\xi^M}/eff^D_{\xi^U}$	2.18(0.40)	1.22(0.07)	1.16(0.05)	1.81(0.34)
$eff^A_{\xi^M}$	0.65(0.28)	0.82(0.13)	0.81(0.13)	0.78(0.11)
$eff^A_{\xi^U}$	0.22	0.36	0.55	0.45
$eff^A_{\xi^M}/eff^A_{\xi^U}$	2.91(1.27)	2.25(0.37)	1.46(0.22)	1.76(0.24)

Tables 5 and 6 show that *D*-efficiencies of the adaptive optimal designs are just moderate, range from 0.67 to 0.76 for $(\theta_1, \theta_2, \theta_3, \theta_4)$, and range from 0.60 to 0.71 for $(\theta_1, \theta_2, \theta_3)$. *A*-efficiencies of the adaptive designs are even worse, range from 0.32 to 0.47. Surprisingly, unlike the previous two examples, the adaptive optimal designs do not always perform better than uniform designs. In fact, among 16 cases, the adaptive optimal designs perform better in 5 cases, perform equivalently in 2 cases, and perform worse in 9 cases. We found that the variances of the estimation in this example are much larger than the counterparts of Examples 1 and 2. The poor estimation maybe the reason why the adaptive designs are not so efficient. Another interesting observation is that, when $\theta_1 = 200$, the relative efficiency is 1.9 under *A*-optimality for $(\theta_1, \theta_2, \theta_3, \theta_4)$. Surprisingly, with the same set up except that the parameters of interest are $(\theta_1, \theta_2, \theta_3)$, the relative efficiency becomes 0.59. After we checked the corresponding variance–covariance matrices, we found that the variance of θ_4 under the uniform design is significantly larger than others. This causes the *A*-efficiency of the uniform design in the first case to be significantly lower than that of the second case.

5. Discussion

Adaptive optimal design approach is a practical remedy of "locally" optimal designs for nonlinear models. Focusing on three different Sigmoid E_{max} models, we derive the simple formats for adaptive optimal designs regardless of the optimality criteria or parameters of interest. The simple formats are helpful for us to derive some specific adaptive optimal designs numerically. A modified general equivalence theorem is developed to verify the optimality.

Simulation studies show that the adaptive optimal designs perform better than the corresponding uniform designs in general, but not always. The performance of the adaptive optimal designs depends on the quality of the initial estimation of parameters, which relies on the initial design and the sample size. How to select a good initial design? Without good knowledge of unknown parameters, either Bayesian optimal design or minimax design can be a good choice. Such optimal problems are in general

Table 5
D- and *A*-efficiencies for $(\theta_1, \theta_2, \theta_3, \theta_4)$.

θ_3	200	300	400	500
$eff^D_{\xi^M}$	0.67(0.17)	0.75(0.12)	0.76(0.12)	0.75(0.11)
$eff^D_{\xi^U}$	0.39	0.77	0.81	0.79
$eff^D_{\xi^M}/eff^D_{\xi^U}$	1.78(0.45)	0.97(0.16)	0.94(0.14)	0.95(0.14)
$eff^A_{\xi^M}$	0.42(0.22)	0.45(0.24)	0.47(0.19)	0.45(0.18)
$eff^A_{\xi^U}$	0.22	0.36	0.55	0.44
$eff^A_{\xi^M}/eff^A_{\xi^U}$	1.90(1.00)	1.25(0.67)	0.86(0.34)	1.00(0.41)

Table 6
D- and *A*-efficiencies for $(\theta_1, \theta_2, \theta_3)$.

θ_3	200	300	400	500
$eff^D_{\xi^M}$	0.60(0.12)	0.66(0.15)	0.70(0.11)	0.71(0.09)
$eff^D_{\xi^U}$	0.64	0.61	0.71	0.73
$eff^D_{\xi^M}/eff^D_{\xi^U}$	0.95(0.18)	1.07(0.24)	0.98(0.15)	0.97(0.13)
$eff^A_{\xi^M}$	0.32(0.16)	0.36(0.23)	0.41(0.18)	0.41(0.18)
$eff^A_{\xi^U}$	0.54	0.29	0.48	0.41
$eff^A_{\xi^M}/eff^A_{\xi^U}$	0.59(0.29)	1.23(0.78)	0.85(0.38)	1.00(0.44)

challenging and worthy of further investigation. On the other hand, although allocating more samples in the initial stage is helpful for better quality of initial parameter estimation, this also means we have less samples to allocate in the remaining stages, which may not be the best strategy for the efficiency of the whole experiment. A general strategy for allocating sample sizes in adaptive designs is not yet available. Some comprehensive simulation studies may be helpful.

Acknowledgments

We received numerous constructive and helpful comments from the Editors, two referees, and Professor Ryan Martin. These comments and suggestions substantially changed and improved the paper. We thank them. This research was supported by NSF grants DMS-0707013 and DMS-1322797, and FDA MCM Challenge Grant. The contents are solely the responsibility of the authors and do not necessarily represent the official views of FDA.

References

Biedermann, S., Dette, H., Zhu, W., 2006. Geometric construction of optimal designs for dose-response models with two parameters. Journal of the American Statistical Association 101, 747–759.
Chaloner, K., Verdinelli, I., 1995. Bayesian experimental design: a review. Statistical Science, 273–304.
Chernoff, H., 1953. Locally optimal designs for estimating parameters. Annals of Mathematical Statistics 24, 586–602.
Dette, H., Melas, V.B., 2011. A note on the de la Garza phenomenon for locally optimal designs. The Annals of Statistics 39, 1266–1281.
Dette, H., Studden, W.J., 1994. Optimal designs with respect to Elfving's partial minimax criterion in polynomial regression. Annals of the Institute of Statistical Mathematics 46, 389–403.
Dragalin, V., Hsuan, F., Padmanabhan, S.K., 2007. Adaptive designs for dose-finding studies based on sigmoid $E_{\max}$ model. Journal of Biopharmaceutical Statistics 17, 1051–1070.
Elfving, G., 1959. Design of linear experiments. In: Grenander, Ulf (Ed.), Probability and Statistics: The Harald Cramr Volume, Almqvist & Wiksell, Stockholm, John Wiley & Sons, New York, pp. 58–74.
Holford, N., Sheiner, L., 1981. Understanding the dose-effect relationship: clinical application of pharmacodynamics models. Clinical Pharmacokinetics 6, 429–453.
Karlin, S., Studden, W.J., 1966. Optimal experimental designs. Annals of Mathematical Statistics 37, 783–815.
Krafft, O., Schaefer, M., 1995. Exact Elfving-minimax designs for quadratic regression. Statistica Sinica 5, 475–484.
Leonov, S., Miller, S., 2009. An adaptive optimal design for the $E_{\max}$ model and its application in clinical trials. Journal of Biopharmaceutical Statistics 19, 360–385.
MacDougall, J., 2006. Analysis of dose-response studies – $E_{\max}$. In: Ting, N. (Ed.), Dose Finding in Drug Development. Springer, New York.
Maloney, A., Karlsson, M.O., Simonsson, U.S.H., 2007. Optimal adaptive design in clinical drug development: a simulation example. Journal of Clinical Pharmacology 66, 319–320.
Murty, V.N., 1971. Minimax designs. Journal of the American Statistical Association 47, 1231–1243.
Pukelsheim, F., 2006. Optimal Design of Experiments. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.
Silvey, S.D., 1980. Optimal Design: An Introduction to the Theory for Parameter Estimation. Chapman & Hall, London.
Stufken, J., Yang, M., 2012. On locally optimal designs for generalized linear models with group effects. Statistica Sinica 22, 1765–1786.
Torsney, B., López-Fidalgo, J., 1995. MV-optimization in simple linear regression. In: Kitsos, C.P., Müller, W.G. (Eds.), MODA4-Advances in Model-Oriented Data Analysis, pp. 71–76.
Yang, M., 2010. On the de la Garza phenomenon. The Annals of Statistics 38, 2499–2524.
Zamuner, S., Di Iorio, V., Nyberg, J., Gunn, R., Cunningham, V., Gomeni, R., Hooker, A., 2010. Adaptive-optimal design in PET occupancy studies. Clinical Pharmacology & Therapeutics 87, 563–571.