Statistics
in Medicine

Special Issue Paper

Received 11 July 2011, Accepted 3 July 2012 Published online 3 August 2012 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.5555

On the efficiency of two-stage response-adaptive designs

Holger Dette,[a*†] Björn Bornkamp[b] and Frank Bretz[b]

In this paper, we investigate the efficiency of response-adaptive locally optimum designs. We focus on two-stage adaptive designs, where after the first stage the accrued data are used to determine a locally optimum design for the second stage. On the basis of an explicit expansion of the information matrix, we compare the variance of the maximum likelihood estimates obtained from a two-stage adaptive design and a fixed design without adaptation. For several one-parameter models, we provide explicit expressions for the relative efficiency of these two designs, which is seen to depend sensitively on the statistical problem under investigation. In particular, we show that in non-linear regression models with moderate or large variances the first-stage sample size of an adaptive design should be chosen sufficiently large in order to address variability in the interim parameter estimates. These findings support the results of recent simulation studies conducted to compare adaptive designs in more complex situations. We finally present an application to a real clinical dose-finding trial aiming at the estimation of the smallest dose achieving a certain percentage of the maximum treatment effect by using a three-parameter $E_{\max}$ model.

Keywords: dose estimation; dose finding; mean squared error; optimal design; maximum likelihood estimation; non-linear regression

1. Introduction

It is well known that good study designs can substantially improve the efficiency of statistical analyses, and numerous authors have worked on the problem of constructing optimal designs for regression models. Optimal designs for non-linear regression models usually depend on the unknown model parameter, leading to so-called locally optimal designs [1]; see also [2–5] among many others. Locally optimal designs require a specification of the unknown model parameter at the planning stage of a study and might thus be sensitive with respect to an initial misspecification of that parameter. More advanced design strategies have been developed instead to overcome this sensitivity, such as Bayesian or other robust designs; see [6–9] among others.

Sequential designs are an attractive alternative that update the information about the unknown parameter sequentially after each observation; see [10–12] for early references. Several authors have proved efficiency of sequentially optimal designs in the sense that sequential designs converge asymptotically to the locally optimal designs and that the corresponding parameter estimates are asymptotically efficient; see [13–16] among others. However, these results usually refer to specific models, and fully sequential designs are often not feasible in practice because of logistic restrictions (necessity of real-time data capture, automated data analyses, highly flexible drug supply, etc.).

Response-adaptive designs with several cohorts of subjects (adaptive designs, in short) are often used instead: After each stage, the accumulated data of the ongoing study are used to update the initial guess of the underlying model parameters [17–19]. These designs continue to gain popularity in biopharmaceutical applications. For example, in clinical studies addressing dose-finding objectives, trial designs that enable adaptations based on accrued data of an ongoing trial can be more efficient than fixed designs without adaptations [20]. Several adaptive designs have been introduced in the recent past;

[a] *Ruhr-Universität Bochum, 44780 Bochum, Germany*
[b] *Novartis Pharma AG, Lichtstrasse 35, 4002 Basel, Switzerland*
* *Correspondence to: Holger Dette, Ruhr-Universität Bochum, 44780 Bochum, Germany.*
† *E-mail: holger.dette@ruhr-uni-bochum.de*

see, for example, [21–24] for approaches in the context of dose-finding clinical trials. In order to investigate the operating characteristics of a given adaptive design, in particular in comparison with a traditional fixed design, extensive simulations are typically necessary [25]. Theoretical comparisons often fall short because of the complicated structure in the data generating process of adaptive (or sequential) designs. As pointed out by [24], the main reason for the lack of theoretical results is that sequential and adaptive designs generate dependent observations. Consequently, the precision of the estimators cannot be measured by the Fisher information matrix as formally done in classical statistical theory (which requires independent observations).

In this paper, we compare the asymptotic efficiency of two-stage adaptive designs with fixed designs. To this end, we derive an explicit expression for the (asymptotic) Fisher information of these designs. These results are used for a comparison of the variances of the maximum likelihood estimates (MLE) obtained from adaptive and fixed designs. We derive explicit expressions for the relative efficiency comparing these two designs for several one-parameter models. We illustrate the methodology with several examples and demonstrate that the approximations derived by the asymptotic theory are accurate for realistic sample sizes. Moreover, we show that, in non-linear regression models with a moderate variance of the responses, the first-stage sample size of an adaptive design should be chosen sufficiently large in order to address variability in the interim parameter estimate. In particular, we demonstrate that the superiority of an adaptive or a fixed design depends sensitively on the statistical problem under investigation. These findings support the results of recent simulation studies comparing several adaptive designs in more complex situations [20, 25].

Accordingly, this paper is organized as follows. In Section 2, we present the main asymptotic results. In particular, we give an explicit expression for the relative efficiency comparing adaptive designs with fixed designs for one-parameter models. In Section 3, we illustrate these expressions for the exponential, logistic, and Poisson regression models. In Section 4, we present an application to a real clinical dose-finding trial aiming at the estimation of the smallest dose achieving a certain percentage of the maximum treatment effect using a three-parameter $E_{\max}$ model. We give concluding remarks in Section 5.

2. Main results

In this section, we present the main results for comparing the asymptotic efficiency of two-stage adaptive designs with fixed designs. In Section 2.1, we introduce the basic notation used throughout this paper. In Section 2.2, we describe the two design options and state the main asymptotic efficiency result. We give a sketch of the proof in the Appendix and refer to a Technical Report for the extensive but relatively standard asymptotic expansions. Finally, we give in Section 2.3 an explicit expression for the relative efficiency comparing adaptive designs with fixed designs for one-parameter models.

2.1. Notation

For the sake of concreteness, we describe the methods and examples in the context of clinical dose-finding studies, although the results of this paper remain valid for other applications. We consider a clinical outcome Y observed at dose level $d \in \mathcal{D}$. The variable Y may represent efficacy or safety, and the dose range is given by $\mathcal{D} = [\underline{d}, \overline{d}]$, where $\underline{d}$ and $\overline{d}$ denote the lowest and highest doses under investigation, respectively. In many cases, $\underline{d} = 0$ is the placebo dose. Assume that Y has the density $f(y, d, \theta)$, where $\theta \in \Theta$ denotes the finite dimensional unknown parameter vector. We further assume that N independent observations $Y_1, \ldots, Y_N$ are available and denote by $\hat{\theta}$ the MLE based on the full sample. Finally, let

$$\xi = \begin{pmatrix} d_1 & \ldots & d_k \\ w_1 & \ldots & w_k \end{pmatrix}$$

denote an experimental design with relative patient allocation w_i at dose $d_i, i = 1, \ldots, k$. Following [26], the weights $w_i \geqslant 0$, with $\sum_{i=1}^k w_i = 1$, are not necessarily multiples of $1/N$. In practice, for a given total sample size N, a design ξ is implemented by rounding the quantities $w_i N$ to integers, say n_i, with $\sum_{i=1}^k n_i = N$.

The statistic $\sqrt{N}\left(\hat{\theta}-\theta\right)$ is, under standard regularity assumptions, asymptotically normal distributed with mean 0 and covariance matrix $M^{-1}(\xi,\theta)$, where

$$M(\xi,\theta)=\int\int\left(\frac{\partial}{\partial\theta}\log f(y,d,\theta)\right)^{T}\left(\frac{\partial}{\partial\theta}\log f(y,d,\theta)\right)f(y,d,\theta)\mathrm{d}y\mathrm{d}\xi(d) \tag{1}$$

denotes the information matrix of the given design ξ. The matrix $M(\xi,\theta)$ can be interpreted as a precision measure of the parameter estimate $\hat{\theta}$ based on the design ξ. 'Larger' values of $M(\xi,\theta)$ indicate better (i.e., more precise) estimates of θ. A locally optimal design maximizes an appropriate functional of this matrix, the so-called optimality criterion; see [27,28] among others. Finally, let

$$I(\theta,\tau):=M(\xi_\tau,\theta) \tag{2}$$

denote the information matrix of the locally optimal design ξ_τ for the parameter $\tau\in\Theta$ if the 'true' parameter is given by θ.

2.2. *Asymptotic efficiency*

We now introduce two major design options: a fixed, that is, non-adaptive, design ξ_F, where observations are taken at pre-specified doses, and a two-stage adaptive design ξ_A, where after the first stage the accrued data are used to determine the second-stage design. In the following, let θ_0 denote a preliminary guess for the unknown parameter θ. In the context of clinical dose-finding trials, such preliminary information is often available from previous trials (animal studies, proof-of-concept studies, etc.) to generate a best guess θ_0.

(F) *Fixed design* ξ_F: Take all N observations according to the locally optimal design ξ_{θ_0} based on the best guess θ_0. The resulting estimate of θ is denoted by $\hat{\theta}_F$.

(A) *Two-stage adaptive design* ξ_A: Split the total sample N in two parts and proceed as follows.

- Take N_0 observations according to the locally optimal design ξ_{θ_0}. For the asymptotic considerations in the following text, we let $p_0=\frac{N_0}{N}$ and assume that $\lim_{N\to\infty}N_0/N\in[0,1)$ is a fixed constant.
- Estimate the parameter θ by maximum likelihood (ML) estimation from these N_0 observations, resulting in $\hat{\theta}_1$.
- Take $N_1=N-N_0$ observations according to the locally optimal design $\xi_{\hat{\theta}_1}$ and estimate the parameter θ by ML estimation from all $N=N_0+N_1$ observations. The final estimate is denoted by $\hat{\theta}_A$ and depends on the first-stage data through the random variable $\hat{\theta}_1$, that is, $\hat{\theta}_A=\hat{\theta}_A\left(\hat{\theta}_1\right)$. Finally, we define $p_1=\frac{N_1}{N}$ and note that $p_0+p_1=1$.

In the following, we provide an analytical comparison of the two design options. Note that under standard assumptions in non-linear regression the variance of the MLE is of the order $O(1/N)$, whereas the squared bias is of the order $O(1/N^2)$ [29], which implies that the mean squared error is dominated by the variance. Therefore, we approximate the mean squared error (MSE) of the MLE by its variance, that is,

$$\mathrm{MSE}\left(\hat{\theta}\right)=E\left[\left(\hat{\theta}-\theta\right)\left(\hat{\theta}-\theta\right)^{T}\right]\approx\mathrm{Var}\left(\hat{\theta}\right). \tag{3}$$

We aim at deriving asymptotic expansions for the variances $\mathrm{Var}\left(\hat{\theta}_F\right)$ and $\mathrm{Var}\left(\hat{\theta}_A\right)$ in order to compare the two design options ξ_F and ξ_A for a given statistical problem.

In general, the explicit calculation of the variances $\mathrm{Var}\left(\hat{\theta}_F\right)$ and $\mathrm{Var}\left(\hat{\theta}_A\right)$ for a given non-linear model is very cumbersome. However, an asymptotic expansion can be calculated using computer algebra systems such as MATHEMATICA or MATLAB. First note that the discussion at the beginning of this section shows that the variance of an estimator $\hat{\theta}$ from the total sample can be approximated by $\mathrm{Var}\left(\hat{\theta}\right)\approx M^{-1}(\xi,\theta)/N$, where $M(\xi,\theta)$ is the information matrix defined in (1). In Appendix A, we

provide a sketch of a proof for the following main results. The variance of the MLE from the fixed design ξ_F using the locally optimal design ξ_{θ_0} can be approximated by

$$\mathrm{Var}\left(\hat{\theta}_F\right) \approx \frac{1}{N} M(\xi_F, \theta),$$

where the information matrix of ξ_F satisfies

$$M(\xi_F, \theta) \approx I(\theta, \theta_0) + \frac{1}{\sqrt{N_0}} K(\theta, \theta_0) + \frac{1}{N_0} L(\theta, \theta_0), \tag{4}$$

and the matrices K and L depend on the specific model under consideration as well as the initial guess θ_0 for the unknown parameter θ. Furthermore, we obtain for the variance of the MLE from the adaptive design ξ_A the approximation

$$\mathrm{Var}\left(\hat{\theta}_A\right) \approx \frac{1}{N} M(\xi_A, \theta)$$

where

$$M(\xi_A, \theta) \approx H(\theta, \theta_0) + \frac{1}{\sqrt{N_0}} \bar{K}(\theta, \theta_0) + \frac{1}{N_0} \bar{L}(\theta, \theta_0), \tag{5}$$

with $H(\theta, \theta_0) = p_0 I(\theta, \theta_0) + p_1 I(\theta, \theta)$ and appropriate matrices $\bar{L}, \bar{K}$. Note that the matrix $H(\theta, \theta_0)$ is a weighted average of the information matrices corresponding to the locally optimal designs ξ_{θ_0} and ξ_θ. Therefore, this matrix can be interpreted as a mixture of information matrices corresponding to two locally optimal designs: one for the 'true' parameter θ and another one for the preliminary guess θ_0. The weights p_0 and p_1 in this mixture correspond to the relative proportions of subjects treated in the first and second stages, respectively. Consequently, for 'small' values of p_0, the dominating term in (5) becomes 'close' to the Fisher information matrix of the locally optimal design ξ_θ. Similarly, the adaptive design ξ_A is approximately given by $\xi_A \approx p_0 \xi_{\theta_0} + p_1 \xi_\theta$, and the remainder corresponds to the error in these approximations. Note that the expansion (5) refers to an asymptotic analysis where we assume the first-stage sample size N_0 to be of the same order as the total sample size $N \to \infty$, that is, $\lim_{N\to\infty} N_0/N \in (0, 1)$.

In general, the matrices $K, \bar{K}$ and $L, \bar{L}$ in (4) and (5) are neither positive nor negative definite, and therefore it is not clear whether for finite sample sizes the matrix $M(\xi_A, \theta)$ is smaller (with respect to the Loewner ordering) than $M(\xi_F, \theta)$ corresponding to the locally optimal design ξ_θ. Because $H(\theta, \theta_0) \geq I(\theta, \theta_0)$, however, it follows that asymptotically the adaptive design ξ_A is always better than the fixed design ξ_F. For finite sample sizes, we have to factor in the correction terms of order $1/\sqrt{N_0}$ and $1/N_0$, and the relationship is not obvious anymore.

The prior arguments remain valid for any differentiable optimality criterion ϕ. To be precise, assume that a (locally) ϕ-optimal design minimizes $\phi(M(\xi, \theta))$ in the class of all designs. When comparing the efficiency of the two designs ξ_F and ξ_A, this gives

$$\mathrm{eff}_\phi(\xi_F, \xi_A) = \frac{\phi(M(\xi_F, \theta))}{\phi(M(\xi_A, \theta))} \approx \frac{\phi(I(\theta, \theta_0))}{\phi(p_0 I(\theta, \theta_0) + p_1 I(\theta, \theta))} + \frac{c}{\sqrt{N_0}} + \frac{d}{N_0} \tag{6}$$

for the ϕ-efficiency of the fixed design ξ_F with respect to the adaptive design ξ_A, where no information regarding the sign of the constants c and d is available in general. If $\mathrm{eff}(\xi_F, \xi_A) < 1$, the design ξ_F is preferable as it yields smaller values of the optimality criterion. If $\mathrm{eff}(\xi_F, \xi_A) > 1$, the opposite is true, and the design ξ_A is preferable. In general, a conclusion about the superiority of a design depends on the underlying regression model; see Section 3 for examples. A common application is the problem of estimating a function of the unknown parameter θ, say $\psi(\theta)$. For example, a frequent problem in dose response studies is the estimation of relevant target doses as a function of the parameters of a regression model; see [30, 31] among others. In such situations, the asymptotic variance of the canonical estimate $\hat{\psi} = \psi\left(\hat{\theta}_A\right)$ from a two-stage design ξ_A is given by

$$\begin{aligned}\mathrm{Var}(\hat{\psi}) &\approx \nabla\psi(\theta) M^{-1}(\xi_A, \theta)(\nabla\psi(\theta))^T \\ &\approx \nabla\psi(\theta) H^{-1}(\theta, \theta_0)(\nabla\psi(\theta))^T + \frac{1}{\sqrt{N_0}} \tilde{K}(\theta, \theta_0) + \frac{1}{N_0} \tilde{L}(\theta, \theta_0)\end{aligned}$$

with appropriate constants $\tilde{K}$, $\tilde{L}$ and where ∇g denotes the gradient of a real valued function g. Note that we can extend (6) accordingly. We come back to these results when estimating the smallest dose,

achieving a certain percentage of the maximum treatment effect using a three-parameter $E_{\max}$ model in Section 4.

2.3. *Efficiencies for one-parameter models*

If the parameter θ in the non-linear regression model from Section 2.1 satisfies $\theta \in \Theta \subset \mathbb{R}$, the formulae simplify substantially, and we can explicitly give the constants in (6). In this case, the information matrix of a given design ξ is one-dimensional, and an optimal design maximizes this matrix (or minimizes its inverse). Assume that for each $\theta \in \Theta$ a one-point design, say ξ_θ, maximizes $M(\xi, \theta)$ in the class of all designs on the dose range $\mathcal{D}$. Let $d(\theta)$ denote the corresponding support point of the locally optimal design ξ_θ, which we assume to be an interior point of the dose range $\mathcal{D} \subset \mathbb{R}$. In this case, we can give an explicit expression for the asymptotic efficiency of a fixed design ξ_F compared with an adaptive design ξ_A as the ratio of the asymptotic variances $\mathrm{Var}\left(\hat{\theta}_F\right)$ and $\mathrm{Var}\left(\hat{\theta}_A\right)$ (see Appendix B for some details). Because the mean squared error is dominated by the variance, it follows from (3) that

$$\mathrm{eff}(\xi_F, \xi_A) = \frac{\mathrm{MSE}\left(\hat{\theta}_F\right)}{\mathrm{MSE}\left(\hat{\theta}_A\right)} \approx \frac{\mathrm{Var}\left(\hat{\theta}_F\right)}{\mathrm{Var}\left(\hat{\theta}_A\right)} \approx \left\{ \frac{I(\theta, \theta_0)}{H(\theta, \theta_0)} - p_1 \frac{g(\theta)(5p_0 I(\theta, \theta_0) + p_1 I(\theta, \theta))}{2N_0 H^3(\theta, \theta_0)} \right\}^{-1}. \quad (7)$$

Note again that in (7) the dominating term $I(\theta, \theta_0)/H(\theta, \theta_0) < 1$, because $H(\theta, \theta_0) = p_0 I(\theta, \theta_0) + p_1 I(\theta, \theta) \geqslant I(\theta, \theta_0)$. Therefore, for large first-stage sample sizes N_0, we have $\mathrm{eff}(\xi_F, \xi_A) > 1$ and expect the adaptive design ξ_A to be more efficient than the fixed design ξ_F. However, the second term in (7) is positive, and this contribution may be substantial for finite sample sizes as illustrated with examples in the following section.

3. Examples

In this section, we illustrate the asymptotic theory with three examples by considering an exponential, a logistic, and a Poisson regression model.

3.1. *Exponential regression model*

We consider the one-parameter exponential regression model with homoscedastic errors,

$$E[Y|d] = \eta(d, \theta) = e^{-\theta d}, \qquad \mathrm{Var}\,(Y|d) = \sigma^2 > 0, \quad (8)$$

where $\mathcal{D} = [0, \infty)$ and $\theta > 0$. In this case, we have $\frac{\partial}{\partial\theta}\eta(d, \theta) = -de^{-\theta d}$, and the Fisher information matrix at the point d is obtained from [32]. Numerous authors [33] have considered optimal design problems for this model. In particular, the performance of two-stage sequential designs for this model has also been investigated in [14]. These authors considered one strategy for the first step and two fixed design strategies for the second step of the adaptive procedure and determined the optimal allocation of the observations between the two stages. For this purpose, they introduced a prior distribution for the unknown parameter and calculated a first-order approximation for the expected value of the optimality criterion. In contrast to the work of this paper, our approach is based on maximum likelihood estimation and additionally uses higher order expansions as considered in (4). The local D-optimal design for the model (8) is a one-point design with $d(\theta) = 1/\theta$. Consequently,

$$I(\theta, \theta) = \frac{1}{\sigma^2}(e\theta)^{-2}, \quad I(\theta, \theta_0) = \frac{1}{\sigma^2}\left(e^{\theta/\theta_0}\theta_0\right)^{-2}, \quad \text{and} \quad g(\theta) = -\frac{2}{\sigma^2}\left(\theta^2 e\right)^{-2}.$$

Therefore, it follows from (7) that

$$\mathrm{eff}(\xi_F, \xi_A) \approx \left\{ \frac{1}{p_0 + (1-p_0)\left\{(\theta e)^2\left(e^{\theta/\theta_0}\theta_0\right)^{-4}\right\}^{-1}} + \frac{\sigma^2}{e^2 N \theta^4} \frac{1-p_0}{p_0} \frac{5p_0\left(e^{\theta/\theta_0}\theta_0\right)^{-4} + \frac{1-p_0}{(\theta e)^2}}{\left[p_0\left(e^{\theta/\theta_0}\theta_0\right)^{-4} + (1-p_0)\{(\theta e)^2\}^{-1}\right]^3} \right\}^{-1}. \quad (9)$$

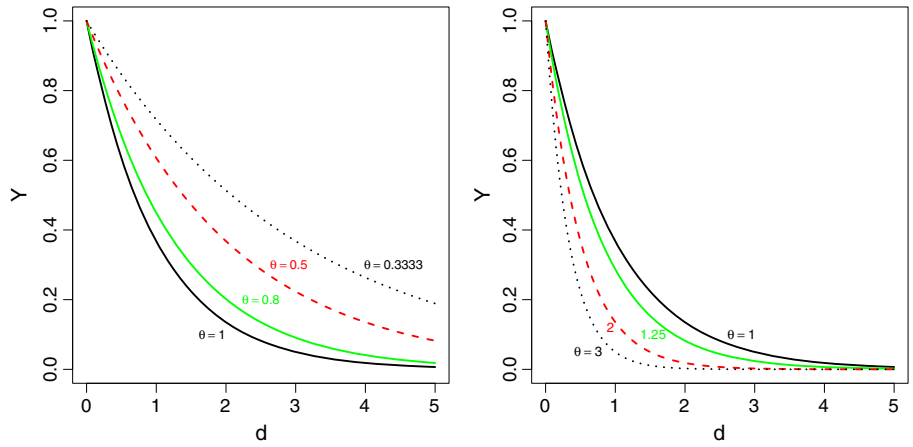

Figure 1. Plots of the exponential model (8) for different values of θ reflecting the degree of misspecification of the true parameter value $\theta = 1$.

Note that the second term in this expression is always positive. This means that—because of the adaptive structure of the design—there is a further contribution in the efficiency of order $O(1/N)$, which is negligible for large N. In the following, we investigate this approximation in several concrete finite sample scenarios. In Figure 1, we display the resulting exponential models if the parameter θ is underestimated ($\theta_0 = 0.8, 0.5, 0.3333$ in the left panel) or overestimated ($\theta_0 = 1.25, 2, 3$ in the right panel). In both panels, we display the true exponential model with $\theta = 1$ as a solid line.

In Figure 2, we plot the approximated efficiency from (9) as function of p_0 for $\sigma^2 = 1, 0.1$ and the θ_0 values displayed in Figure 1, where the total sample size is $N = 100$ and the true parameter value is $\theta = 1$. A ratio larger than 1 means that the adaptive design ξ_A yields smaller MSEs and is therefore better than the fixed design ξ_F. On the other hand, if the ratio is smaller than 1, the non-adaptive design ξ_F shows a better performance. In the top panels, we display the results for $\theta_0 < \theta$, whereas the case $\theta_0 > 0$ is shown in the bottom panel. In the left panels of each row, we show the results for a realistic signal-to-noise ratio with $\sigma^2 = 1$, whereas in the right panels, we display the results for a rather unrealistically small variance $\sigma^2 = 0.1$.

Comparing the two designs ξ_F and ξ_A reveals that for large variances the fixed design ξ_F has often a competitive or even better performance for a broad range of p_0 values. This observation can be nicely explained by the fact that the term of order $1/N_0$ in the approximation (9) is always positive and increasing with σ^2. Heuristically, a large error variance leads to a highly variable first-stage estimate $\hat{\theta}_1$ if the initial sample size N_0 is not sufficiently large. Therefore, the corresponding design $\xi_{\hat{\theta}_1}$ used in the second stage may not be efficient in some cases. On the other hand, for small variances or large first-stage sample sizes, we can estimate the parameter θ rather precisely from the data collected in the first stage. Consequently, updating the initial parameter guess θ_0 based on the first-stage data will lead to a better second-stage design and to an overall better performance for most p_0 values. Note also that the degree of initial misspecification of the parameter θ (through θ_0) has only limited impact when the variance is large. Overall, the differences between the designs ξ_F and ξ_A are small for the situations considered here, except in the case of very small variances σ^2 and where the initial guess θ_0 deviates substantially from θ.

3.2. Logistic regression

Consider a logistic regression model, where the responses are independent Bernoulli random variables with probability of success

$$p(d, \theta) = E[Y|d] = \frac{1}{1 + e^{d-\theta}} ,$$

and $\mathcal{D} = \mathbb{R}$. This model is sometimes called the one-parameter Rasch model and is used to model the item characteristic curve in item response theory [34]. Several authors [10, 16, 35] have discussed

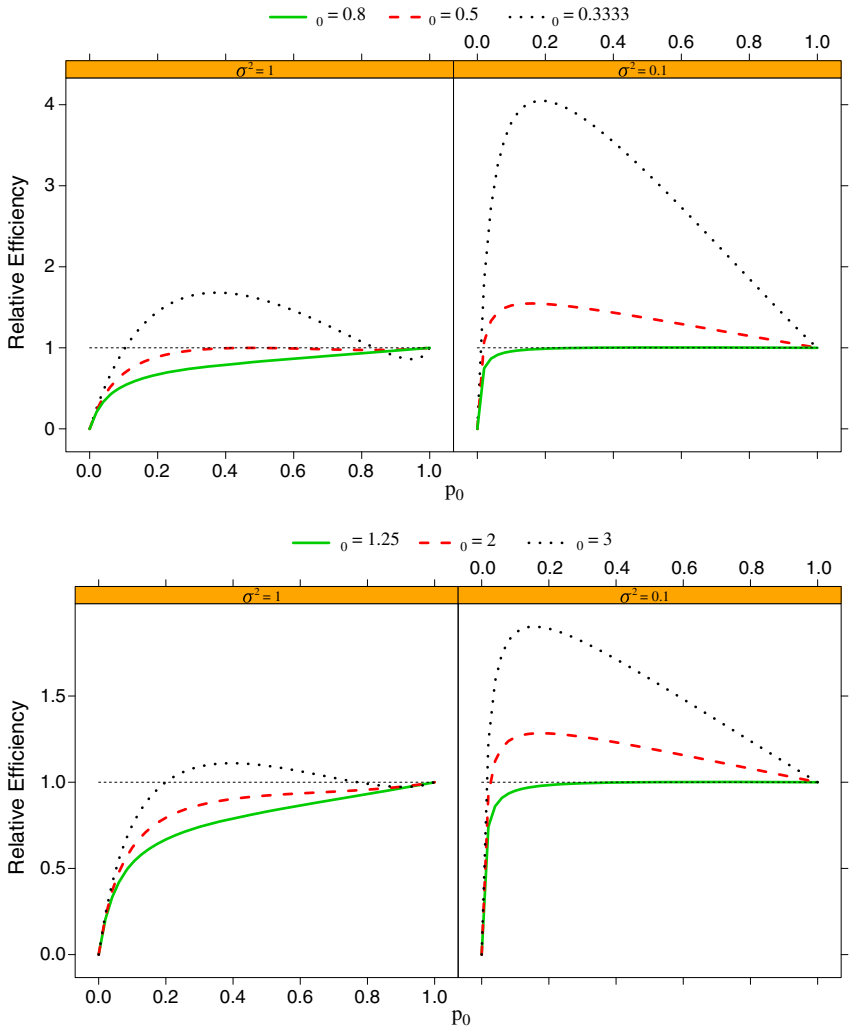

Figure 2. Plot of the approximation defined in (9) for the MSE ratio as function of p_0 under the exponential model (8) for $N = 100$, $\sigma^2 = 1$ (left panels), $\sigma^2 = 0.1$ (right panels), and different θ_0 values. Top: $\theta_0 = 0.8, 0.5, 0.3333$; bottom: $\theta_0 = 1.25, 2, 3$.

sequential optimal designs for the Rasch model. It follows from [36] that the Fisher information matrix for a one-point design δ_d at the point d is given by

$$M(\delta_d, \theta) = \frac{e^{d-\theta}}{\left(1 + e^{d-\theta}\right)^2}.$$

Standard calculation shows that the design concentrating its mass at the point $d(\theta) = \theta$ is locally optimal. Therefore, we obtain

$$I(\theta, \tau) = \frac{e^{\tau-\theta}}{\left(1 + e^{\tau-\theta}\right)^2},$$

which implies

$$I(\theta, \theta) = \frac{1}{4}, \quad g(\theta) = -\frac{1}{8}, \quad \text{and} \quad H(\theta, \theta_0) = p_0 \frac{e^{\theta_0-\theta}}{\left(1 + e^{\theta_0-\theta}\right)^2} + p_1 \frac{1}{4}.$$

Consequently, it follows from (7) that

$$\text{eff}(\xi_F, \xi_A) \approx \left\{ \left(p_0 + \frac{p_1(1+e^\gamma)^2}{4e^\gamma} \right)^{-1} + p_1 \frac{(20 p_0 e^\gamma + p_1(1+e^\gamma)^2)(1+e^\gamma)^4}{N_0(4p_0 e^\gamma + p_1(1+e^\gamma)^2)^3} \right\}^{-1}, \tag{10}$$

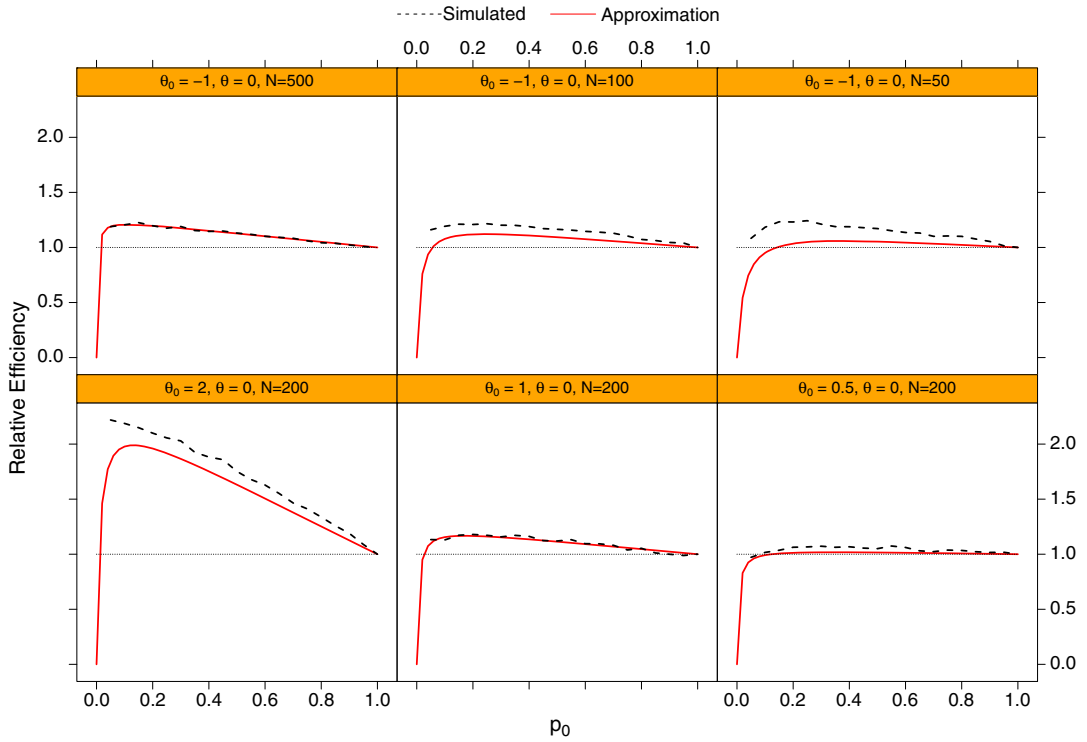

Figure 3. Plot of the approximation (10) for the MSE ratio (solid line) and corresponding simulation results (dashed line) as a function of p_0 under the logistic model for different configurations of θ, θ_0, and N.

where $\gamma = \theta_0 - \theta$ denotes the degree of initial misspecification of θ through θ_0. In Figure 3, we plot the approximation defined in (10) together with the corresponding simulation results as a function of p_0 for different values of θ_0, θ, and N. In order to investigate the accuracy of the asymptotic results for finite sample sizes, we have also performed simulations to calculate the MSE ratio for the MLEs obtained from the designs ξ_F and ξ_A (dashed lines in Figure 3; based on 20,000 simulation runs). We observe a rather precise approximation of the simulated MSE ratios by the asymptotic theory.

We observe that in most situations the adaptive design shows a better performance, although the improvement remains small, except for large values of $|\gamma|$ and N. We can explain these results by the fact that the variance in the logistic regression model is relatively small. For example, if $\theta_0 - \theta = -1$, the variance of individual observations in the first stage is $p(\theta_0, \theta)(1 - p(\theta_0, \theta)) = 0.197$. As a consequence, the parameter estimate $\hat{\theta}_1$ obtained from the first stage is rather accurate, and the corresponding design $\xi_{\hat{\theta}_1}$ is already close to the locally optimal design. If $\theta_0 - \theta = -2$, the variance of the observations from the first stage is even smaller (roughly 0.105), which explains the superiority of the adaptive design in this case.

3.3. *Poisson regression model*

In our final example, we consider the Poisson regression model

$$P(Y = k|d) = \frac{\left(e^{\theta d}\right)^k}{k!} e^{-e^{\theta d}}$$

for $d \in [0, \infty)$ and $\theta \in (-\infty, 0]$ [37]. A straightforward calculation shows that the Fisher information at the point d is $d^2 e^{\theta d}$. A locally optimal design based on the initial first guess θ_0 advises the experimenter to take all observations at the point $d(\theta_0) = -2/\theta_0$. Consequently,

$$g(\theta) = -\frac{8}{\theta^4 e^2}, \quad I(\theta, \theta_0) = \frac{4}{\theta_0^2} e^{-2\theta/\theta_0},$$

and it follows from (7) that for $\gamma = \theta/\theta_0$,

$$\text{eff}(\xi_F, \xi_A) \approx \left\{ \frac{e^\gamma}{p_0 e^\gamma + p_1 \gamma^{-2}} + \frac{p_1 e^2 (5 p_0 e^\gamma \gamma^2 + p_1)}{4\theta^4 N_0 (p_0 e^\gamma \gamma^2 + p_1)^3} \right\}^{-1}. \tag{11}$$

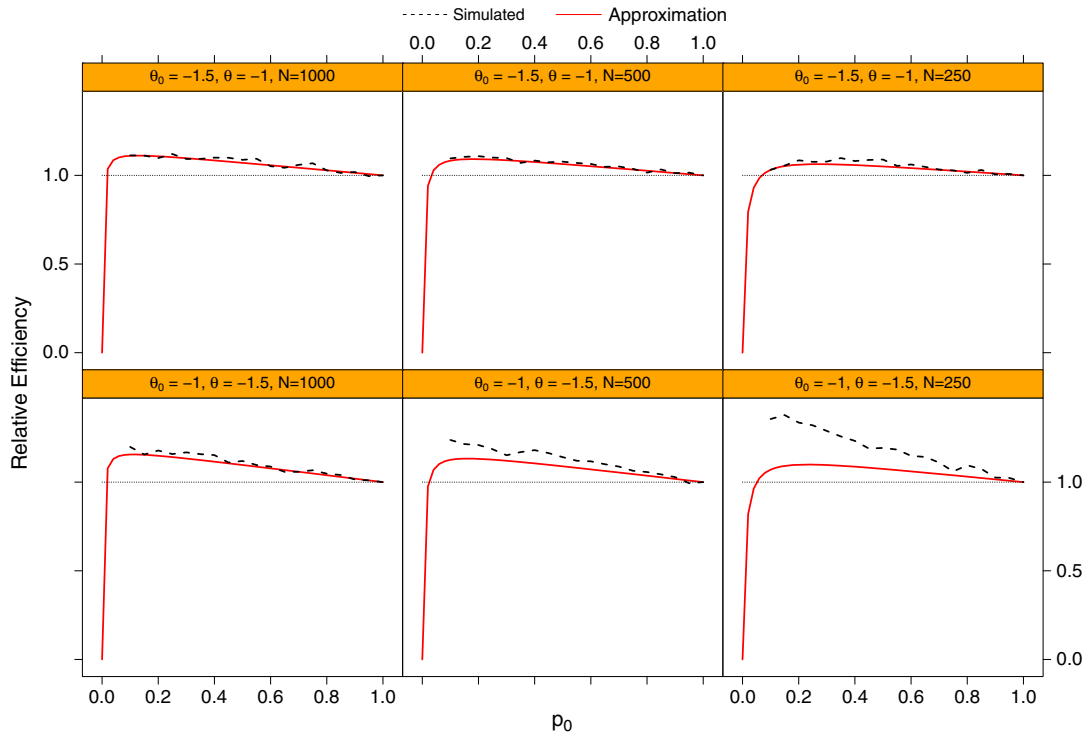

Figure 4. Plot of the approximation defined in (11) for the MSE ratio (solid line) and corresponding simulation results (dashed line) as a function of p_0 under the Poisson model for different configurations of θ, θ_0, and N.

In Figure 4, we plot the MSE ratio as a function of p_0 for two parameter specifications: $\theta_0 = -1.5$, $\theta = -1$ and $\theta_0 = -1$, $\theta = -1.5$. In both cases, we chose rather large sample sizes to avoid situations where the interim MLE $\hat{\theta}_1 = 0$ and the optimal design point for the second stage cannot be calculated. The results are qualitatively similar to those for the logistic model discussed in Section 3.2. Again, the simulated MSE ratio is approximated well by the asymptotic theory.

4. Clinical trial example to estimate the ED_p

We now present an application to the clinical dose-finding study for an anti-anxiety drug described in [38]. The primary endpoint is the change in an anxiety scale score from baseline at the end of the study. In the following, we focus on the homoscedastic $E_{\max}$ model

$$E[Y|d] = \eta(d,\theta) = \theta_0 + \theta_1 \frac{d}{\theta_2 + d}, \qquad \text{Var}\,(Y|d) = \sigma^2 > 0.$$

Here, θ_0 denotes the placebo effect, θ_1 the asymptotic maximum treatment effect achieved at an infinite dose, and θ_2 the ED_{50}, that is, the dose that gives 50% of the maximum treatment effect. The motivation to focus on the $E_{\max}$ model is its ubiquitous use in clinical practice. For example, it can be justified on the relationship of drug–receptor interactions and therefore deduced from the chemical equilibrium equation [39]. On the basis of clinical information available before the start of the study, [38] assumed that the average placebo effect is $\theta_0 = 0$, the maximum treatment effect within the dose range $\mathcal{D} = [\underline{d}, \overline{d}] = [0\text{ mg}, 150\text{ mg}]$ under investigation is 0.4 (i.e., $\theta_1 = 0.467$), and $\theta_2 = 25$. Furthermore, they assumed that all dose levels within $\mathcal{D}$ are safe.

Unlike [38], we consider the problem of estimating the smallest dose ED_p achieving $100p\%$, $0 < p < 1$, of the maximum treatment effect in the observed dose range $\mathcal{D}$. Let $h(d,\theta) = \eta(d,\theta) - \eta(\underline{d},\theta)$ denote the effect difference at $d \in (\underline{d}, \overline{d}]$ and $\underline{d}$. Following [40], we define the ED_p as

$$ED_p = \inf\left\{ d \in (\underline{d}, \overline{d}] : \frac{h(d,\theta)}{h(d_{\max},\theta)} \geqslant p \right\},$$

where $d_{\max} = \overline{d}$. We estimate ED_p by

$$\widehat{\mathrm{ED}}_p = \inf\left\{ d \in (\underline{d}, \overline{d}] : \frac{h\left(d, \hat{\theta}\right)}{h\left(\hat{d}_{\max}, \hat{\theta}\right)} \geqslant p \right\},$$

where $\hat{d}_{\max} = \mathrm{argmax}_{d \in (\underline{d}, \overline{d}]} h\left(d, \hat{\theta}\right)$ denotes the dose corresponding to the observed maximum effect difference in the interval $(\underline{d}, \overline{d}]$.

In the following, we compare the two design options from Section 2.2 with respect to their relative efficiency of estimating ED_p, focusing on $p = 0.9$. Following [40], the optimal design for estimating ED_{90} allocates one-fourth of the patients on each of $\underline{d}$ and $\overline{d}$ and the remaining one-half of the patients on the intermediate dose

$$\frac{\overline{d}(\underline{d} + \theta_2) + \underline{d}(\overline{d} + \theta_2)}{\underline{d} + \overline{d} + 2\theta_2}.$$

In total, 100 patients will be allocated for each simulated trial.

In Figure 5, we plot the ratio of the simulated MSEs as a function of p_0 for different configurations of the true parameter θ_2, the initial guess $\theta_2^{(0)}$, and σ. As the ML parameter estimate might not always exist

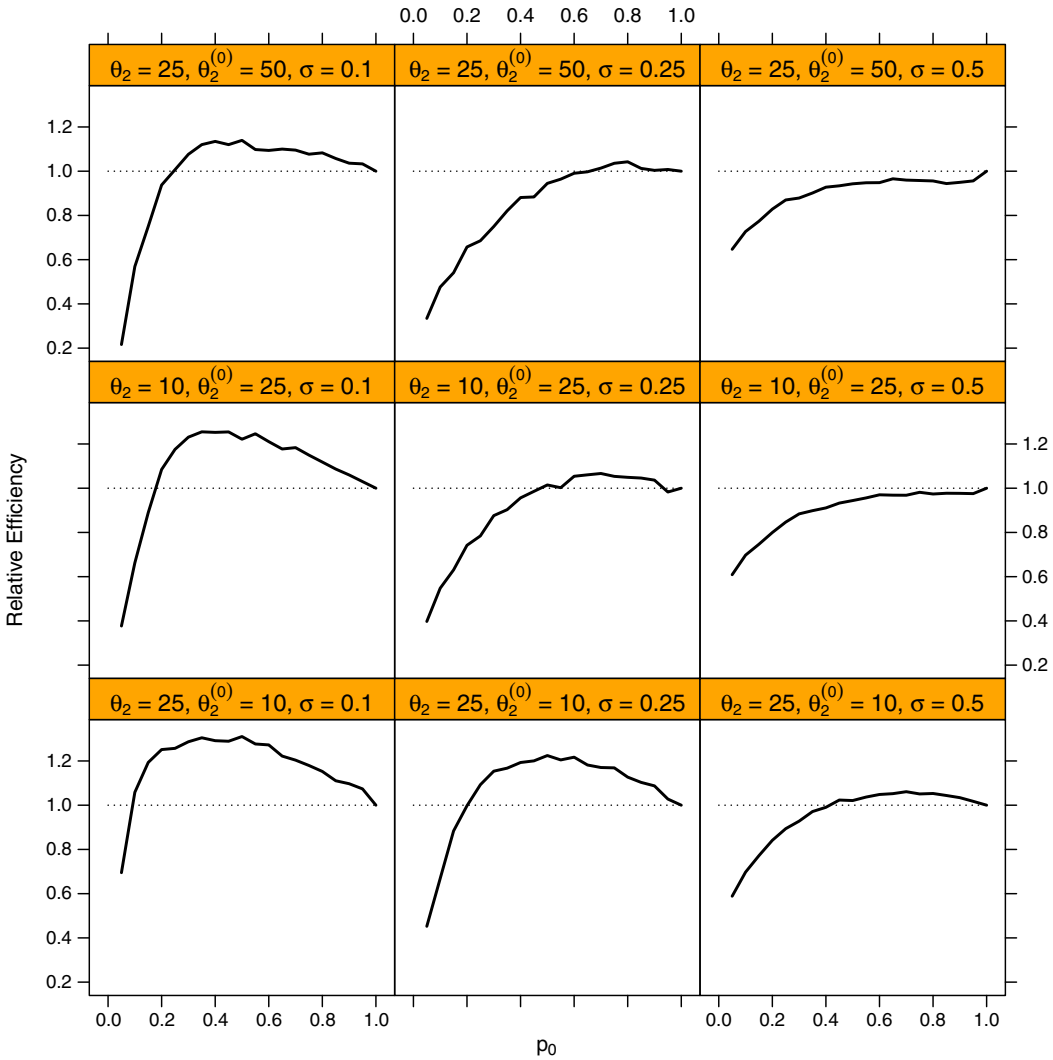

Figure 5. Relative efficiency in estimating the ED_{90} as a function of p_0 for the $E_{\max}$ model for different configurations of θ_2, $\theta_2^{(0)}$, and σ; the total sample size was 100, and the results are based on 20,000 simulations.

in the simulations, particularly at interim analyses with small p_0, we bounded the θ_2 values by 0.015 and 1500 to ensure existence of the MLE. Similar to the results in Figure 2, one can observe that the benefit of the adaptive design depends on the variability. A small variability results in a benefit for the adaptive design, whereas for larger variability the benefit of an adaptive design diminishes.

5. Conclusions

A major motivation for this work was the observation from simulation studies that the benefit of adaptive designs in terms of estimation efficiency is sometimes less in magnitude than intuitively expected and crucially depends on the underlying models and assumptions [25]. This paper provides a theoretical confirmation of these empirical results in a well-controlled situation, where we can safely ignore additional factors. We derive analytic expansions for the MSE of the ML parameter estimates on the basis of an adaptive design, which enables the analytical comparison of adaptive with fixed designs.

One main result of this paper is that one can theoretically expect a benefit of adaptive designs for sufficiently large sample sizes for a broad class of non-linear regression models. When the sample size is small, however, the remainder in (6) is non-negligible. This can lead to situations where the fixed design outperforms the adaptive design, as illustrated with three practical examples. In some applications, we can derive further general conclusions. For example, the efficiency ratio (9) reveals that adaptive designs are always more efficient than fixed designs for sufficiently small variances. In practice, more complex models than those considered in this paper are often used. The methodology presented in this paper remains applicable in this case, but the calculations become very cumbersome.

In particular, we illustrate in the problem of estimating the smallest dose achieving a certain percentage of the maximum treatment effect using a three-parameter $E_{\max}$ model that a sufficiently large sample size is necessary in the first stage for a better performance of the adaptive two-stage design. Moreover, the theoretical results enable us to understand the relationship of key factors impacting the relative efficacy of adaptive designs compared with fixed designs. For example, in the logistic regression example from Section 3.2, the relative efficiency depends only on three factors: the unknown degree of misspecification γ and the two design parameters p_0 and N_0. Using analytical methods, closed form expressions can be derived for the relationship between these factors, giving insight into their impact on efficiency performance. By contrast, simulation studies, even if performed comprehensively, do not provide theoretical explanations and are mainly used to provide empirical evidence.

Appendix A. Technical arguments

In this appendix, we give a sketch of a proof for the general expression (5). The representation (4) follows by similar arguments. Some details can be found in [41]. In order to derive the asymptotic distribution of $\sqrt{N_0}\left(\hat{\theta}_1 - \theta\right)$, we use a standard argument and the fact that the MLE $\hat{\theta}_1$ based on the observations $(d_1, Y_1), \ldots, (d_{N_0}, Y_{N_0})$ is a solution of the equation $0 = \sum_{i=1}^{N_0} \frac{\partial}{\partial\theta} \log f\left(Y_i, d_i, \hat{\theta}_1\right)$. Assume that the observations are taken according to a design ξ with $k \geqslant d$ different experimental conditions, say $t_1, \ldots, t_k$, and positive weights $w_1, \ldots, w_k$. Because $N_0 w_i \to \infty$ $(i = 1, \ldots, k)$, it follows by the strong law of large numbers that

$$\frac{1}{N_0} \sum_{i=1}^{N_0} \frac{\partial^2}{\partial^2\theta} \log f(Y_i, d_i, \theta) \longrightarrow \int\int \frac{\partial^2}{\partial^2\theta} \log f(y, d, \theta) f(y, d, \theta) \mathrm{d}y \, \mathrm{d}\xi(d),$$

and a standard argument shows that the right-hand side is equal to the matrix $-M(\xi, \theta)$, where $M(\xi, \theta)$ is defined in (1). A Taylor expansion yields

$$0 \approx \sqrt{N_0}\left(\hat{\theta}_1 - \theta\right) \frac{1}{N_0} \sum_{i=1}^{N_0} \frac{\partial^2}{\partial^2\theta} \log f(Y_i, d_i, \theta) + \frac{1}{\sqrt{N_0}} \sum_{i=1}^{N_0} \frac{\partial}{\partial\theta} \log f(Y_i, d_i, \theta).$$

This gives for any design ξ with positive masses at $k \geqslant d$ points

$$\sqrt{N_0}\left(\hat{\theta}_1 - \theta\right) \approx M^{-1}(\xi, \theta) \frac{1}{\sqrt{N_0}} \sum_{i=1}^{N_0} \frac{\partial}{\partial\theta} \log f(Y_i, d_i, \theta).$$

On the right-hand side, we have a sum of independent random variables, and the central limit theorem shows that it has an asymptotic (d-dimensional) normal distribution with mean 0 and covariance matrix $M^{-1}(\xi, \theta)$. Therefore, it follows using the locally optimal design ξ_{θ_0} and observing the definition of the matrix $I(\theta, \tau)$ in (2) that $\sqrt{N_0}\left(\hat{\theta}_1 - \theta\right) \approx I^{-1/2}(\theta, \theta_0) \cdot Z_{0,N_0}$, where

$$Z_{0,N_0} = \frac{1}{\sqrt{N_0}}\, I^{-1/2}(\theta, \theta_0) \sum_{i=1}^{N_0} \frac{\partial}{\partial \theta} \log f(Y_i, d_i(\theta_0), \theta) \xrightarrow{\mathcal{D}} \mathcal{N}_d(0, I_d). \tag{12}$$

The MLE $\hat{\theta}_A$ from the total sample depends on the sample from the first stage through the random variable $\hat{\theta}_1$, that is, $\hat{\theta}_A = \hat{\theta}_A\left(\hat{\theta}_1\right)$, and satisfies

$$0 = \sum_{i=1}^{N_0} \frac{\partial}{\partial \theta} \log f\left(Y_i, d_i(\theta_0), \hat{\theta}_A\right) + \sum_{i=n_0+1}^{N_0+N_1} \frac{\partial}{\partial \theta} \log f\left(Y_i, d_i\left(\hat{\theta}_1\right), \hat{\theta}_A\right),$$

where $d_{N_0+1}\left(\hat{\theta}_1\right), \ldots, d_{N_0+N_1}\left(\hat{\theta}_1\right)$ are the design points from the second stage (which depend on the parameter estimate $\hat{\theta}_1$ obtained in the first stage). The same argument as in the first part of the proof yields

$$\begin{aligned}
0 \approx\ & \frac{1}{N} \sum_{i=1}^{N_0} \frac{\partial^2}{\partial^2 \theta} \log f\,(Y_i, d_i(\theta_0), \theta)\, \sqrt{N}\left(\hat{\theta}_A - \theta\right) + \frac{1}{\sqrt{N}} \sum_{i=1}^{N_0} \frac{\partial}{\partial \theta} \log f\,(Y_i, d_i(\theta_0), \theta) \\
& + \frac{1}{N} \sum_{i=N_0+1}^{N_0+N_1} \frac{\partial^2}{\partial^2 \theta} \log f\left(Y_i, d_i\left(\hat{\theta}_1\right), \theta\right) \sqrt{N}\left(\hat{\theta}_A - \theta\right) + \frac{1}{\sqrt{N}} \sum_{i=N_0+1}^{N_0+N_1} \frac{\partial}{\partial \theta} \log f\left(Y_i, d_i\left(\hat{\theta}_1\right), \theta\right) \\
\approx\ & -\left(p_0\, I(\theta, \theta_0) + p_1\, I\left(\theta, \tilde{\theta}_1\right)\right) \sqrt{N}\left(\hat{\theta}_A - \theta\right) + \sqrt{p_0}\, I^{1/2}(\theta, \theta_0) Z_{0,N_0} + \sqrt{p_1}\, I^{1/2}\left(\theta, \tilde{\theta}_1\right) Z_{1,N_1},
\end{aligned}$$

where $\tilde{\theta}_1 = \theta_1 + \frac{1}{\sqrt{N_0}} I^{-1/2}(\theta, \theta_0) Z_{0,N_0}$ and Z_{0,N_0} and Z_{1,N_1} are defined by (12) and

$$Z_{1,N_1} = \frac{1}{\sqrt{N_1}}\, I^{-1/2}\left(\theta, \tilde{\theta}_1\right) \sum_{i=N_0+1}^{N_0+N_1} \frac{\partial}{\partial \theta} \log f\left(Y_i, d_i\left(\tilde{\theta}_1\right), \theta\right),$$

respectively. Note that we have used the fact that the design points $d_i(\theta)$ and the density f of the locally optimal design are continuously differentiable with respect to θ and x, respectively. This gives

$$\sqrt{N}\left(\hat{\theta}_A - \theta\right) \approx \left(p_0 I(\theta, \theta_0) + p_1 I\left(\theta, \tilde{\theta}_1\right)\right)^{-1} \left(\sqrt{p_0} I^{1/2}(\theta, \theta_0) Z_{0,N} + \sqrt{p_1} I^{1/2}\left(\theta, \tilde{\theta}_1\right) Z_{1,N_1}\right).$$

The variance of the random variable $\sqrt{N}\hat{\theta}_A$ can be calculated using the variance decomposition formula

$$\mathrm{Var}\left(\sqrt{N}\hat{\theta}_A\right) = \mathrm{E}\left[\mathrm{Var}\left(\sqrt{N}\hat{\theta}_A \mid Y_1, \ldots, Y_{N_0}\right)\right] + \mathrm{Var}\left(\mathrm{E}\left[\sqrt{N}\hat{\theta}_A \mid Y_1, \ldots, Y_{N_0}\right]\right), \tag{13}$$

and from the calculations of the previous paragraph, we obtain for the conditional expectation and variance given $Y_1, \ldots, Y_{N_0}$

$$E\left[\sqrt{N}\hat{\theta}_A \mid Y_1, \ldots, Y_{N_0}\right] \approx \left(p_0 I(\theta, \theta_0) + p_1 I\left(\theta, \tilde{\theta}_1\right)\right)^{-1} \sqrt{p_0} I^{1/2}(\theta, \theta_0) Z_{0,N_0},$$

$$\mathrm{Var}\left(\sqrt{N}\hat{\theta}_A \mid Y_1, \ldots, Y_{N_0}\right) \approx p_1 \left(p_0 I(\theta, \theta_0) + p_1 I\left(\theta, \tilde{\theta}_1\right)\right)^{-1} I\left(\theta, \tilde{\theta}_1\right) \left(p_0 I(\theta, \theta_0) + p_1 I\left(\theta, \tilde{\theta}_1\right)\right)^{-1}.$$

Here, we used the fact that $E[Z_{1,N_1} | Y_1, \ldots, Y_{N_0}] = 0$, $\mathrm{Var}(Z_{1,N_1} | Y_1, \ldots, Y_{N_0}) = I_d$ and that $\tilde{\theta}_1$ depends only on $Y_1, \ldots, Y_{N_0}$.

Next, we consider a Taylor expansion of the function $I\left(\theta, \tilde{\theta}_1\right)$, which gives for the element in the position (i, j) of the matrix $I(\theta, \theta+\tau)$

$$(I(\theta, \theta+\tau))_{ij} = (I(\theta, \theta))_{ij} + \nabla(I(\theta, \theta))_{ij}\tau + \frac{1}{2}\tau^T \nabla^2 (I(\theta, \theta))_{ij}\tau + o\left(\tau^2\right), \tag{14}$$

where the derivatives are taken with respect to the second argument of the matrix $I(\theta, \tau)$ and evaluated at $\tau = \theta$. Writing the expansion (14) in matrix form and using the notation $\tau = \frac{1}{\sqrt{N_0}} I^{-1/2}(\theta, \theta_0) Z_{0,N_0}$ yields

$$I\left(\theta, \tilde{\theta}_1\right) \approx I(\theta, \theta) + \frac{1}{\sqrt{N_0}} D_1(\theta, Z_{0,N_0}) + \frac{1}{2N_0} D_2(\theta, Z_{0,N_0}) + o_p\left(\frac{1}{N_0}\right) = I(\theta, \theta) + R(\theta, Z_{0,N_0}),$$

where

$$\begin{aligned} R(\theta, Z_{0,N_0}) &= \frac{1}{\sqrt{N_0}} D_1(\theta, Z_{0,N_0}) + \frac{1}{2N_0} D_2(\theta, Z_{0,N_0}), \\ D_1(\theta, Z_{0,N_0}) &= \left(\nabla(I(\theta, \theta))_{ij} I^{-1/2}(\theta, \theta_0) Z_{0,N_0}\right)_{i,j=1}^{d}, \\ D_2(\theta, Z_{0,N_0}) &= \left(Z_{0,N_0}^T I^{-1/2}(\theta, \theta_0) \nabla^2 (I(\theta, \theta))_{i,j} I^{-1/2}(\theta, \theta_0) Z_{0,N_0}\right)_{i,j=1}^{d} \end{aligned} \tag{15}$$

and ∇g and $\nabla^2 g$ denote the gradient and the Hessian matrix of a real valued function g. Assuming $A, B \in \mathbb{R}^{d\times d}$ with $\det A \neq 0$ and letting $\varepsilon \to 0$, we use the expansion

$$(A+\varepsilon B)^{-1} = \left(I_d + \varepsilon A^{-1} B\right)^{-1} A^{-1} = A^{-1}\left(I_d - \varepsilon B A^{-1} + \varepsilon^2 B A^{-1} B A^{-1}\right) + o\left(\varepsilon^2\right)$$

and obtain

$$\begin{aligned} E\left[\sqrt{N}\hat{\theta}_A \mid Y_1, \ldots, Y_{N_0}\right] \approx\ & H^{-1}(\theta, \theta_0)\left\{I_d - p_1 R(\theta, Z_{0,N_0}) H^{-1}(\theta, \theta_0) + p_1^2 S(\theta, Z_{0,N})\right\} \\ & \times \sqrt{p_0} I^{1/2}(\theta, \theta_0) Z_{0,N_0}, \end{aligned} \tag{16}$$

$$\begin{aligned} \mathrm{Var}\left[\sqrt{N}\hat{\theta}_A \mid Y_1, \ldots, Y_{N_0}\right] \approx\ & p_1 H^{-1}(\theta, \theta_0)\left\{I_d - p_1 R(\theta, Z_{0,N_0}) H^{-1}(\theta, \theta_0) + p_1^2 S(\theta, Z_{0,N})\right\} \\ & \times \{I(\theta, \theta) + R(\theta, Z_{0,N_0})\} \\ & \times \left\{I_d - p_1 R(\theta, Z_{0,N_0}) H^{-1}(\theta, \theta_0) + p_1^2 S(\theta, Z_{0,N})\right\}^T H^{-1}(\theta, \theta_0), \end{aligned} \tag{17}$$

where the matrix $S(\theta, Z_{0,N_0})$ is given by

$$S(\theta, Z_{0,N_0}) = \frac{1}{N_0} D_1(\theta, Z_{0,N_0}) H^{-1}(\theta, \theta_0) D_1(\theta, Z_{0,N}) H^{-1}(\theta, \theta_0). \tag{18}$$

The general structure of the information in (5) now follows from (13), observing the definition of the matrices R and S in (15) and (18), respectively.

Appendix B

For a proof of (7), we have to find the asymptotic variances of the estimators from the adaptive and fixed designs. For the fixed design, we have $\mathrm{Var}(\sqrt{N}\hat{\theta}_F) \approx I^{-1}(\theta.\theta_0)$. In order to obtain a similar expression for the MLE under the adaptive design, we observe the representation

$$I(\theta, \tau) = \int \left(\frac{\partial}{\partial\theta} \log f(y, \mathrm{d}(\tau), \theta)\right)^2 f(y, \mathrm{d}(\tau), \theta)\mathrm{d}y, \tag{19}$$

where we have used the fact that ξ_τ is a one-point design supported at the point $x(\tau)$. For each $\tau \in \Theta$, a one-point design, say ξ_τ, maximizes $M(\xi, \theta)$, and the corresponding support point $\mathrm{d}(\tau)$ is an interior

point of the dose range $\mathcal{D} \subset \mathbb{R}$. Consequently, it follows from (1) that for each $\tau \in \Theta$, the point $\mathrm{d}(\tau)$ is a solution of the equation

$$\frac{\partial}{\partial d} \int f(y, d, \theta) \left(\frac{\partial}{\partial \theta} \log f(y, d, \theta) \right)^2 \mathrm{d}y = 0. \tag{20}$$

This yields for the derivative of the first order in (14)

$$\nabla I(\theta, \tau)\Big|_{\tau=\theta} = \frac{\partial}{\partial d} \int f(y, d, \theta) \left(\frac{\partial}{\partial \theta} \log f(y, d, \theta) \right)^2 \mathrm{d}y \, \Big|_{d=d(\theta)} \cdot \frac{\partial}{\partial \tau} \mathrm{d}(\tau)\Big|_{\tau=\theta} = 0,$$

where the last identity follows from (20). Similarly, we obtain for the second derivative

$$g(\theta) := \nabla^2 I(\theta, \tau)\Big|_{\tau=\theta} = \frac{\partial^2}{\partial^2 d} \int f(y, d, \theta) \left(\frac{\partial}{\partial \theta} \log f(y, d, \theta) \right)^2 \mathrm{d}y \, \Big|_{d=d(\theta)} \cdot \left(\frac{\partial}{\partial \tau} \mathrm{d}(\tau)\Big|_{\tau=\theta} \right)^2,$$

where $g(\theta) < 0$ because $d(\theta)$ maximizes the function in (19). Consequently, we have $D_1(\theta, Z_{0,N_0}) = 0$, $S(\theta, Z_{0,N_0}) = 0$ and obtain for the matrix $R(\theta, Z_{0,N_0})$ defined by (15)

$$R(\theta, Z_0) = \frac{1}{2N_0} \frac{g(\theta) Z_{0,N_0}^2}{I(\theta, \theta_0)},$$

which, together with (16) and (17), yields as approximation for the variance of

$$\begin{aligned}
\mathrm{Var}\left(\sqrt{N}\hat{\theta}_A\right) &= \mathrm{Var}\left(E\left[\sqrt{N}\hat{\theta}_A \mid Y_1, \ldots, Y_{N_0}\right]\right) + E\left[\mathrm{Var}\left(\sqrt{N}\hat{\theta}_A \mid Y_1, \ldots, Y_{N_0}\right)\right] \\
&\approx E\left[\frac{p_0 Z_{0,N_0}^2 I(\theta, \theta_0)}{H^2(\theta, \theta_0)} \left(1 - \frac{p_1 g(\theta) Z_{0,N_0}^2}{2N_0 I(\theta, \theta_0) H(\theta, \theta_0)} \right)^2 \right] \\
&\quad + p_1 E\left[\frac{1}{H^2(\theta, \theta_0)} \left(1 - \frac{p_1 g(\theta) Z_{0,N_0}^2}{2N_0 I(\theta, \theta_0) H(\theta, \theta_0)} \right)^2 \left(I(\theta, \theta) + \frac{Z_{0,N_0}^2 g(\theta)}{2N_0 I(\theta, \theta_0)} \right) \right] \\
&= \frac{p_0 I(\theta, \theta_0)}{H^2(\theta, \theta_0)} \left(1 - \frac{3 p_1 g(\theta)}{N_0 I(\theta, \theta_0) H(\theta, \theta_0)} \right) \\
&\quad + \frac{p_1 I(\theta, \theta)}{H^2(\theta, \theta_0)} \left\{ 1 + \frac{g(\theta)(p_0 I(\theta, \theta_0) - p_1 I(\theta, \theta))}{2N_0 I(\theta, \theta_0) H(\theta, \theta_0) I(\theta, \theta)} \right\} \\
&= \frac{1}{H(\theta, \theta_0)} - \frac{g(\theta) p_1 (5 p_0 I(\theta, \theta_0) + p_1 I(\theta, \theta))}{2N_0 H^3(\theta, \theta_0) I(\theta, \theta_0)}.
\end{aligned}$$

Calculating the ratio of the two asymptotic variances proves (7).

Acknowledgements

The authors thank Martina Stein, who typed parts of this manuscript with considerable technical expertise. This work has been supported in part by the Collaborative Research Center 'Statistical modeling of non-linear dynamic processes' (SFB 823, Teilprojekt C2) of the German Research Foundation (DFG). The authors would also like to thank two referees for their constructive comments on an earlier version of this paper.

References

1. Chernoff H. Locally optimal designs for estimating parameters. *Annals of Mathematical Statistics* 1953; **24**:586–602.
2. Ford I, Torsney B, Wu CFJ. The use of canonical form in the construction of locally optimum designs for nonlinear problems. *Journal of the Royal Statistical Society, Series B* 1992; **54**:569–583.
3. He Z, Studden WJ, Sun D. Optimal designs for rational models. *Annals of Statistics* 1996; **24**:2128–2142.
4. Fang X, Hedayat AS. Locally D-optimal designs based on a class of composed models resulted from blending Emax and one-compartment models. *Annals of Statistics* 2008; **36**:428–444.

5. Yang M. On the de la Garza Phenomenon. *Annals of Statistics* 2010; **38**(4):2499–2524.
6. Chaloner K, Verdinelli I. Bayesian experimental design: a review. *Statistical Science* 1995; **10**(3):273–304.
7. Dette H. Designing experiments with respect to "standardized" optimality criteria. *Journal of the Royal Statistical Society, Series B* 1997; **59**:97–110.
8. Müller CH, Pázman A. Applications of necessary and sufficient conditions for maximum efficient designs. *Metrika* 1998; **48**:1–19.
9. Wiens DP. Robust discrimination designs, with MATLAB code. *Journal of the Royal Statistical Society, Series B* 2009; **71**:805–829.
10. Wetherill GB. Sequential estimation of quantal response curves (with discussion). *Journal of the Royal Statistical Society, Series B* 1963; **25**(1):1–48.
11. Box GEP, Hunter WG. Sequential design of experiments for nonlinear models. In *Proceedings of IBM Scientific Computing Symposium*, Korth JJ (ed.). IBM: White Plains, 1965; 113–137.
12. Ford I, Silvey SD. A sequentially constructed design for estimating a nonlinear parametric function. *Biometrika* 1980; **67**(2):381–388.
13. Rosenberger WF, Hughes-Oliver JM. Inference from a sequential design: proof of conjecture by Ford and Silvey. *Statistics and Probability Letters* 1999; **44**:177–180.
14. Gautier R, Pronzato L. Some results on optimal allocation in two-step sequential design. *Biometrical Letters* 1999; **36**:15–30.
15. Roy A, Ghosal S, Rosenberger W. Convergence properties of sequential Bayesian D-optimal designs. *Journal of Statistical Planning and Inference* 2008; **139**:425–440.
16. Chang H, Ying Z. Nonlinear sequential designs for logistic item response theory models with applications to computerized adaptive tests. *Annals of Statistics* 2009; **37**:1466–1488.
17. Zhu W, Wong WK. Multiple-objective designs in a dose–response experiment. *IMS Lecture Notes* 1998; **34**:73–82.
18. Haines LM, Perevozskaya I, Rosenberger WF. Bayesian optimal design for phase I clinical trials. *Biometrics* 2003; **59**:591–600.
19. Montepiedra G, Yeh AB. Two-stage designs for identification and estimation of polynomial models. *Computational Statistics and Data Analysis* 2004; **46**:531–546.
20. Bornkamp B, Bretz F, Dmitrienko A, Enas G, Gaydos B, Hsu CH, Koenig F, Krams M, Liu Q, Neuenschwander B, Parke T, Pinheiro J, Roy A, Sax R, Shen F. Innovative approaches for designing and analyzing adaptive dose-ranging trials (with discussion). *Journal of Biopharmaceutical Statistics* 2007; **17**:965–995.
21. Miller F, Guilbaud O, Dette H. Optimal designs for estimating the interesting part of a dose–effect curve. *Journal of Biopharmaceutical Statistics* 2007; **17**:1097–1115.
22. Dragalin V, Hsuan F, Padmanabhan SK. Adaptive designs for dose-finding studies based on sigmoid E_{max} model. *Journal of Biopharmaceutical Statistics* 2007; **17**(6):1051–1070.
23. Bornkamp B, Bretz F, Dette H, Pinheiro J. Response-adaptive dose-finding under model uncertainty. *Annals of Applied Statistics* 2011; **5**(2B):1611–1631.
24. Fedorov VV, Wu Y, Zhang R. Optimal dose-finding designs with correlated continuous and discrete responses. *Statistics in Medicine, to appear,* 2012; **31**:217–234.
25. Dragalin V, Bornkamp B, Bretz F, Miller F, Padmanabhan SK, Patel N, Perevozskaya I, Pinheiro J, Smith JR. A simulation study to compare new adaptive dose-ranging designs. *Statistics in Biopharmaceutical Research* 2010; **2**:487–512.
26. Kiefer J. General equivalence theory for optimum designs (approximate theory). *Annals of Statistics* 1974; **2**:849–879.
27. Pukelsheim F. *Optimal Design of Experiments*. SIAM: Philadelphia, 2006.
28. Atkinson AC, Donev A, Tobias RD. *Optimum Experimental Designs, with SAS*. Oxford University Press: Oxford, 2007.
29. Box MJ. Bias in nonlinear estimation. *Journal of the Royal Statistical Society, Series B* 1971; **33**:171–201.
30. Dette H, Bretz F, Pepelyshev A, Pinheiro J. Optimal designs for dose finding studies. *Journal of the American Statistical Association* 2008; **103**(483):1225–1237.
31. Dette H, Kiss C, Bevanda M, Bretz F. Optimal designs for the EMAX, log-linear and exponential models. *Biometrika* 2010; **97**(2):513–518.
32. Jennrich RI. Asymptotic properties of non-linear least squares estimators. *Annals of Mathematical Statistics* 1969; **40**(2):633–643.
33. Dette H, Neugebauer HM. Bayesian optimal one point designs for one parameter nonlinear models. *Journal of Statistical Planning and Inference* 1996; **52**:17–31.
34. Rasch G. *Probabilistic Models for Some Intelligence and Attainment Tests*. University of Chicago Press: Chicago, 1960.
35. Silvey SD. *Optimal Design*. Chapman & Hall: London, 1980.
36. Wu CFJ. Optimal design for percentile estimation of a quantal-response curve. In *Optimal Design and Analysis of Experiments*, Dodge Y, Fedorov VV, Wynn HP (eds). Elsevier: North Holland, 1988; 213–223.
37. Cameron AC, Trivedi PK. *Regression Analysis of Count Data*. Cambridge University Press: Cambridge, 1998.
38. Pinheiro J, Bornkamp B, Bretz F. Design and analysis of dose-finding studies combining multiple comparisons and modeling procedures. *Journal of Biopharmaceutical Statistics* 2006; **16**:639–656.
39. Boroujerdi M. *Pharmacokinetics: Principles and Applications*. McGraw Hill: New York, 2002.
40. Bretz F, Dette H, Pinheiro J. Practical considerations for optimal designs in clinical dose finding studies. *Statistics in Medicine* 2010; **29**:731–742.
41. Dette H, Bornkamp B, Bretz F. On the efficiency of adaptive designs. *Technical Report*, University of Dormund, 2010. SFB 823 Discussion Paper **40/10**, Available from: http://www.statistik.tu-dortmund.de/sfb823-dp2010.html.