Optimal designs for estimating the interesting part of a dose-effect curve

Frank Miller[1][*], Holger Dette[2] and Olivier Guilbaud[1]

[1]Clinical Information Science, AstraZeneca, Södertälje, Sweden
[2]Fakultät für Mathematik, Ruhr-Universität Bochum, Germany

ABSTRACT

We consider a dose-finding trial in phase IIB of drug development. For choosing an appropriate design for this trial the specification of two points is critical: an appropriate model for describing the dose-effect relationship and the specification of the aims of the trial (objectives), which will be the focus in the present paper.

For many practical situations it is essential to have a robust trial objective that has little risk of changing during the complete trial due to external information. An important and realistic objective of a dose-finding trial is to obtain precise information about the interesting part of the dose-effect curve. We reflect this goal in a statistical optimality criterion and derive efficient designs using optimal design theory. In particular we determine non-adaptive Bayesian optimal designs, i.e. designs which are not changed by information obtained from an interim analysis. Compared with a traditional balanced design for this trial it is shown that the optimal design is substantially more efficient. This implies either a gain in information or essential savings in sample size. Further, we investigate an adaptive Bayesian optimal design that uses two different optimal designs before and after an interim analysis, and we compare the adaptive with the non-adaptive Bayesian optimal design. The basic concept is illustrated using a modification of a recent AstraZeneca trial.

[*]Correspondance: Frank Miller, Clinical Information Science, AstraZeneca, S–15185 Södertälje, Sweden. E-mail: frank.miller@astrazeneca.com

Key Words: Adaptive design; Bayesian design; Clinical trial; Dose-finding; Optimal design; Interesting part of dose-effect curve; Prior knowledge; Two-stage design.

1 Introduction

Before the decision is taken to move into phase III of drug development, a dose-finding trial is performed to determine the dose (or doses) appropriate for the phase III trials and to obtain a solid basis for the decision of continuing to phase III or not. A typical dose-finding trial consists of a few hundred patients, a few doses and the treatment time is often somewhat shorter than in the succeding phase III trials. This trial is performed at the end of phase II in the development process and usually called phase IIB trial. For a general overview about dose-finding studies, we refer to recent monographs by Chevret (2006) and Ting (2006).

One general problem for the pharmaceutical industry is that there is often a rather poor understanding about the dose-effect profile of a new drug when going into phase III despite a quite long development process. The dose-finding trial of phase IIB provides some information on this profile. But it is often felt that there is still insufficient knowledge afterwards and that one could have done better. These difficulties are partially caused by the fact that traditional (balanced) designs are used for this trial, which may be suboptimal. It is clear that a good knowledge of the dose-effect properties is very important for the chances of success of the chosen doses for phase III. Furthermore, phase IIB is often the last possibility before launch of the new drug to investigate this dose-effect relation in detail: only one or two doses are usually used in phase III. Moreover, knowledge of the dose-effect relation would be also useful for physicians and patients when the drug is finally on the market.

For these reasons more innovative designs for dose-finding trials are of particular interest. Running a clinical trial by a traditional approach using parallel groups with placebo and a number of dose arms with equal allocation throws away possibilities to obtain information in a more efficient way. A more innovative way of designing experiments for dose-finding trials is to think carefully about useful doses for this trial, to modify the allocation ratios for the doses and to think about the advantages or disadvantages of a possible change of the design after interim data of the trial has been evaluated. The last point of a potential adaptation got recently very much interest. The European regulatory authorities (Committee for Medicinal Products for Human Use; 2006) developed a reflection paper (currently in draft version) for confirmatory trials (phase III) with flexible analysis plans. For the phase IIB trial considered

in this paper we will discuss the possibility of an adaptive choice of doses or allocation ratios. A recommendation of the Pharmaceutical Research and Manufacturers of America working group regarding adaptive dose-response trials was presented by Gaydos et al. (2006). They reviewed non-adaptive and adaptive dose-response designs both for early (phase I) and later-stage (phase II) drug development. Dose-finding trials conducted with adaptive design are described by Krams et al. (2003) and Smith et al. (2006).

When designing a trial, especially if non-traditional approaches are considered, the specification of two points is critical: what knowledge about the dose-effect relationship (a priori knowledge) is available and what are the aims of the trial (objectives). There are different possible sources to gain a priori knowledge: for example earlier trials with the same drug, preclinical information, published data with similar drugs. All information has to be analysed carefully and it is often possible to derive guesses for an anticipated effect - even if these are very rough.

In this paper we focus more on the second point mentioned above, the specification of trial objectives. Sometimes, the main objective of a dose-finding trial is defined as identification of a single dose for phase III. However, for many practical situations, it is necessary to specify a broader objective for the trial. Usually it takes a long time to perform such a dose-effect trial, in many cases several years. During this period, a lot of new information can be obtained from sources external to the trial. For example, there are other trials ongoing with the compound under investigation and new information might be obtained by those including further information about safety and tolerability. Or, a competing company may launch a new drug for the same indication and requirements for a clinical relevant effect may change.

Since many things can happen during a phase IIB trial, we need a more flexible objective. We think a realistic objective for a dose-finding trial is to get information about the interesting part of the dose-effect curve. We make the notion "interesting part" more concrete in this paper and describe it in statistical terms. Based on the specification of this objective, we derive and investigate appropriate designs using optimal design theory. A brief overview how optimal design theory can be applied to dose-finding trials is included in the article of Dragalin (2006).

The remaining part of this paper is organised as follows. In the next section we describe a specific situation of a phase IIb trial which motivated our investigations. For the sake of a transparent presentation of the basic concepts we concentrate on this specific trial, but it is emphazised that the main ideas presented in this paper are applicable in other situations.

In Section 3 we describe the objectives of the dose-finding trial under consideration and reflect these in statistical terms. We apply optimal design theory in Section 4 to derive a non-adaptive Bayesian optimal design, i.e. a design which is not changed by information obtained from an interim analysis of the available data. It is demonstrated that this non-adaptive design provides answers to our objectives in a more efficient way than the commonly used balanced design. In Section 5, we develop the Bayesian optimal design further by taking into account that an interim analysis is often conducted in dose-finding trials, which provides additional information about the dose response relationship. A modification of the allocation ratios for the dose arms based on this information leads to an adaptive Bayesian optimal design. This concept has been discussed before in the context of model discrimination for nested linear models by Montepiedra and Yeh (1998, 2003). The conduct of such a design requires some further logistical and operational considerations, see e.g. Quinlan and Krams (2006). We compare the efficiencies of the non-adaptive and the adaptive Bayesian optimal design in Section 6 and conclude the paper with a brief discussion in Section 7. Statistical details are postponed to the Appendix.

2   The dose-finding trial

The motivation for the investigation described in this paper was the planning for an AstraZeneca phase IIB trial. For confidentiality reasons, we cannot present all details about the medical background of this trial and we can only publish modified values about assumptions. However, we have modified our example in a way that essential properties and conclusions carry over from the "real" example to our partially artificial example.

The planning for the design of this trial was done when data from phase I trials (single ascending dose and multiple ascending dose trial for healthy volunteers) were available. As result of these trials it was concluded that doses higher than 100 mg have too high safety risks. For doses up to this threshold, good safety experiences were gained and the drug was well tolerated.

In the trials that will be performed in phase II and III for patients, a primary variable will be used measuring the effect of the drug. We do not specify this variable here, but we use in this paper a continuous variable where higher values indicate a positive effect of the drug. Until now, there are no effect-results for the drug available. However, pharmacometric modellers could use various sources of information to get possible dose-effect-relationships within our dose range between 0 and 100 mg. Pharmacokinetic analyses from the phase I trials were available, preclinical data could be used and clinical data from another compound

with similarities to the drug under investigation was another important source to get some anticipations about a possible dose-effect curve. As first result of these investigations, an $E_{\max}$-sigmoid model was seen as an adequate and flexible dose-effect-model. This model has the form:

$$f(x) = E_0 + \frac{E_{\max} x^\alpha}{ED_{50}^\alpha + x^\alpha},$$

where x is the dose in mg and $f(x)$ denotes the true effect for dose x. The parameter E_0 describes the placebo response, $E_{\max}$ is the maximum effect for a very high dose compared with placebo, ED_{50} is the dose with half of the maximum effect and α is a parameter, which influences the shape of the dose-effect-curve. For more details regarding this model we refer to Holford and Sheiner (1981) and Bezeau and Endrenyi (1986), for example, while some optimal designs for estimating the parameters in this model can be found in Dette, Melas and Wong (2005). However estimation of the parameters is usually not the main objective of a dose-finding trial.

In the trial under consideration the experts were quite sure that the dose-effect curve is monotone in the dose range of interest (up to 100 mg). Hence, curves with an umbrella shape were not likely and it seems that the $E_{\max}$-sigmoid model is a rather robust model for the possible monotone curves. As prior guess for the values of the parameters in this model based on the available information, the pharmacometric modellers obtained $E_0 = 22$, $E_{\max} = 11.2$, $ED_{50} = 70$ mg and $\alpha = 1$. There are of course several uncertainties in these guesses. The preclinical evidence suggests that $E_{\max}$ values up to $1.5 \cdot 11.2 = 16.8$ may also be possible. The determination of the ED_{50} was somewhat difficult, therefore half of the prior guess value (35 mg) is used as an optimistic bound and 200 mg as a conservative value. A value of $\alpha = 1$ seemed to be a quite good choice based on the information, however, dose-effect-curves were of interest where low doses have almost no effect. These curves were achieved with higher α values, hence $\alpha = 2$ and 4 were used. A dose-effect-curve which is more flat for higher doses but has some effect overall should also be taken into account and such a curve was modelled by a low $E_{\max}$ in combination with a low ED_{50}. From these considerations, seven anticipated scenarios were set up. The parameters corresponding to these scenarios are listed in Table 1 and the different curves are visualised in Figure 1. Finally we assume a standard deviation of $\sigma = 10$ for all cases, although this will have no effect on the optimal designs derived below (but it will have some effect in the simulation study performed in Section 6).

It appeared to be important to make also a rough judgement about the probabilities of these scenarios. The prior guess scenario should have the largest probability. The Scenarios 4 and 7 seem also very likely, followed by Scenario 6 with an effect for higher doses. The

Table 1: Parameters for anticipated dose-effect scenarios

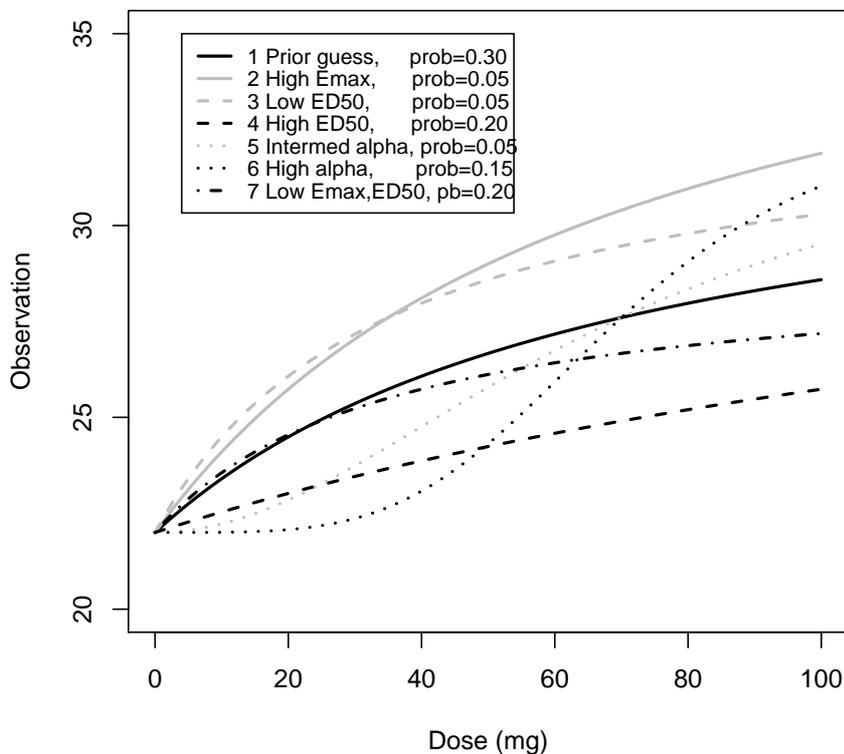|	Scenario	E_0	$E_{\max}$	ED_{50}	α	probability π_j
1	Prior guess	22	11.2	70	1	0.30
2	High $E_{\max}$	22	16.8	70	1	0.05
3	Low ED_{50}	22	11.2	35	1	0.05
4	High ED_{50}	22	11.2	200	1	0.20
5	Intermed α	22	11.2	70	2	0.05
6	High α	22	11.2	70	4	0.15
7	Low $E_{\max}, ED_{50}$	22	7.0	35	1	0.20

Figure 1: Anticipated dose-effect scenarios

other scenarios, Scenario 5 and the optimistic Scenarios 2 and 3, seem to be less likely. These considerations were translated to rough values for probabilities for the scenarios, see the last column in Table 1 and Figure 1. We call these probabilities "a priori probabilities", since they are based on the a priori knowledge of the clinical team. The a priori probability for Scenario j is denoted by π_j $(j = 1, \ldots, 7)$.

In the dose-finding trial for which the design was constructed, we have multiples of 20 mg as available strengths of the new drug and there is also a placebo available. Therefore, possible treatment arms are placebo and the doses 20, 40, 60, 80, and 100 mg. A balanced design assigning an equal proportion of patients to each arm is a common choice in this context. In our case it allocates 1/6 of the patients to placebo and each of the five active doses. We investigate in the following sections unbalanced designs and compare the efficiency of these designs with the balanced design which we treat as a reference design.

Based on a balanced design, a linear-contrast-test of the null-hypothesis that all dose-arms have the same effect as placebo (with one-sided significance level $\alpha = 0.025$, power $1 - \beta = 0.8$) versus the alternative of Scenario 7 with a maximal effect of about 5 units compared to placebo and standard deviation $\sigma = 10$, a sample size of $n_{\text{total}} = 300$ patients was determined. For convenience, we use this sample size for all designs under investigation.

3 Trial objectives

The main purpose of the trial is, roughly speaking, to characterise the interesting part of the dose-effect curve. But what means interesting in the present context? Based on results of competing drugs on the market, effects lower than $\delta = 5$ compared with the placebo-effect are of no medical interest. If there is a dose with an effect lower than δ, it is not worth to investigate this dose further since there exist already better treatments on the market.

Let x_δ be the unknown dose which has an effect of δ and x_{max} be the maximum tolerable dose, in our case $x_{\text{max}} = 100mg$. The objectives of the trial are (see also Figure 2):

1. To estimate the dose-effect (compared to the placebo-effect) for doses between x_δ and x_{max}.

2. To estimate the effect at the maximal dose x_{max} compared to the placebo-effect.

3. To estimate the dose x_δ with effect δ.

4. To prove that there is a positive effect at some active dose compared to placebo.
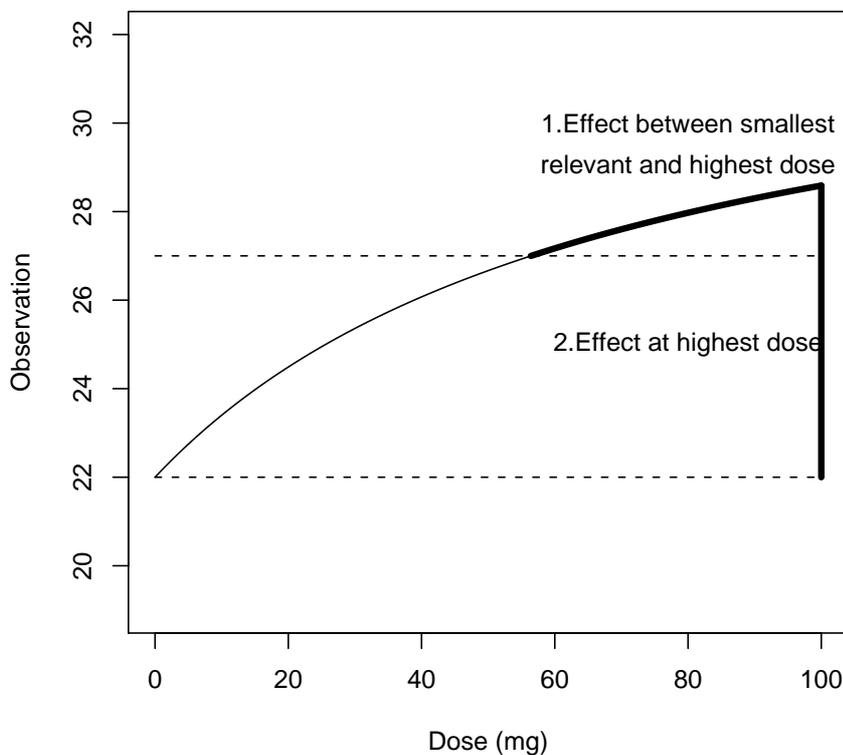
Figure 2: Objectives of the trial

It is worth to highlight that all these objectives are defined via the dose-effect curve itself and not in terms of the four parameters in the $E_{\max}$-sigmoid model. This is important from the point of view of estimation. In our situation we anticipate that the $E_{\max}$-sigmoid curve will not reach its plateau. Dutta et al. (1996) investigate this situation and conclude that if the highest dose in the trial is below ED_{95}, the parameters in this model are poorly estimated. Despite this, the curve itself can be well estimated via the least squares method within the range of observations. Since our interest is to obtain knowledge about effects for doses up to 100 mg but not the model parameters, we have no problem with this issue.

Having the Objectives 1-4 defined above, the question is how we should choose the design of the trial to get the best answers (estimates and test decision) for our objectives. As long as we use a non-adaptive design, it is specified by the total number of patients n, the number of doses k (inclusive placebo), the dose levels $x_0, \ldots, x_{k-1}$ (in mg) and the proportions of

patients $w_0, \ldots, w_{k-1}$ allocated to each dose level. We assume that the total sample size n is already determined. For our example, n was determined in Section 2 ($n_{\text{total}} = 300$) for the balanced design and the goal is now to get better estimates with other designs using the same sample size. We assume further that it is already decided which doses are possible, in our case we have $k = 6$ and $x_0 = 0, x_1 = 20, \ldots, x_5 = 100$. So we focus here on the optimisation of the proportions of patients allocated to each dose level. We don't exclude that a proportion is 0, hence it is possible to use actually less than 6 different dose levels.

For the construction of an efficient experimental design (i.e. the appropriate allocation proportions), we will apply optimal design theory [see Silvey (1980), for example]. As developed in the appendix, a function $d(x, \text{design}, \text{scenario})$ can be specified which is asymptotically proportional to the variance of the nonlinear least squares estimate $\hat{f}(x) - \hat{f}(0)$ of the dose-effect function $f(x) - f(0)$ at dose x compared to placebo, given a fixed design and a fixed scenario. Because these estimates are asymptotically unbiased, this quantity measures the precision of the estimate, and it can be shown that the function $d(x, \text{design}, \text{scenario})$ depends sensitively on the design of the experiment. Therefore the goal in this context is to minimize this variance by the choice of the design or, alternatively, to maximize the information $1/d(x, \text{design}, \text{scenario})$.

In this paper, we focus on Objective 1 and 2 and use the following optimality criteria for these objectives.

Criterion 1: For the first objective, the function $x \mapsto d(x, \text{design}, \text{scenario})$ should be small for all x between x_δ and $x_{\max}$. We use the I_L-criterion proposed in Dette and O'Brien (1999) with $L = 1$, which determines the design such that the function

$$\Phi_1(\text{design}, \text{scenario}) = \left\{ \int_{x_\delta}^{x_{\max}} d(x, \text{design}, \text{scenario}) \, dx \right\}^{-1}$$

is maximal. It would be possible, to use unequal weightings of different doses, if a good estimation is more important for certain doses than for others. Further, an alternative would be to minimize the largest variance instead of the integrated variance (a type of G-optimality) or to use other types of integrals, see Dette and O'Brien (1999).

Criterion 2: This criterion is simpler than the one before, and we use

$$\Phi_2(\text{design}, \text{scenario}) = 1/d(x_{\max}, \text{design}, \text{scenario}),$$

which corresponds to the minimization of the variance of the estimate for the effect of the maximal tolerable dose $x_{\max}$.

For the sake of simplicity we do not consider the other two objectives (3 and 4) mentioned above for the construction of optimal designs. We mention briefly how one can define optimality criteria in these cases and refer to literature for further information. It will be clear from the following discussion how the method can extended to incorporate these criteria as well.

> **Criterion 3:** We estimate x_δ by $\hat{x}_\delta$ defined via $\hat{f}(\hat{x}_\delta) - \hat{f}(0) = \delta$. Dette et al. (2007) applied the *delta*-method (Van der Vaart, 1988) to derive the asymptotic variance of $\hat{x}_\delta$ and used this formula to define an optimality criterion, which is called MED-optimality.

> **Criterion 4:** For a precise definition of the optimality criterion for Objective 4 it is necessary to specify the test, which will be used to prove that there is a positive effect of the drug. In this context, tests based on contrasts are very common, see Stewart and Ruberg (2000) and Bretz et al. (2005). For example, if we use a single contrast test for a prespecified and fixed contrast, the test statistic has a noncentral t-distribution with a noncentrality parameter $\tau = \tau(\text{design}, \text{scenario})$ depending on both the design and the scenario, see Bretz et al. (2005). We can define the optimality criterion as the square of the noncentrality parameter.

A commonly used design is the balanced design with 1/6 of the patients allocated to placebo and each of the five active doses. We define the relative efficiency of an arbitrary design with respect to the balanced design by

$$\text{Eff}_c(\text{design}, \text{scenario}) = \Phi_c(\text{design}, \text{scenario}) / \Phi_c(\text{balanced design}, \text{scenario}), \qquad (3.1)$$

where c is the criterion. An efficiency > 1 means that the design is more efficient than the balanced design. Moreover, we can directly interprete the efficiency in terms of sample size: for example, an efficiency of 1.25 means that the balanced design would need 25% of patients more than the design under consideration to obtain estimates with approximately the same precision (where we compare different designs by the criterion "c").

4 Bayesian optimal design

Criterion 1 (dose-effect in the dose range between x_δ and $x_{\max}$) was judged to be the most important criterion. We want to choose a design for the trial which has maximal efficiency according to this criterion. However, this criterion makes only sense if there exists a dose with effect δ between 0 and $x_{\max}$ *mg*. If there is no such dose, we want to have a good

estimate for the effect at the highest dose (Criterion 2), since this is cruical for the decision of whether it is time to stop the development of the drug or whether there is any hope that the drug could have some value. For our trial, this means that we look at Criterion 1 for all scenarios except for Scenario 4. For Scenario 4, we look at Criterion 2.

Since the true scenario is unknown, we determine an overall optimal design by maximising the weighted mean of the efficiencies for certain scenarios. According to the description above, we define our main function to be optimised by

$$\Psi(\text{design}) = \sum_{c=1}^{2} \sum_{j=1}^{7} v_{j,c} \text{Eff}_c(\text{design}, \text{scenario} = j) \tag{4.1}$$

where

$$v_{j,c} = \begin{cases} \pi_j, & \text{if } (j \neq 4 \text{ and } c = 1) \text{ or } (j = 4 \text{ and } c = 2), \\ 0, & \text{else.} \end{cases}$$

The π_j are the a priori probabilities for Scenario j, see Table 1. Note that it is clear from the definition of Ψ in (4.1) how other criteria (like the criteria mentioned in Section 3) could be incorporated. We could calculate the sum over the efficiencies for these criteria as well after we have specified appropriate weights $v_{j,c}$. This optimality criterion of the weighted mean of the efficiencies for the scenarios is called Bayesian- or multiobjective optimality criterion in the literature on optimal designs [see e.g. Dette, Haines and Imhof (2005) or Zhu and Wong (2000)]. We call the resulting design "Bayesian optimal design" and sometimes "non-adaptive Bayesian optimal design" in contrast to the adaptive designs considered in Section 5. As mentioned before, we assume that the only available doses are 20, 40, 60, 80, and 100 *mg*.

Table 2: Bayesian optimal design

Dose (in *mg*)	0	20	40	60	80	100
Weight	0.417	0.023	0.023	0.126	0.112	0.299
n	125	7	7	38	33	90

For the trial discussed in this paper, the Bayesian optimal design was computed numerically, and the allocation weights are shown in Table 2. More weight is put on the doses 60, 80, and 100 *mg* and on placebo than on 20 and 40 *mg*. To have a higher proportion of patients assigned to placebo is important since we are interested in the differences between the placebo and the other doses. Allocation of patients to the maximal tolerable dose 100 *mg* is important to get information about the right end of the dose-effect curve. The "interesting"

part of the dose-effect curve (i.e. the interval between x_δ and $x_{\max}$) starts for most of the scenarios (Scenario 1, 5, 6, 7) around 60 *mg*.

Since the sample size for the trial is 300 patients, a rounding procedure is applied to obtain the integers from the not necessarily integer valued quantities, which are obtained by multiplying the weights from Table 2 with the sample size $n = 300$ [see Pukelsheim and Rieder (1992)]. We try to get these number of patients for each dose as closely as possible, but depending on the randomisation procedure of the trial, we can't be sure to achieve these numbers exactly.

Compared with the balanced design, the Bayesian optimal design has an average efficiency of $\Psi(\text{design}) = 1.55$. In other words, if Ψ is used to measure the quality of an experimental design, the balanced design requires 55% more patients than the Bayesian optimal design to obtain estimates with a comparable precision.

For a more detailed comparison of the two designs one has to check the efficiency of this design also under Criterion 2 and under the different scenarios and criteria seperately. For this purpose we determined the efficiencies defined by (3.1) for all scenarios and the two criteria. The corresponding results are listed in Table 3. As mentioned before, there exists no dose with an effect of δ for Scenario 4 and consequently there is no efficiency for Criterion 1. For the computation of the overall efficiency for Criterion 2, we have used formula (4.1), with $v_{j,1} = 0$ and $v_{j,2} = \pi_j$, the a priori probability of Scenario j. We observe in nearly all cases a substantial improvement in efficiency if the Bayesian optimal design is used instead of the balanced design. In particular the Bayesian optimal design is between 70% and 106% more efficient for estimating the effect of the maximal tolerable dose $x_{\max}$ (Criterion 2). Under the Criterion 1 the adavantages of the Bayesian optimal design are less substantial but still clearly visible. The overall efficiency is 1.55 and only under the Scenario 6 the commonly used balanced design is 11% more efficient for estimating the effect of the dose in the range $[x_\delta, x_{\max}]$. In all other cases the Bayesian optimal design should be perferred.

5 Adaptive Bayesian optimal designs

The a priori probabilities for the seven scenarios were quite rough guesses based on the information available before the trial. Since in the trial under consideration there was the opportunity to perform an interim analysis, it seems to be reasonable to revise these probabilities in the light of the interim results. With the modified probabilities a new optimal design can be calculated. Two-stage designs of this type have been discussed before in the context of constructing optimal designs for model discrimination [see e.g. Montepiedra and

Table 3: Efficiencies of the Bayesian optimal design compared to the balanced design

		Criterion	
Scenario		$c = 1$	$c = 2$
1	Prior guess	1.48	1.97
2	High $E_{\max}$	1.10	1.97
3	Low ED_{50}	1.08	1.93
4	High ED_{50}	-	2.02
5	Intermed α	1.36	2.06
6	High α	0.89	1.71
7	Low $E_{\max}, ED_{50}$	1.98	1.93
Overall		1.55	1.93

Yeh (1998, 2003) and the references in these papers].

In the trial discussed in this paper an interim analysis was performed after $1/3$ of the patients has been investigated, i.e. the sample size in the first stage is $n^{(1)} = 100$. Until the interim analysis, the patients will be assigned to treatments according to the (non-adaptive) Bayesian optimal design determined in the previous section. We call this first part of the trial stage I. During the time necessary for interim analysis recruitment for the trial continues and we still assign the patients to treatment according to the (non-adaptive) Bayesian optimal design. We assume that $n^{(1o)} = 40$ patients will be treated during the time of interim analysis and we call these patients "stage I overrun". In the interim analysis we calculate a posteriori probabilities for the seven scenarios using the Bayes formula. We call the part of the trial with the remaining $n^{(2)} = 160$ patients stage II. For stage II, we will use another design than in stage I and stage I overrun, based on the optimal design for the a posteriori probabilities, see below. Let $n_i^{(1)}$ be the number of patients for dose level $i = 0, 1, \ldots, 5$ in stage I (corresponding to $0, 20, \ldots, 100mg$, respectively), $n_i^{(1o)}$ the number of patients in the stage I overrun and $n_i^{(2)}$ the number of patients in stage II.

Under the assumption of normally distributed responses the means of the observations for each dose level , say $\bar{Y}_0^{(1)}, \bar{Y}_1^{(1)}, \ldots, \bar{Y}_5^{(1)}$, are independent and $\bar{Y}_i^{(1)} \sim N(f(x_i), \sigma^2/n_i^{(1)}), i = 0, \ldots, 5$ (if the assumption of a normal distribution cannot be made, this statement is still correct in an asymptotic sense). We are interested in differences to placebo and therefore, we focus on the differences of the dose means with the placebo mean, i.e. $\bar{Y}_i^{(1)} - \bar{Y}_0^{(1)}, i = 1, \ldots, 5$.

These have a multivariate normal distribution, $N(\mu, \sigma^2\Sigma)$ with mean

$$\mu = (f(x_1) - f(x_0), \ldots, f(x_5) - f(x_0))^\top$$

and covariance matrix Σ which is specified in the appendix. Let $h_j(a_1, \ldots, a_5)$ be a common density of this multivariate distribution for Scenario j (note that the distribution depends on the scenario via the means $f(x_j) - f(x_0)$). Then, the updated probabilities can be computed according to

$$\pi_{j,\text{updated}} = \pi_j \cdot h_j(\bar{Y}_1^{(1)} - \bar{Y}_0^{(1)}, \ldots, \bar{Y}_5^{(1)} - \bar{Y}_0^{(1)}) / \left[\sum_{t=1}^{7} \pi_t \cdot h_t(\bar{Y}_1^{(1)} - \bar{Y}_0^{(1)}, \ldots, \bar{Y}_5^{(1)} - \bar{Y}_0^{(1)}) \right]$$

where $\pi_j, \pi_{j,\text{updated}}$ are the a priori and a posteriori probability for Scenario j, respectively.

We compute now numerically an optimal design using the probabilities $\pi_{j,\text{updated}}$ for Scenario $j = 1, \ldots, 7$. Since we have already patients from stage I and stage I overrun, the proportion $w_i^{(1)} = (n_i^{(1)} + n_i^{(1o)})/n_{\text{total}}$ of all patients has already been allocated to dose level x_i, $i = 0, \ldots, 5$. Therefore, computation of the optimal design means that we compute the optimal weights $w_i \in [w_i^{(1)}, 1]$ with $\sum_{i=0}^{5} w_i = 1$.

We compute then n_i by $w_i n_{\text{total}}$ and application of a rounding procedure (ensuring that still $n_i \geq n_i^{(1)} + n_i^{(1o)}$ after rounding). The number of patients for dose x_i in stage II is then $n_i^{(2)} = n_i - (n_i^{(1)} + n_i^{(1o)})$, $i = 0, \ldots, 5$. We illustrate this procedure, in particular the calculation of the a posteriori probabilities, with a simulated example.

Example 5.1 *Based on the results of $n^{(1)} = 100$ patients in stage I, see Table 4, a posteriori probabilities are calculated and shown in Table 5. Since the observed results are closer to the optimistic scenarios, the a posteriori probabilities are larger than the a priori probabilities for Scenario 2 and 3. For Scenario 1, 5 and 6 there are only small differences between the a priori and a posteriori probabilities, whereas for the other scenarios, the a posteriori probabilities are smaller than the a priori probabilities.*

The Bayesian optimal design for the a posteriori probabilities is calculated and the required number of patients per dose is shown in the last row in Table 6. The numbers of patients required for stage II are calculated by subtracting the number of patients in stage I and stage I overrun from the totally required patients, see Table 6. Because the patients already assigned to dose have to be taken into account, we have calculated the optimal design in the subset of designs which have at least weight $w_0^{(1)} = 0.193, w_1^{(1)} = 0.013, w_2^{(1)} = 0.010, w_3^{(1)} = 0.057, w_4^{(1)} = 0.053, w_5^{(1)} = 0.140$ for the doses 0, 20, 40, 60, 80 and 100mg, respectively. These weights correspond to the number of patients in stage I and stage I overrun for a dose divided by the total number of patients, $n_{total} = 300$.

Table 4: Simulated result for stage I ($n^{(1)} = 100$)

Dose (in mg)	0	20	40	60	80	100	Total
$n_i^{(1)}$	41	3	2	13	11	30	100
$\bar{Y}_i^{(1)} - \bar{Y}_0^{(1)}$	-	9.48	4.93	8.26	14.03	9.87	-

Table 5: Change of probabilities for simulated data

Scenario	1	2	3	4	5	6	7
a priori probability	0.30	0.05	0.05	0.20	0.05	0.15	0.20
a posteriori probability	0.29	0.28	0.20	0.01	0.05	0.12	0.06

Table 6: Final design for the trial for simulated stage I data

Dose (in mg)		0	20	40	60	80	100	Total
stage I	$n_i^{(1)}$	41	3	2	13	11	30	100
stage I overrun	$n_i^{(1o)}$	17	1	1	4	5	12	40
stage II	$n_i^{(2)}$	63	12	22	40	10	14	160
Total	n_i	121	16	25	57	26	56	300

In the described adaptive Bayesian optimal design, we perform an interim analysis with the aim to recalculate the allocation ratios for each dose. For the trial which motivated this research, further possibilities were discussed. If for all doses there is only an insufficient effect of the drug observable, the trial should be stopped for futility immediatly after interim analysis. Further, if we see in the interim analysis that the variance differs from our pre-trial anticipation, we could change the sample size n_{total} (sample-size re-estimation). However, although these are interesting possibilities, we focus here on recalculation of allocation ratios. Futility stop or sample-size re-estimation are not within the scope of this paper.

In contrast to the approach described in this section, which uses only the a posteriori probabilities for the determination of the design for stage II, there is it at least one alternative approach to obtain an adaptive design, which is briefly indicated here. From the available data of stage I the dose-effect curve could be estimated and these parameter estimates could serve as inital guess for the construction of a design for stage II. This design is determined optimising $\text{Eff}_1(\text{design}, \text{estimated scenario})$ if there exists a dose with effect δ between 0 and

$x_{\max}$ and $\mathrm{Eff}_2(\text{design, estimated scenario})$ otherwise. Although this approach is attractive on a first glance, it is necessary to point out that for the interim analysis only 100 observations are available. As a consequence there is still a rather large uncertainty in the estimates obtained from the interim analysis. Therefore, it seems to be more appropriate to construct the design for stage II also in a robust sense. We recommend not to optimise with the assumption that the observed scenario is the true one, but to express the still existing uncertainty using the a posteriori probabilities calculated from the interim analysis.

6   Comparison between non-adaptive and adaptive designs

In order to compare the non-adaptive with the adaptive Bayesian optimal design described in Section 5, we performed a simulation study. We have no asymptotic formula to compute an efficiency for the adaptive design as it is available for the non-adaptive design. Therefore, we compute mean squared errors (MSE) of the estimates in the simulations. More precisely, for a certain design and a certain simulation scenario, we obtain the MSE at dose x for estimation of $f(x) - f(0)$ by

$$\mathrm{MSE}(x, \text{design, scenario}) = \frac{1}{s} \sum_{l=1}^{s} \left(\hat{f}_l(x) - \hat{f}_l(0) - (f(x) - f(0)) \right)^2,$$

where s is the total number of simulations and $\hat{f}_l$ denotes the estimated function in the lth run of the simulation. Because the estimates are asymptotically unbiased, the MSE can also be approximated by the variance of $\hat{f}_l(x) - \hat{f}_l(0)$, which is proportional to $d(x, \text{design, scenario})$. Therefore we replace in the numerical calculation of Φ_c the function $d(x, \text{design, scenario})$ by the mean squared error $\mathrm{MSE}(x, \text{design, scenario})$ of the estimates for $f(x) - f(0)$. For each scenario, we performed 5000 simulations for the non-adaptive and the adaptive Bayesian optimal design. We focus on the important criteria, Criterion 1 and 2, and calculate the relative efficiency of the adaptive with respect to the non-adaptive Bayesian optimal design by

$$\frac{\Phi_c(\text{adaptive Bayesian optimal design, scenario})}{\Phi_c(\text{non-adaptive Bayesian optimal design, scenario})}.$$

A value > 1 means here that the the adaptive Bayesian optimal design has a better performance than the non-adaptive Bayesian optimal design.

The results are presented in Table 7 and show that in many cases the non-adaptive and the adaptive Bayesian optimal design are comparable. Under Criterion 1 the adaptive Bayesian optimal design performs better in the optimistic Scenario 2 and slightly better for scenorios 3 and 6, while the non-adaptive Bayesian design has slight advantages in Scenario 1, 5 and

Table 7: Simulated relative efficiencies of the adaptive Bayesian optimal design compared to the non-adaptive Bayesian optimal design (based on MSE calculation)

		Criterion	
Scenario		$c = 1$	$c = 2$
1	Prior guess	0.96	0.95
2	High $E_{\max}$	1.25	0.82
3	Low ED_{50}	1.08	0.89
4	High ED_{50}	-	0.98
5	Intermed α	0.95	0.91
6	High α	1.05	0.82
7	Low $E_{\max}, ED_{50}$	0.94	0.97
Overall		1.00	0.93

7. On the other hand, under Criterion 2, the non-adaptive design yields always a smaller MSE than the adaptive design. While these advantages are less substantial in Scenario 4 and 7, the differences in the other scenarios vary between 9% and 18%. Overall, assuming the given a priori distribution and observing that Criterion 1 reflects the more important objective, the two designs have a comparable efficiency (relative efficiency=1.00 in Table 7).

Generally, when using design adaptations during the conduct of a trial, it is not clear if statistical properties of e.g. estimates are affected from adaptation. Therefore, we show the empirical distribution function for the effect estimates at dose 100 *mg* (difference to placebo effect) of our simulations for Scenario 1 for the non-adaptive (Figure 3) and the adaptive Bayesian optimal design (Figure 4), based on 5000 simulation runs. The empirical distribution functions are compared in the figures with the cumulative distribution function of the normal distribution (dotted line) with the same mean and variance as the simulated data. For the non-adaptive design, the empirical distribution function cannot be distinguished from the normal cummulative distribution function in the figure. For the adaptive design, we observe a slight deviation. For values above the 0.9-quantile, the simulated results are larger than the values from the fitted normal distribution.

In summary we did not observe a convincing advantage of the adaptive over the non-adaptive Bayesian optimal design in terms of efficiency. On the other hand, an adaptive design is more challenging from an operational point of view. Moreover, there is a slight effect on the distributional properties of the estimates caused by the adaptation, which can
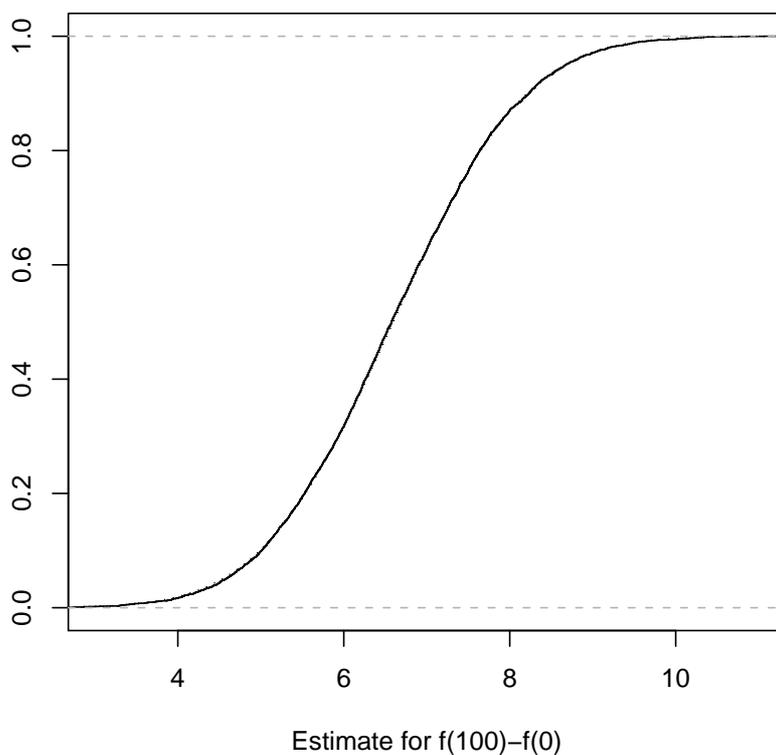
Figure 3: Simulated distribution function of the estimate of the effect at the maximum tolerable dose $x_{\max}$ based on the non-adaptive Bayesian optimal design.

be avoided with a non-adaptive design. Therefore, we recommended to use the non-adaptive Bayesian optimal design in this situation.

On the other hand the deviations from the normal distribution caused by adaptation are not too serious. So if we would have seen a good efficiency gain with the adaptive design, we probably would have accepted this fact. The same is true for the operational aspects: if there would have been justification from an efficiency point of view, it would have been worth to conduct an adaptive design at the cost of different distributional properties of the resulting estimates.

If an interim analysis is required because of ethical or other reasons, this could also be performed on the basis of the non-adaptive Bayesian optimal design without changing the design for the second stage.
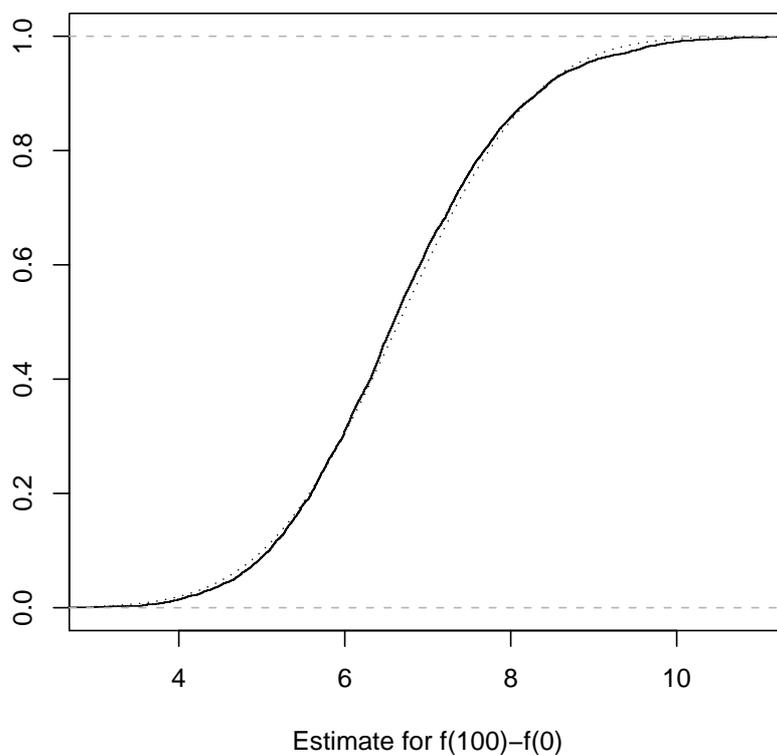
Figure 4: Simulated distribution function of the estimate of the effect at the maximum tolerable dose $x_{\max}$ based on the adaptive Bayesian optimal design.

7   Concluding remarks

An advantage when thinking about possible innovative or adaptive designs is that one is more forced to specify the knowledge available before conduction experiments and the objectives of the trial. We could propose and specify in statistical terms the objective to get good information about the interesting part of the dose-effect curve. This objective is connected to the desires of the project team and the sponsor but is hopefully also appreciated by physicians and patients after marketing.

However, it is essential to discuss the statistical criteria extensively with other experts involved in the trial to be really sure that the specified objectives reflect the underlying aim of the trial appropriately. It is important that all objecives have been taken into account and that there are no implicit additional objectives which the clinical team has in mind but

which are hard to express in common terms. If essential objectives have not been considered in the construction of an statistical criterion, one would optimise the design on a wrong basis and the resulting design coud be efficient for the specified objective but very inefficient for the implicit objective.

It is remarkable and was somewhat unexpected that we have seen an impressive improvement in efficiency when a balanced design is replaced by a (non-adaptive) Bayesian optimal design. This underlines the usefulness of optimal design theory for clinical trial applications. In contrast, we have not observed remarkable differences between the non-adaptive and the adaptive Bayesian optimal design. In the trial under consideration the non-adaptive design had a better performance in many cases. Advantages of non-adaptive designs in context of model discrimination have also been observed by other authors [see Dette and Kwiecien (2004), for example]. Roughly speaking the answer to the question, whether or not an adaptive design has advantages over an optimised non-adaptive design depends mainly on the correct specification of the a priori knowledge. If this a priori information has been completely misspecified, the application of adaptive designs may have advantages. Consequently, if there is absolutely no knowledge about the dose-effect curve available in advance, either a separate pilot trial has to be performed or the first part in the trial itself (stage I) has to be used to derive the final design of the trial. On the other hand, in phase IIB trials, there is often a priori information available and the application of non-adaptive optimal designs is well justified.

In the trial considered in this paper the anticipated scenarios differed not too much and the more extreme scenarios were judged as more unlikely. Therefore the uncertainty was – relatively – low which gives a partial explanation that we are in a situation where we have no gain from changing allocation ratios during the trial.

8 Appendix

The function d(x,design,scenario)

Let ϑ be the vector of parameters in the $E_{\max}$-sigmoid model, i.e. $\vartheta = (E_0, E_{\max}, ED_{50}, \alpha)^\top$, which characterises a specific scenario. Let $f(x, \vartheta) = f(x)$ be the true dose-effect curve and σ^2 be the variance of the observations. We obtain from Seber and Wild (1989) for the nonlinear least squares estimates $(\hat\vartheta, \hat\sigma^2)$ for (ϑ, σ^2):

$$\sqrt{n}\left\{ \begin{pmatrix} \hat\vartheta \\ \hat\sigma^2 \end{pmatrix} - \begin{pmatrix} \vartheta \\ \sigma^2 \end{pmatrix} \right\} \xrightarrow{\mathcal{D}} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 M^{-1}(\text{design}, \vartheta) & 0 \\ 0 & 2\sigma^4 \end{pmatrix} \right),$$

where n is the sample size, $\xrightarrow{\mathcal{D}}$ denotes the convergence in distribution for $n \to \infty$, $\mathcal{N}(\mu, \Sigma)$ is a two-dimensional normal distribution,

$$M(\text{design}, \vartheta) = \sum_{j=1}^{k} w_j g(x_j, \vartheta) g^\top(x_j, \vartheta)$$

is the information matrix [see Silvey (1980)] and

$$g(x, \vartheta) = \left(\frac{\partial f(x, \vartheta)}{\partial E_0}, \frac{\partial f(x, \vartheta)}{\partial E_{\max}}, \frac{\partial f(x, \vartheta)}{\partial ED_{50}}, \frac{\partial f(x, \vartheta)}{\partial \alpha} \right)^\top$$
$$= \left(1, \frac{x^\alpha}{ED_{50}^\alpha + x^\alpha}, \frac{-E_{\max}\alpha ED_{50}^{\alpha-1} x^\alpha}{(ED_{50}^\alpha + x^\alpha)^2}, \frac{E_{\max} ED_{50}^\alpha x^\alpha (\log x - \log ED_{50})}{(ED_{50}^\alpha + x^\alpha)^2} \right)^\top$$

is the gradient of the dose-effect-function with respect to ϑ.

The variance of the estimated effect of a certain dose x is approximately proportional to

$$g^\top(x, \vartheta) M^{-1}(\text{design}, \vartheta) g(x, \vartheta).$$

The variance of the estimated difference in the effect between a dose x and placebo (dose 0) is approximately proportional to

$$d(x, \text{design}, \vartheta) = (g(x, \vartheta) - g(0, \vartheta))^\top M^{-1}(\text{design}, \vartheta) (g(x, \vartheta) - g(0, \vartheta)).$$

Covariance matrix for differences of the dose means with the placebo mean in the interim analysis

The differences of the dose means with the placebo mean in the interim analysis, $\bar{Y}_i^{(1)} - \bar{Y}_0^{(1)}, i = 1, \ldots, 5$ have a multivariate normal distribution, $\mathcal{N}(\mu, \sigma^2\Sigma)$ with $\mu = (f(x_1) - f(x_0), \ldots, f(x_5) - f(x_0))^\top$ and covariance matrix

$$\Sigma = \begin{pmatrix} 1/n_0^{(1)} + 1/n_1^{(1)} & 1/n_0^{(1)} & \cdots & 1/n_0^{(1)} \\ 1/n_0^{(1)} & 1/n_0^{(1)} + 1/n_2^{(1)} & & 1/n_0^{(1)} \\ \vdots & & \ddots & \\ 1/n_0^{(1)} & \cdots & 1/n_0^{(1)} & 1/n_0^{(1)} + 1/n_5^{(1)} \end{pmatrix}.$$

References

Bezeau, M., Endrenyi, L. (1986). Design of experiments for the precise estimation of dose-response parameters: the Hill equation. *J. Theor. Biol.* 123:415-430.

Bretz, F., Pinheiro, J., Branson, M. (2005). Combining multiple comparisons and modeling techniques in dose-response studies. *Biometrics* 61:738-748.

Chevret, S. (2006). *Statistical methods for dose-finding experiments.* Chichester: Wiley.

Committee for Medicinal Products for Human Use (2006). Reflection paper on methodological issues in confirmatory Clinical trials with flexible design and analysis plans. Draft. London: European Medicines Agency.

Dette, H., Bretz, F., Pepelyshev, A., Pinheiro, J. (2007). Optimal designs for dose finding studies. Technical report, Ruhr-Universität Bochum. Available at: http://www.ruhr-uni-bochum.de/mathematik3/quellen/draft14.pdf

Dette, H., Kwiecien, R. (2004). A comparison of sequential and non-sequential designs for discrimination between nested regression models. *Biometrika* 91:165-176.

Dette, H., Haines, L.M., Imhof, L. (2005) Bayesian and maximin optimal designs for heteroscedastic regression models. *Canadian Journal of Statistics* 33:221-241.

Dette, H., Melas, V.B., Wong, W.K. (2005). Optimal design for goodness-of-fit of the MichaelisMenten enzyme kinetic function. *Journal of the American Statistical Association* 100:1370-1381.

Dette, H., O'Brien, T.E. (1999). Optimality criteria for regression models based on predicted variance. Biometrika 86:93-106.

Dragalin, V. (2006). Adaptive designs: terminology and classification. *Drug Inf. J.* 40:425-435.

Dutta, S., Matsumoto, Y., Ebling, W.F. (1996). Is it possible to estimate the parameters of the sigmoid E_{max} model with truncated data typical of clinical studies? *Journal of Pharmaceutical Sciences* 85:232-239.

Gaydos, B., Krams, M., Perevozskaya, I., Bretz, F., Liu, Q., Gallo, P., Berry, D., Chuang-Stein, C., Pinheiro, J., Bedding, A. (2006). Adaptive dose-response studies. *Drug Inf. J.* 40:451-461.

Holford, N.H.G., Sheiner, L.B. (1981). Understanding the dose-effect relationship: clinical application of pharmacokinetic - pharmacodynamic models. *Clin. Pharmacokinet.* 6:429-453.

Krams, M., Lees, K.R., Hacke, W., Grieve, A.P., Orgogozo, J.M., Ford, G.A. (2003). Acute

stroke therapy by inhibition of neutrophils (ASTIN). *Stroke* 34:2543-2548.

Montepiedra, G., Yeh, A.B. (1998). A two stage strategy for the construction of *D*-optimal experimental designs. *Commun. in Statistics, Simul. and Comput.* 27:377-402.

Montepiedra, G., Yeh, A.B. (2003). Two-stage designs for identification and estimation of polynomial models. *Computational Statistics & Data Analysis* 46:531-546.

Pukelsheim, F., Rieder, S. (1992). Efficient rounding of approximate designs. *Biometrika* 79: 763-770.

Quinlan, J.A., Krams, M. (2006). Implementating adaptive designs: logistical and operational considerations. *Drug Inf. J.* 40:437-444.

Seber, G.A.F., Wild, C.J. (1989). *Nonlinear regression.* New York: Wiley.

Silvey, S. D. (1980). *Optimal design.* London: Chapman and Hall.

Smith, M.K., Jones, I., Morris, M.F., Grieve A.P., Tan, K. (2006). Implementation of a Bayesian adaptive design in a proof of concept study. *Pharmaceut. Statist.* 5:39-50.

Ting, N. (2006). *Dose finding in drug development.* Springer.

Van der Vaart, A. (1998). *Asymptotic statistics.* Cambridge: University Press.

Zhu, W., Wong, W.K. (2000). Multiple-objective designs in a dose-response experiment. *Journal of Biopharmaceutical Statistics*, 10:1-14.