CrossMark

ORIGINAL PAPER

Experiment design for nonparametric models based on minimizing Bayes Risk: application to voriconazole[1]

David S. Bayard[1] · Michael Neely[1,2]

Abstract An experimental design approach is presented for individualized therapy in the special case where the prior information is specified by a nonparametric (NP) population model. Here, a NP model refers to a discrete probability model characterized by a finite set of support points and their associated weights. An important question arises as to how to best design experiments for this type of model. Many experimental design methods are based on Fisher information or other approaches originally developed for parametric models. While such approaches have been used with some success across various applications, it is interesting to note that they largely fail to address the fundamentally discrete nature of the NP model. Specifically, the problem of identifying an individual from a NP prior is more naturally treated as a problem of classification, i.e., to find a support point that best matches the patient's behavior. This paper studies the discrete nature of the NP experiment design problem from a classification point of view. Several new insights are provided including the use of Bayes Risk as an information measure, and new alternative methods for experiment design. One particular method, denoted as MMopt (multiple-model optimal), will be examined in detail and shown to require minimal computation while having distinct advantages compared to existing approaches. Several simulated examples, including a case study involving oral voriconazole in children, are given to demonstrate the usefulness of MMopt in pharmacokinetics applications.

Keywords Bayes Risk · Experiment design · Nonparametric · Population model · Voriconazole

✉ Michael Neely
mneely@chla.usc.edu

1 Laboratory of Applied Pharmacokinetics and Bioinformatics, The Saban Research Institute, Children's Hospital of Los Angeles, Los Angeles, CA, USA

2 Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

Introduction

Nonparametric (NP) population models are specified by a discrete parameter probability distribution. The richness of the NP approach is in its ability to not simply estimate central tendency and dispersion, but rather to estimate the entire population parameter joint density. NP models arise naturally when applying maximum likelihood (ML) estimation to population modeling problems due to a remarkable fact that the optimal ML estimate is known to be discrete [34, 36]. NP models also arise in pharmacokinetic (PK) applications when computations are simplified by making discrete approximations to continuous distributions [28]. Software programs for generating NP PK population models are readily available, e.g., NPEM [16, 46], NPAG [3, 32], and USC*PACK [26], which are now all available within the Pmetrics package for R [10]. The use of NP models for design of PK dosage regimens is discussed in [6].

An important question arises as to how to best design experiments with respect to NP population model priors. Ideally, the experiment should be designed to extract the most information using the least number of samples. A widely used method for experiment design is to optimize the D-optimality criterion introduced by Box and Lucas [12],

which consists of maximizing the determinant of the Fisher information matrix M [14, 19, 50]. One criticism of D-optimal (Dopt) design is that it only uses the mean of the population model and ignores other distributional information. An improved class of approaches incorporating the full prior distribution are the so-called "robust" sampling designs, which more generally optimize the expectation of a specified scalar function of M [15, 41, 52, 57]. Choices for this scalar function include the determinant of M (giving rise to ED-optimal design), the determinant of the matrix inverse of M (giving rise to EID-optimal design), and the log of the determinant of M (giving rise to ELD design). The ELD design is sometimes called the API design. Algorithms and software for computing the ED, EID and ELD designs are described in Pronzato and Walter [41], Tod and Rocchisani [51, 52] and Walter and Pronzato [57].

While such Dopt approaches have been used with success across various applications, it is interesting to note that they largely fail to address the fundamentally discrete nature of the NP problem. Specifically, the problem of identifying an individual from a NP prior is more naturally treated as a problem of classification, i.e., of finding the support point that best matches the patient's behavior. Looking to the classification literature for inspiration, a classifier's performance is most often scored in terms of how well it minimizes the Bayes Risk. Bayes Risk is defined as the probability that a maximum a posteriori (MAP) estimator will misclassify a subject [18]. Hence, an experiment design minimizing Bayes Risk will ensure that the MAP estimate (which is one of the support points), is representative of the true patient with high probability. This is especially important for PK applications in light of the work of Sheiner et al. [47–49] and others [24, 44, 55, 56], who have demonstrated the value of MAP Bayesian control, i.e., where the MAP estimate is used as the basis for determining subsequent dosing regimens.

Interestingly, the authors are not aware of a single instance where Bayes Risk has been used as an information measure for experiment design purposes in a PK context. It is expected that this is because its computation is generally unwieldy. As pointed out in Fedorov (cf., [19], p. 233), in the context of evaluating Bayes Risk for model discrimination "...even for comparatively simple response surfaces, it is necessary to deal with serious computation difficulties, which make construction of the optimal design practically impossible". The current paper makes several strides in this direction. Specifically, use is made of a theoretical upper bound on the Bayes Risk that has recently appeared in the literature in Blackmore et al. [11]. The upper bound is important because it can be computed completely algebraically without requiring multidimensional integration or stochastic analysis. Because of its computational advantages, this paper defines multiple-model optimal (MMopt) as the experiment design approach that minimizes this Bayes

Risk upper bound. It will be shown that MMopt provides a flexible and effective method for designing experiments for NP models across a wide range of PK applications.

Background on experiment design is given in "Background" section, with discussion of D-optimality, the ED, EID, and ELD robust designs, and the globally optimal Bayes optimal classifier (Bopt). In "Multiple model optimal (MMopt) design" section, the Bayes Risk upper bound of [11] is introduced. MMopt is defined as an experiment design approach that minimizes this upper bound. In "Two-support-point example" section, a simple two-support-point example is used to compare performance between the various designs. Sampling designs are applied in "Four-support-point example" section to a more complicated four-support-point example. The MMopt experiment design approach is demonstrated on a simulated PK example in "Pharmacokinetic example" section, and its usefulness in a voriconazole case study is demonstrated in "Voriconazole case study" section. "Discussion" section summarizes the properties of MMopt relative to Bopt, ED, EID, and ELD. Appendices and a Supplemental File are included that provide detailed analysis on certain aspects of the various designs. Conclusions are postponed until "Conclusions" section.

This paper focuses on sampling times, but MMopt can be easily adapted to optimize over any other aspects of experiment design (i.e., dose timing, dose amounts, etc.)

Background

Dynamic model and measurements

A nonlinear dynamic model is defined of the form,

$$\dot{x}(t) = f(x(t), d(t), \theta), \tag{1}$$

where x is the state, d is the input, and $\theta \in R^p$ is the parameter vector, and f is a function of these three quantities. The system output at time t_k is given as,

$$\mu_k = h(x(t_k), \theta), \tag{2}$$

where the observed function h depends on both the state x and parameter vector θ. A noisy measurement of μ_k is taken of the form

$$y_k = \mu_k + \sigma_k n_k, \tag{3}$$

where $n_k \sim N(0, 1)$ is the measurement noise, and σ_k scales n_k to its desired level. The experiment design problem is formulated to find the n optimal sampling times

$$U = \{t_1, \ldots, t_n\}, \tag{4}$$

that act to best identify the parameter vector θ.

The parameter vector θ can have either continuous or discrete values. When applied to NP problems, θ is

assumed to belong to a discrete set $\theta \in \Theta \stackrel{\Delta}{=} \{\theta_1, \ldots, \theta_m\}$, having Bayesian prior probabilities $p(\theta_i)$, $i = 1, \ldots, m$. The true value of θ is denoted as θ^*. A slight generalization of (3) is to allow σ_k to be a function of μ_k according to an "assay polynomial" of the form,

$$\sigma_k(\mu_k) = c_0 + c_1\mu_k + c_2\mu_k^2 + c_3\mu_k^3. \tag{5}$$

Here the coefficients $c_0, \ldots, c_3$ are typically determined from a separate laboratory assay error calibration.

D-optimal design

The D-optimal experiment design (denoted here as Dopt), is defined by maximizing the determinant of the Fisher information matrix [14, 19, 50].

D-optimal design

$$U_D \stackrel{\Delta}{=} \arg\max_U |M|, \tag{6}$$

where $|(\cdot)|$ denotes the matrix determinant, and the Fisher information matrix M is given by,

$$M(\theta, U) = \sum_{k=1}^{n} \frac{1}{\sigma_k^2} \left[\frac{\partial \mu_k}{\partial \theta} \frac{\partial \mu_k}{\partial \theta^T} \right]\bigg|_{\theta=\theta^*}. \tag{7}$$

Dopt designs have the property that they minimize the volume of confidence ellipsoids for the parameters. Dopt design has become a useful tool for optimizing experiments over a wide range of applications [20, 30].

The partial of μ in (7) is evaluated on the true parameter value θ^*. This is cause for some concern, since the true parameter value θ^* is not known, and in fact is the main reason for doing experiment design in the first place. However, for problems where $\mu_k = h(x(t_k), \theta)$ is linear in θ,

$$\mu_k = h(x(t_k))^T \theta, \tag{8}$$

where h is a column vector, the partial becomes,

$$\frac{\partial \mu_k}{\partial \theta} = h(x(t_k)). \tag{9}$$

In this case the dependence on θ vanishes, and the problem goes away. Hence, Dopt is best suited to problems for which μ_k has the form (8) which is a linear function of θ. Furthermore, when the Fisher matrix M is singular its determinant is zero. This implies that Dopt sampling design is only defined when at least p samples are taken, where p equals the number of parameters.

Robust D-optimal designs

A drawback to using Dopt designs with nonlinear models of the form (1)–(3), is that μ is nonlinear in θ. This means

that the Dopt design becomes a function of the unknown true parameter vector θ^*, which presents a problem since it is not known beforehand. A more principled approach is to use robust Dopt designs which optimize the expectation of a cost function, where the expectation is taken over the random variable θ, and where the cost function is a specified as a scalar function of the Dopt cost. The most common robust Dopt designs are based on the ED, EID, ELD experiment design criteria defined as (cf., [13, 15, 41, 52, 57]),

ED optimal design

$$\arg\max_\xi E_\theta(|M|). \tag{10}$$

EID optimal design

$$\arg\min_\xi E_\theta\left(\frac{1}{|M|}\right). \tag{11}$$

ELD optimal design

$$\arg\max_\xi E_\theta(\log|M|). \tag{12}$$

Again, since the Fisher matrix determinant is only non-zero when the Fisher matrix M is nonsingular, these robust designs have the same restriction as Dopt designs in that they must contain at least p samples, where p equals the number of parameters.

Other robust Dopt approaches are possible. One variation is to add the precision matrix Ω^{-1} to M in any of the above expressions, where Ω is the prior covariance matrix of θ. Such modifications are discussed in [13]. Another variation is to reverse the order of the determinant and expectation in the ED design (10) to give $|E_\theta(M)|$. Combining both of these latter two modifications gives the design criterion $|E_\theta(M) + \Omega^{-1}|$ which has been applied to PK problems in Hennig et al. [23] and Merle and Mentre [37].

In this paper, MMopt experiment designs will be compared to ED, EID, and ELD optimal robust designs under the assumption of an NP prior. This comparison is not intended to be exhaustive as other experiment design methods are possible. Comparisons with methods such as D_s optimal design [22], T-optimal design [2, 54], alternative robust Dopt designs [23, 37], and Bayesian design [13, 43], are left for future studies.

Bayes optimal classifier

Given a vector of observations $Y \in \mathcal{R}$ and a vector of control inputs $U \in \mathcal{U}$, the posterior probability $p(H_i|Y, U)$ of the ith hypothesis H_i in an m-category classification problem is calculated using Bayes rule as [29],

$$p(H_i|Y, U) = \frac{p(Y|H_i, U)p(H_i)}{p(Y|U)}, \tag{13}$$

where H_i, $i = 1, \ldots, m$ represent the m individual hypotheses, and $p(H_i)$, $i = 1, \ldots, m$ represents their corresponding prior probabilities.

Remark 1 When applied to discrete priors, H_ℓ is understood to be the hypothesis that the true value of θ corresponds to θ_ℓ. Bayesian prior support point probabilities are denoted as $p_i = p(\theta_i)$, $i = 1, \ldots, m$.

After computing the posterior probabilities from (13), the most probable hypothesis is defined by H_i where,

$$H_i = \arg\max_j p(H_j|Y, U). \tag{14}$$

This selection rule is denoted as the **Bayes optimal classifier**. Since the Bopt classifier maximizes the posterior probability, it is also a MAP estimator.

Since (14) is only an estimate, it will sometimes be correct and sometimes be incorrect. The probability of it being incorrect is denoted as the **Bayes Risk**. For a given U, Eq. (14) can be thought of as a prescription for dividing up the domain $\mathcal{R}$ into at most m regions defined by the following relation,

$$\mathcal{R}_i \overset{\Delta}{=} \{Y | p(H_i|Y, U) > p(H_\ell|Y, U) \text{ for all } \ell \neq i\}. \tag{15}$$

This definition allows one to state the Bopt classifier (14) simply as an algorithm that selects model H_i as the MAP estimate if and only if Y falls into region $\mathcal{R}_i$.

Multiplying both sides of the inequality in (15) by $p(Y|U) > 0$ and rearranging allows region $\mathcal{R}_i$ to be equivalently defined as,

$$\mathcal{R}_i \overset{\Delta}{=} \{Y | p(H_i, Y|U) > p(H_\ell, Y|U) \text{ for all } \ell \neq i\}. \tag{16}$$

The Bayes Risk is calculated by summing over all possible ways that a given classifier can make a classification error, i.e., by summing over the probabilities of selecting H_j when in fact H_i is true,

$$p(error) = \sum_{i=1}^{m} \sum_{j \neq i} p(Y \in \mathcal{R}_j, H_i|U), \tag{17}$$

$$= \sum_{i=1}^{m} \sum_{j \neq i} p(Y \in \mathcal{R}_j|H_i, U)p(H_i), \tag{18}$$

$$= \sum_{i=1}^{m} \sum_{j \neq i} \int_{\mathcal{R}_j} p(Y|H_i, U)p(H_i)dY. \tag{19}$$

The experiment design that minimizes the Bayes Risk (19) is defined as the Bopt design, or Bopt for short,

Bayes optimal design (Bopt)

$$U_B \overset{\Delta}{=} \arg\min_U p(error). \tag{20}$$

Multiple model optimal (MMopt) design

Bayes Risk overbound

Computing the Bayes Risk is generally a very difficult problem. However, a recent overbound on the Bayes Risk has appeared in the literature that is relevant to the special case where the likelihood functions are Gaussian, i.e.,

$$p(Y|H_i, U) = N(\mu(i), \Sigma(i)), \quad i = 1, \ldots, m. \tag{21}$$

The overbound is given in the following result.

Theorem 1 (Blackmore et al. [11]) *When performing hypothesis selection between m hypotheses, for Gaussian observation distributions such that $p(Y|H_i) = N(\mu(i), \Sigma(i))$, $i = 1, \ldots, m$, the Bayes Risk is upper bounded as follows,*

$$p(error) \leq \sum_{i=1}^{m} \sum_{j > i} p(H_i)^{\frac{1}{2}} p(H_j)^{\frac{1}{2}} e^{-\kappa(i,j)}, \tag{22}$$

where,

$$\kappa(i, j) = \frac{1}{4} (\mu(j) - \mu(i))^T [\Sigma(i) + \Sigma(j)]^{-1} (\mu(j) - \mu(i))$$
$$+ \frac{1}{2} \ln \frac{\left|\frac{1}{2}\Sigma(i) + \frac{1}{2}\Sigma(j)\right|}{|\Sigma(i)|^{\frac{1}{2}}|\Sigma(j)|^{\frac{1}{2}}}. \tag{23}$$

Proof See Blackmore et al. [11]. □

MMopt design

Theorem 1 is important because the expression (22) overbounds the Bayes Risk and hence represents a form of pre-posterior analysis [42]. However, in contrast to many of the pre-posterior analysis expressions found in modern Bayesian optimal design (cf., [43]), expressions (22), (23) can be calculated analytically without requiring multidimensional integration or stochastic analysis. For the purpose of this paper, a design that minimizes the Bayes Risk overbound (22) is denoted as the multiple model optimal experiment design (MMopt). Specifically,

MM optimal design

$$U_{MM} = \arg\min_U \sum_{i=1}^{m} \sum_{j > i} p(H_i)^{\frac{1}{2}} p(H_j)^{\frac{1}{2}} e^{-\kappa(i,j)}. \tag{24}$$

MMopt designs can be computed for many problems of practical interest. The overbound was first developed and used for active estimation in Blackmore et al. [11]. These ideas were later extended by the present authors in a series of workshops on optimal experiment design [7, 8]. Since Bayes Risk is always non-negative, the minimization of an

overbound on the Bayes Risk will intuitively act to drive it toward lower values, as desired.

The minimization of Bayes Risk acts to make the MAP estimate a better predictor (i.e., classifier), of the true patient. For example, if the Bayes Risk is reduced to 0.1 by proper experiment design, it means that the MAP estimate (which itself is always a support point in the discrete model) will correspond to the true patient with probability 0.9. This interpretation can be very useful in PK applications where the MAP estimate is used to guide subsequent dosing and/or therapeutic recommendations. Currently, a large number of successful dosing and therapeutic drug monitoring approaches are based on MAP Bayesian control (cf., [44, 47–49, 55, 56]). The efficacy of such methods is directly based on the accuracy of the MAP estimate. Since MMopt acts to minimize the Bayes Risk, it improves the discriminating power of the MAP Bayesian estimate, which in turn, has the potential advantage of improving control performance across a wide range of PK applications.

Two-support-point example

A simple two-support-point example is given to demonstrate the difference between treating an experiment design problem as a parameter estimation problem (i.e., using D, ED, EID, ELD designs) versus treating it as a classification problem (i.e., using MMopt).

System model

A two-support-point example involving an exponential measurement model is defined as

$$y_k = \mu(t_k, a) + \sigma n_k, \tag{25}$$

$$\mu(t, a) = e^{-at}. \tag{26}$$

The Fisher information matrix M for a single sample time is computed as,

$$\frac{\partial \mu}{\partial a} = -t e^{-at}, \tag{27}$$

$$M(a) = \frac{1}{\sigma^2} \left(\frac{\partial \mu}{\partial a} \right)^2 = \frac{1}{\sigma^2} (t e^{-at})^2, \tag{28}$$

and the Dopt cost follows as

$$|M(a)| = t^2 e^{-2at} / \sigma^2. \tag{29}$$

Values for parameter a are given as

Model 1 : $a_1 = 1.5$ (fast support point), (30)

Model 2 : $a_2 = 0.25$ (slow support point), (31)

with $\sigma = 0.3$, and with the prior specified as $p_1 = 0.5$ and $p_2 = 0.5$. The two support point responses are plotted in Fig. 1. For this example, all optimizations are performed
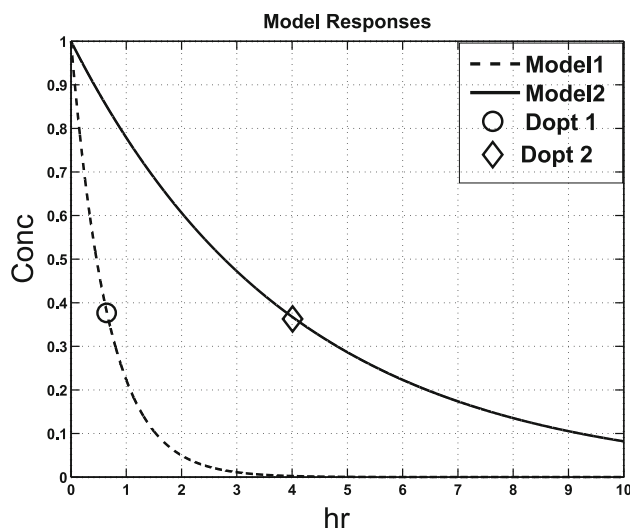

Fig. 1 Responses for the two-support-point example. The Dopt design is indicated for each support point taken separately

over a discrete time grid spanning 0–10 h, with a $\Delta t = 0.0001$ h spacing.

Designs for two-support point example

It is known that the Dopt design for an exponential model of the form (25), (26) is given by $t = 1/a$ [31], i.e., at one time constant of the model response, or equivalently where the response has dropped to $e^{-1} = 0.3679$ of its initial value. Hence, minimizing (29) over t gives the individual Dopt designs as,

$$t_D(\text{fast}) = 1/a_1 = 0.6667, \tag{32}$$

$$t_D(\text{slow}) = 1/a_2 = 4. \tag{33}$$

These two Dopt designs are plotted with a circle and diamond, respectively, in Fig. 1, and indicate the time constants for each response. Clearly, the optimal sample times for the two support points are very different (i.e., 0.6667 vs. 4). This exposes a dilemma when using the Dopt design. What is good for one support point is very different from what is good for the other support point.

ED-optimal design

One method to resolve the above dilemma is to use a robust design approach such as ED design. As discussed earlier, ED design makes use of prior information, and is defined by maximizing over the expected value of $|M|$,

$$t_{ED} = \arg\max_t E[|M|] = \arg\max_t 0.5|M(a_1)| + 0.5|M(a_2)|. \tag{34}$$

Here $|M|$ is given by (29). Optimization of (34) for this example gives the ED design as

$$t_{ED} = 3.9991. \tag{35}$$

The ED cost (34) is a weighted combination of the individual Dopt costs given earlier in (29), with weights of (0.5, 0.5) corresponding to the prior probabilities (p_1, p_2). The situation is depicted in Fig. 2. The ED-optimal sampling design of $t = 3.9991$, represents a compromise between the Dopt values of $t = 0.6667$ and 4. Here, the weighting strongly favors the term $|M(a_2)|$ which is associated with the slow a_2 support point.

Bayes optimal design

Computation of the Bopt design is usually unwieldy. Fortunately, the current example is simple enough to allow its analytic computation.

Responses for the two model support points (30), (31) are shown in Fig. 3, where the response separation is defined by $r(t) = |\mu(t, a_1) - \mu(t, a_2)|$. Since there is additive Gaussian noise associated with each measured response, the Bayes Risk (19) can be represented as the grey area in Fig. 4. Here, Fig. 4 can be thought of as looking at Fig. 3 "sideways", such that the two Gaussian "bumps" represent the two hypotheses associated with the true patient being Model 1 or 2. The Bopt design minimizes this Bayes Risk (i.e., grey area), which can be calculated analytically as

$$\text{Bayes Risk} = \text{Prob}(\text{Type I error}) + \text{Prob}(\text{Type II error}), \tag{36}$$
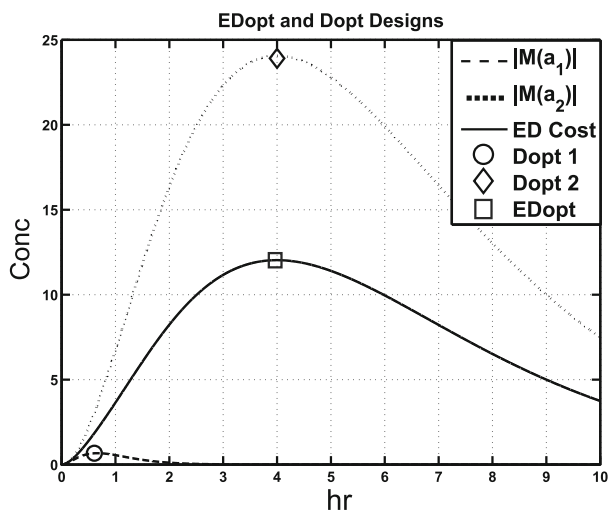

Fig. 2 The ED measure (*solid*) is equal to one half the Dopt measure for the slow a_2 support point $|M(a_2)|$ (*dot*), and one half the Dopt measure for the fast a_1 support point $|M(a_1)|$ (*dash*). Since the slow support point measure dominates the fast support point, the ED sample at $t = 3.9991$ is nearly identical to that of the slow a_2 model's Dopt sample at $t = 4$, but has been drawn slightly lower by the weighted influence of the fast a_1 model's Dopt sample at $t = 0.6667$
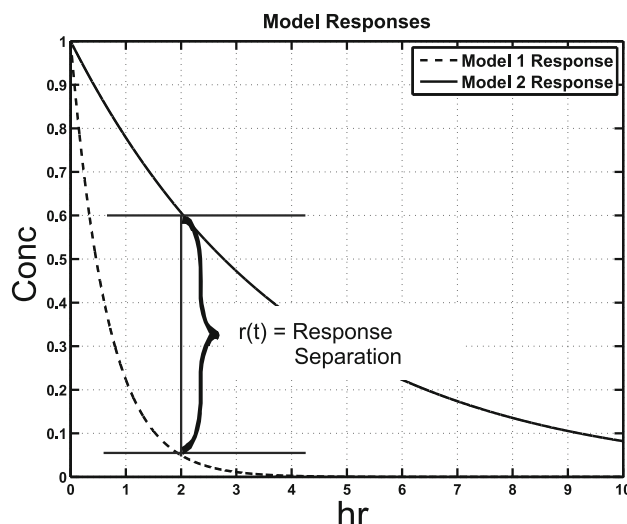
Fig. 3 Depiction of response separation $r(t)$ between responses $\mu(t, a_1)$ and $\mu(t, a_2)$
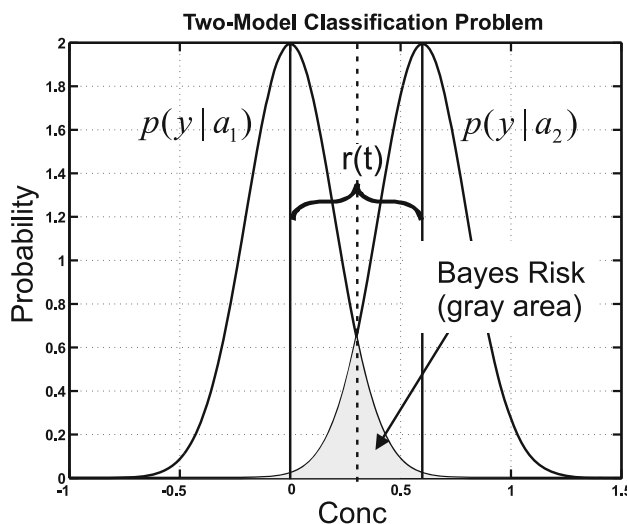

Fig. 4 Bayes Risk is depicted as the *grey area* between curves, corresponding to the sum of Types I and II errors. Bayes Risk decreases as response separation $r(t)$ increases, which "pulls apart" the two Gaussian distributions and shrinks the *grey area* between them

$$= p_1 \int_{y_{cross}}^{\infty} p(y|a_1) dy + p_2 \int_{-\infty}^{y_{cross}} p(y|a_2) dy$$

$$= 1 - \int_{-\infty}^{r(t)/2} N(0, \sigma^2) dy, \tag{37}$$

where y_{cross} is where the two Gaussians intersect in Fig. 4. Intuitively, this calculation is simply adding the standard Types I and II error probabilities that arise in a classical two-hypothesis test (i.e., defined as if we were treating a_1 as the null hypothesis). The response separation $r(t)$ is an important statistic in the two-support-point example.

Equation (37) indicates that the Bayes Risk decreases monotonically with increasing $r(t)$. This is seen graphically in Fig. 4 where the sum of the two grey areas decreases monotonically as the two distributions are "pulled apart". Hence the Bayes Risk is minimized at the time of maximal response separation $r(t)$ given by,

$$t_B = \arg\max_t r(t) = \arg\max_t |\mu(t, a_1) - \mu(t, a_2)| = 1.4334.$$

(38)

MMopt design

The MMopt design is calculated by minimizing the Bayes Risk overbound (22) as

$$t_{MM} = \arg\min_t p_1^{\frac{1}{2}} p_2^{\frac{1}{2}} e^{-\kappa(1,2)} = \arg\min_t p_1^{\frac{1}{2}} p_2^{\frac{1}{2}} e^{-(\mu(t,a_1)-\mu(t,a_2))^2/(8\sigma^2)}.$$

(39)

The minimization is changed to a maximization by taking the negative log of (39) and then is simplified to give,

$$t_{MM} = \arg\max_t |\mu(t, a_1) - \mu(t, a_2)|.$$

(40)

Since MMopt directly maximizes the response separation $r(t) = |\mu(t, a_1) - \mu(t, a_2)|$, it is identical to the Bopt design (38) for this example, and is given by,

$$t_{MM} = 1.4334.$$

(41)

Summary of results

Results are summarized in Table 1. The true Bayes Risk for each design is calculated using Monte Carlo simulation with 1.4×10^6 realizations.

MMopt gives the identical sampling time $t = 1.4334$ as the Bopt design, with the identical Bayes Risk of 0.1659. This is due to the fact that they both maximize response separation, which was seen in (37) to minimize Bayes Risk and be the optimal strategy for discriminating between the two support points. The ED optimal sampling time of 3.9991 is very different from the Bopt, and suffers a Bayes Risk that is 60% larger (i.e., 0.2712 vs. 0.1659). The fact that the ED design discriminates poorly is not surprising

Table 1 Summary of optimal sampling time designs for two-support-point example

Design metric	Sample time (h)	Bayes Risk (Prob)
Bopt	1.4334	0.1659
MMopt	1.4334	0.1659
ED	3.9991	0.2712
EID	0.7498	0.2003
ELD	1.1429	0.1705

since the criteria (34) is a simple average (over the prior) of the Fisher information calculated for each support point separately, which is generally unrelated to response separation. Likewise, the remaining Fisher-information-based designs (EID, and ELD) do not perform as well as Bopt and MMopt (although ELD is a close second), due to their criteria (11), (12), which ignore response separation.

Here, the MMopt sample times ensure that the MAP Bayesian estimate (i.e., the largest support point of the Bayesian posterior), will correctly determine the true patient 83% of the time (i.e., $0.8341 = 1 - 0.1659$). This can be an important advantage in applications to PK control, where the MAP estimate is often used to compute the patient's next drug dose or to decide subsequent treatment.

In summary, is has been shown that the key idea when optimizing experiment design for a NP model (with two support points: the simplest possible case), is to pick a sampling time that maximizes the separation between the two model responses. It is interesting that this notion is completely missing (and should not be expected), from Fisher-information based designs which are derived and justified based on completely different arguments. This example demonstrates the usefulness of MMopt, and clearly underscores the inherent difference between trying to design an optimal experiment to learn a parametric model versus a NP model.

Designs for two-support-point example with closely-spaced parameters

A modification of the two-support-point example is now studied, where the uncertainty in parameter a is chosen very small. In this case, it will be seen that the Bopt, MMopt, and the ED, EID, and ELD designs all give very similar results.

The two-support-point example (25), (26) is used again as in the previous example. However, in the current example the values for a_1 and a_2 are chosen very close

Model 1 : $a_1 = 1.1$ (fast support point), (42)

Model 2 : $a_2 = 1$ (slow support point), (43)

so that there is very little uncertainty in the problem. The two support point responses are plotted in Fig. 5. The Dopt designs for each support point taken separately are calculated as $t = 1/a_1 = 0.9091$ (for fast support point a_1) and $t = 1/a_2 = 1$ (for slow support point a_2). These sample times are relatively close together, in contrast to the previous example. Results are summarized in Table 2. As expected by the theory, the MMopt design again agrees exactly with Bopt design. However, somewhat unexpectedly, the remaining designs (ED, EID, ELD, Dopt) are also
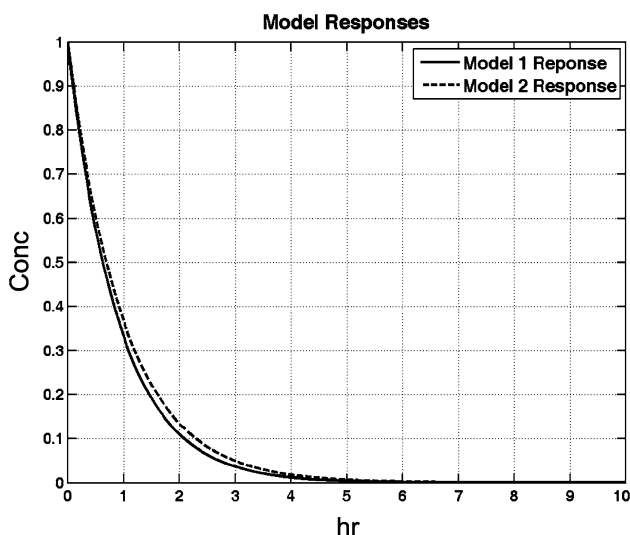
Fig. 5 Responses for two-support-point example with closely-spaced parameters

Table 2 Summary of optimal sampling time designs for two-support-point example with close parameters

Design metric	Sample time (h)	Bayes Risk (Prob)
Bayes optimal	0.9530	0.4767
MMopt	0.9530	0.4767
ED	0.9570	0.4767
EID	0.9480	0.4767
ELD	0.9520	0.4767
D-optimal	0.9524	0.4767

All designs give nearly the same optimal sample time at 0.95 h

very close to Bopt design, with essentially the same optimal sampling time at $t = 0.95$ h.

It is shown in Appendix that the fact that the various designs give the same optimal sampling time in this example is not an accident, but rather, follows from the asymptotic properties of the robust design cost functions as the parameters a_1 and a_2 become close. Specifically, it is proved in Corollary 1 (for the restricted two-support-point problem which is a slightly more general problem than treated here), that as the parameter uncertainty $\Delta a = a_2 - a_1$ becomes vanishingly small, the objectives of maximizing J_{ED}, J_D, J_{ELD} and minimizing J_{EID} asymptotically approach the single common objective of maximizing the response separation $r(t) = |\mu(a_1, t) - \mu(a_2, t)|$ over time t.

In summary, results from this two-support-point example indicate that the ED, EID, ELD, and Dopt designs applied to models having closely-space parameters (i.e., small model uncertainty), tend to agree with the Bopt and MMopt designs in this special case. The result is not just numerical, but is supported by the theory provided in Appendix.

Four-support-point example

System model

In this section the ED, EID, ELD, and Dopt designs are compared with MMopt on a more complicated four-support-point example with two unknown parameters. A two-parameter exponential measurement model is defined as

$$y_i = \mu(t_i, a, b) + \sigma n_i, \tag{44}$$

$$\mu(t, a, b) = be^{-at}, \tag{45}$$

$$n_i \sim N(0, \sigma^2). \tag{46}$$

Discrete parameter values are given in Table 3, with $\sigma = 0.1$ with prior $p_i = 0.25$, $i = 1, \ldots, 4$. The four support point responses are plotted in Fig. 6.

In this example, model parameters are intentionally chosen to help demonstrate the difference between using Fisher information and classification as experiment design objectives. Specifically, Fisher-information-based designs strongly favor putting a sample at the time of the highest output response, which in this example occurs at time $t = 0$. However, from a classification point of view, a sample at time $t = 0$ is partially wasted because it does not help discriminate between all responses. Specifically, three of

Table 3 Values for model parameters a and b

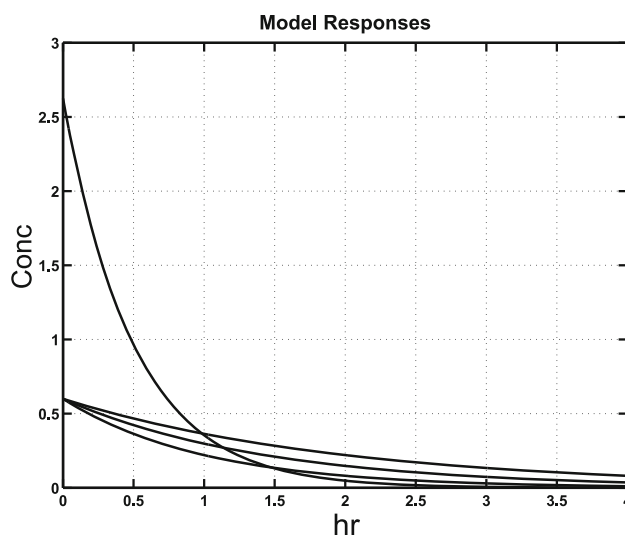| Index	a	b
1	2	2.625
2	1	0.6
3	0.7	0.6
4	0.5	0.6

Fig. 6 Responses for four-support-point example

the four support points shown in Fig. 6 have the same output response at $t = 0$ and are not well separated.

The designs are optimized over a time grid $t \in [0, 4]$, with grid points separated by 0.05 s apart. The Bopt design is too difficult to compute for this example and is not included. The Bayes cost for each design is evaluated using Monte Carlo analysis with 1.4×10^6 runs per estimate. Results are compared for the two and three sample time cases only, since the model has two parameters and the ED, EID and ELD designs do not exist for the one sample design.

Designs for four-support-point example

Results for the various metrics are summarized in Table 4 for a two sample design, and in Table 5 for a three sample design. As expected, the ED, EID, and ELD designs place a sample at $t = 0$ in all cases. The MMopt design improves on these designs by shifting this first sample slightly to the right (to $t = 0.45$ s), at which time the response curves are better separated, and hence better discriminated. It is seen that MMopt has smallest Bayes Risk for all cases studied.

Pharmacokinetic example

MMopt experiment design is demonstrated on a simulated PK example. The example is based on a one-compartment (first-order) model,

$$\dot{x} = -Kx + d(t), \tag{47}$$

$$\mu_k = \frac{x(t_k)}{V}, \tag{48}$$

$$y_k = \mu_k + \sigma_k n_k, \tag{49}$$

$$n_k \sim N(0, 1), \tag{50}$$

$$\sigma_k = 0.1. \tag{51}$$

Here, x is the amount of drug in the central compartment, $d(t)$ is the dose input (time-dependent in general), K is the elimination constant, V is the volume of distribution, y_k is the noisy measurement at time t_k, and σ_k is the standard deviation of its noise. The population model is comprised of $m = 10$ randomly chosen support points, whose values of V and K are given in Table 6.

The Bayesian prior is defined such that all 10 support points have equal probability, i.e., $p_i = 1/10$, $i = 1, \ldots, 10$. The system is driven by a dosing profile $d(t)$ having a value of 300 units that is held constant for 1 h and then shut off. The 10 support point responses to the dose input $d(t)$ are shown in Fig. 7. All optimization is performed over a discrete uniform time grid ranging from $t = 0$ to 24 h, with time points spaced at $\Delta t = 0.25$ ho (i.e., 15 min) apart. This time grid is chosen coarser than the grids used previously in order to simplify the optimization process, while at the same time being consistent with what can reasonably be enforced in a practical clinical

Table 6 Values of V and K in a pharmacokinetic model with 10 support points

#	K	V
1	0.090088	113.7451
2	0.111611	93.4326
3	0.066074	90.2832
4	0.108604	89.2334
5	0.103047	112.1093
6	0.033965	94.3847
7	0.100859	109.8633
8	0.023174	111.7920
9	0.087041	108.6670
10	0.095996	100.3418
Mean	0.0820	102.3852
STD	0.0311	9.8615

Table 4 Summary of optimal two-sample designs for four-support-point example

Design metric	Two-sample times (h)		Bayes Risk (Prob)	99% Bayes Risk conf bound
MMopt	0.45	1.4	0.32839	±0.00070
ED	0	0.8	0.37028	±0.00070
EID	0	1	0.36044	±0.00072
ELD	0	0.95	0.36234	±0.00072

Table 5 Summary of optimal three-sample designs for four-support-point example

Design metric	Three-sample times (h)			Bayes Risk (Prob)	99% Bayes Risk conf bound
MMopt	0.45	1.4	1.4	0.28065	±0.00067
ED	0	0.7	0.9	0.32048	±0.00067
EID	0	0	1	0.36034	±0.00072
ELD	0	0.85	1.05	0.30990	±0.00069
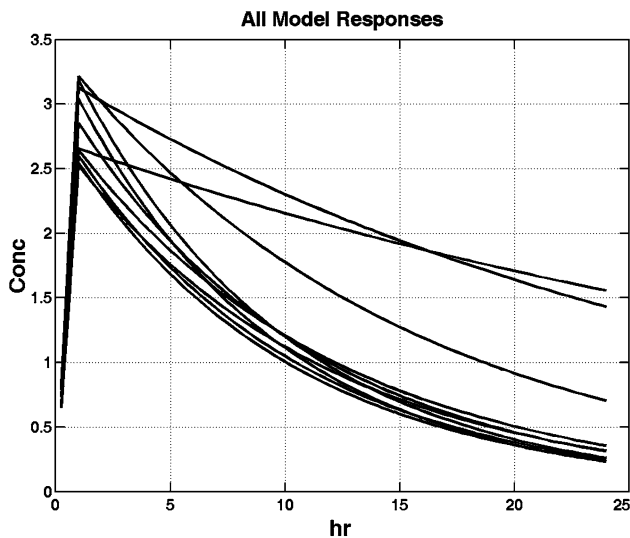
Fig. 7 Ten support point responses to dose input $d(t)$

Fig. 8 Bayes Risk computed using 1×10^6 Monte Carlo runs, showing the one-sample Bayesian design at Bopt = 4.25 h with an optimal value of Prob = 0.5474

application. For each case studied, results are computed using a Monte Carlo simulation having 1×10^6 runs. The corresponding 99% confidence bounds are calculated from the Monte Carlo statistics and reported for completeness.

Design for PK estimation

MMopt design is applied to the problem of estimating the PK parameters V and K in models (47)–(49) by designing experiments having one, two and three optimal sample times. For the one sample design, MMopt is compared to Bopt (the optimal Bayesian design), while for the two- and three-sample designs, MMopt is compared to ED. This choice is motivated by the fact that ED is not defined for a one sample design, while Bopt becomes computationally unwieldy for the two- and three-sample cases.

Performance is summarized in Table 7. For the one-sample design, the true Bayes Risk is computed numerically using 1×10^6 Monte Carlo runs, and is shown in Fig. 8. The minimum is attained at $t = 4.25$ h. The shape of the curve is interesting, having several local minima which would generally complicate its optimization. For the one-sample
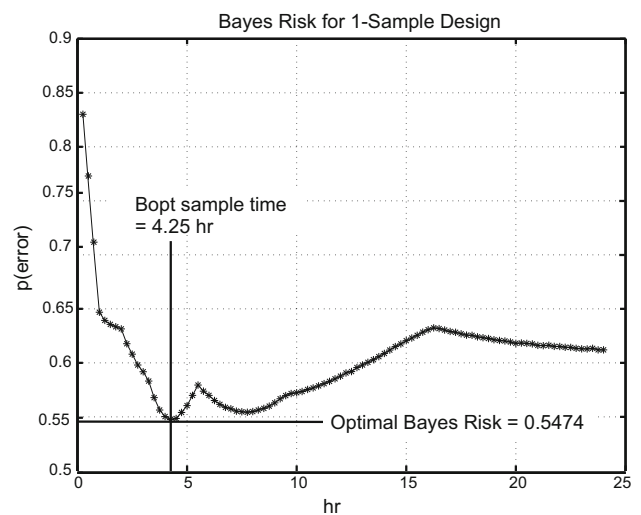
design, MMopt also attains an optimal at $T = 4.25$ h, which is identical to that of Bopt. Hence, MMopt is globally optimal for this example. Of course this result is obtained on a coarse time grid, so they may not agree exactly on a finer grid. For the two and three sample designs, MMopt has lower Bayes Risk than ED design. In both cases, the improvement is statistically significant with p-values smaller than 2.2e−16 (machine precision). The low p-values are due to the high confidence obtained from using a large number of runs in each Monte Carlo evaluation.

Voriconazole case study

We wished to use MMopt to develop an optimal first sampling strategy for oral voriconazole in children. Because this drug has a mixture of linear and non-linear elimination, which is related to but not entirely predicted by age, it is a challenge for practitioners to know when steady state is achieved. This in turn makes comparison of

Table 7 Results for optimal one, two and three sample designs applied to PK estimation

Design metric	Samples (h)			Bayes Risk (Prob)	99% Conf (Prob)
One-sample design					
Bopt	4.25			0.5474	±0.0015
MMopt	4.25			0.5474	±0.0015
Two-sample design					
MMopt	1	9.5		0.2947	±0.0014
ED	1	24		0.3272	±0.0014
Three-sample design					
MMopt	1	1	10.5	0.2325	±0.0013
ED	1	1	24	0.2617	±0.0013

a trough concentration to a pre-established target steady-state range of 1–5.5 mg/L difficult due to the uncertainty about steady-state conditions. For this optimal sampling project, we used the same dataset that we previously used to build a population model of voriconazole in adults and children [39]. The data comprised 141 subjects, with 85 children ranging from 2 to 11 years of age, and 56 adults 19 to 55 years of age. The population model had seven parameters with linear absorption, allometrically scaled Michaelis–Menten clearance and volume of the central compartment, a peripheral tissue compartment, and bioavailability. There were 125 support points in the final NP population model.

We first used the MMopt function that has been incorporated into our Pmetrics NP population modeling and simulation package for R to calculate one and two optimal times to sample after the first maintenance dosage of voriconazole. The MMopt calculations were constrained to 0.5 h intervals on the range from 0 to 36 h. We used the population points and assay error polynomial (5) of $c_0 = 0.02$, $c_1 = 0.1$, $c_2 = c_3 = 0$, with loading dosages of 9 mg/kg two times in the first 24 h, followed by the first maintenance dose of 8 mg/kg at the beginning of day 2, 12 h after the second loading dose. We combined three sets of 125 time–concentration profiles generated from the support points in the model, with each set corresponding to a child aged 2, 7 or 12 years, and weighing the median for each age, according to CDC growth charts: 12.2, 23.3 and 36.1 kg. The single optimal time point to sample (MMopt1) was 1 h before the second maintenance dose, i.e., 35 h into the regimen. The two optimal time points (MMopt2) were a peak and trough, 1 h after the first maintenance dose, and just before the second maintenance dose, i.e., 25 and 36 h into the regimen. This is shown in Fig. 9.

Next, we used the population model with full covariances to simulate 50 new subjects administered oral voriconazole 9 mg/kg q12h for two doses, followed by a single 8 mg/kg dose. The mean (range) simulated weights and ages were 40.6 (12–89) kg and 6.4 (2.2–11.1) years. We simulated hourly concentrations until 24 h after the final dose, and these observations were corrupted with random noise drawn from a Gaussian distribution with mean of 0 and standard deviation based on the assay error coefficients above.

The complete data for each simulated subject was defined as the "full" dataset. We depleted the full data set to include only the MMopt2 observations at times 25 and 36 h. We also depleted it to include only the MMopt1 observation time at 35 h. We used the model to calculate the full Bayesian posterior with concentrations every 12 min from 0 to 48 h for each dataset (full, MMopt2, MMopt1). From each posterior, we calculated the area-under-the-curve (AUC) from 24 to 36 h by trapezoidal approximation, as well as extracted the concentration at time 25 h (1 h post-dose peak) and 36 h (12 h post-dose trough).
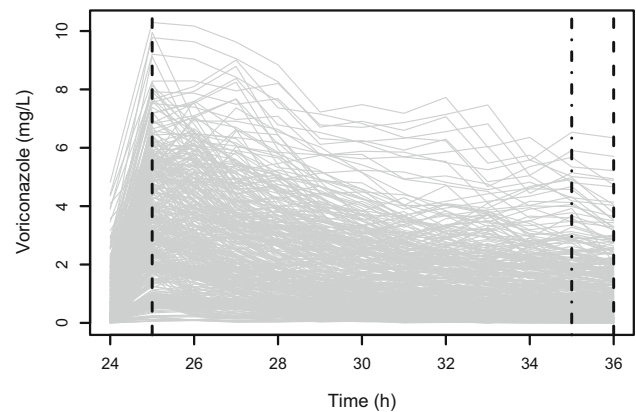

Fig. 9 Oral dosing responses of time versus concentration, simulated from a voriconazole nonparametric model with 125 support points. Patients are given loading doses at 0 and 12 h, and a maintenance dose at 24 h. MMopt samples times are determined on a grid constrained to 0.5 h intervals. The two optimal sample times (MMopt2) are shown in *dashed lines* at 25 and 36 h. The one optimal sample time (MMopt1) is indicated by the *dash-dot line* at 35 h

Using the full dataset to generate a Bayesian posterior, the simulated observations could be predicted as $0.05 + 0.92*$ prediction, $R^2 = 0.87$, indicating a good fit. We therefore defined the AUC, peak and trough from the Bayesian posterior based on the full dataset to be the reference, and compared the AUCs, peaks and troughs from the MMopt2 and MMopt1 posteriors using a paired T-test. The results are shown in Table 8. Also included in the table is the 95th percentile of the absolute prediction difference as a measure of imprecision.

In conclusion, we could show that using two MMopt samples, one obtained 1 h after the first maintenance dose and one just prior to the second maintenance dose, permits estimation of voriconazole AUC, peaks and troughs with reasonably low bias. While the peaks are significantly underpredicted, this is not likely to affect clinical outcome since efficacy has been linked to AUC or trough concentrations [1, 25, 27, 35, 38], and the absolute underprediction is low. A single sample 1 h before the second maintenance dose is very similar in accuracy to two samples. Using one or two MMopt samples with this population model results in an error in trough prediction that is 95% likely to be less than 0.38 or 0.93 mg/L, respectively. Dopt sampling would require seven blood samples, which is not readily feasible even in the inpatient setting. MMopt sampling will permit rapid assessment of the appropriateness of the voriconazole oral dose in children, and subsequent sampling times in individual patients can be guided by further use of MMopt for each patient. Clinicians may now have a tool that overcomes the concerns about steady state, and which uses the patient's own measured

Table 8 Bias and imprecision of predictions from Bayesian posteriors based on two optimal samples (MMopt2) or one (MMopt1) optimal sample compared to posteriors based on hourly samples (Full)

	Full	MMopt2	MMopt1
Bias (MMopt minus Full)			
AUC (mg h/L)	18.53	−0.82 (−3.18 to 1.55), P = 0.49	−1.86 (−4.13 to 0.40), P = 0.11
Peak (mg/L)	2.58	0.37 (0.16 to 0.58), P < 0.001	0.42 (0.04 to 0.79), P = 0.03
Trough (mg/L)	0.8	0.01 (−0.16 to 0.18), P = 0.87	−0.09 (−0.20 to 0.03), P = 0.12
95th Percentile imprecision (|MMopt minus Full|)			
AUC (mg h/L)		11.5	10.75
Peak (mg/L)		1.13	2.55
Trough (mg/L)		0.93	0.38

The MMopt2 samples are taken 1 h after and just before the first and second oral maintenance voriconazole doses, respectively. The MMopt1 sample is taken 1 h before the second oral maintenance voriconazole dose

concentrations to control both the dose and the time to best sample.

Discussion

Several properties of the various experiment design approaches are summarized in Table 9, and are discussed below.

An interesting question arises whether an experiment design is invariant under a change of variables in the model. Let an estimation problem having parameter vector $a \in R^p$ be reparametrized in terms of the vector $b \in R^p$ where a and b are related by the mapping,

$$a = f(b). \tag{52}$$

The mapping (52) is assumed to be *regular* in the sense that it is one-to-one and has continuous partial derivatives. In the case of a regular *linear* reparametrization, the mapping (52) takes the form

$$a = Fb, \tag{53}$$

where $F \in R^{p \times p}$ is a square and invertible matrix.

P1 *Invariance under regular linear reparametrization.*

P2 *Invariance under regular nonlinear reparametrization.*

An obvious property of support point responses is that they are invariant under regular reparametrization. Accordingly, methods such as Bopt and MMopt based purely on support point responses, will also trivially be invariant under a regular reparametrization (of both the linear and nonlinear kind). However the same is not true for the Fisher-based robust designs. An analysis performed in the (Supplemental File, 8/13/2016) proves that while all the robust designs are invariant under linear parametrization, only the ELD design (not the ED and EID designs) are invariant under nonlinear parametrization.

This lack of invariance can have significant consequences in practice. For example, a one-compartment PK model is parameterized by volume V and elimination rate K as

$$\dot{x} = -Kx + d, \tag{54}$$

$$\mu_i = \frac{x(t_i)}{V}. \tag{55}$$

Here, the drug amount in the compartment is x, the dose is d, and μ_i indicates the noiseless part of a drug concentration measurement at time t_i. However, an alternative parametrization is given by using parameters volume V and clearance C (where $C = VK$) [21],

$$\dot{x} = -(C/V)x + d, \tag{56}$$

Table 9 Properties of ED, EID, ELD, MMopt, and Bopt designs, where Y stands for "yes," and N stands for "no"

#	Properties	ED	EID	ELD	MMopt	Bopt
P1	Invariance under regular linear reparametrization	Y	Y	Y	Y	Y
P2	Invariance under regular nonlinear reparametrization	N	N	Y	Y	Y
P3	Avoids dependence on estimator asymptotic properties	N	N	N	Y	Y
P4	Avoids knowing true parameters in nonlinear problems	Y	Y	Y	Y	Y
P5	Allows taking fewer than p samples	N	N	N	Y	Y
P6	Can handle heterogeneous model structures	N	N	N	Y	Y
P7	Gives known optimal solution to two-support point example	N	N	N	Y	Y
P8	Captures main elements of pre-posterior analysis	N	N	N	Y	Y
P9	Solves optimal Bayesian classifier problem	N	N	N	N	Y

$$\mu_i = \frac{x(t_i)}{V}. \tag{57}$$

According to Table 9 property (P2), when using the ED or EID methods, a clinician would have to compute separate experiment designs depending on whether he was using the (V, K) or (V, C) parametrization. This behavior is non-intuitive as the responses are identical under either parametrization and should not call for separate designs. In contrast, the ELD, MMopt and Bopt designs are invariant under this change of variables.

P3 *Avoids dependence on estimator asymptotic properties.*

P4 *Avoids knowing true parameters in nonlinear problems.*

As discussed in Pronzato and Pazman [40], there are two common avenues for circular reasoning to enter into the experimental design process. The first [cf., (P3)], is the dependence on asymptotic properties of the estimator. This represents a contradiction since experiment designs are typically used for placing only a small (non-asymptotic) number of samples, while certain information measures (e.g., Fisher information) are rigorously only valid asymptotically as the number of samples goes to infinity. The second [cf., (P4)], is the assumption that the true parameters are available beforehand for use in computing the optimal design. This represents a contradiction since it is the whole point of experimental design to find these parameters in the first place.

MMopt and Bopt both satisfy (P3) and (P4) and completely avoid these contradictions. However, the robust designs do not satisfy (P3) because they are defined in terms of Fisher information which involves an asymptotic approximation. Nevertheless, all robust designs avoid the second contradiction [and hence satisfy (P4)], due to the fact that they average over the prior distribution.

P5 *Allows taking fewer than p samples.*

Both Bopt and MMopt allow taking fewer than p samples, where p is the number of parameters. The Dopt robust designs do not enjoy this same property since their cost would be zero for all sub-sampled designs. It is worth noting that certain heuristic modifications of Dopt designs have been introduced to help overcome this limitation [17].

P6 *Can handle heterogeneous model structures.*

The Bopt and MMopt experiment designs are based purely on support point responses and as such, are able (trivially) to handle heterogeneous model structures. In particular, the underlying discrete NP population can be a mix of models having different orders or numbers of compartments, linear or

nonlinear dynamics, or any other combination of heterogeneous characteristics. The Dopt designs do not enjoy this same property. However, several extensions relevant to model discrimination have been developed (cf., [2, 54]).

P7 *Gives known optimal solution to two-support-point example.*

It was shown theoretically and confirmed numerically in the two-support-point example of Section , that both Bopt and MMopt maximize the response separation and consequently correspond to the Bopt design. Hence both Bopt and MMopt satisfy (P7). Of course the robust designs are not designed to maximize response separation, and as expected, they generally fail to give the Bopt design even in the two-support-point example. However, an interesting property proved theoretically in Appendix and confirmed by example in Section is that the robust designs act to minimize Bayes Risk in the limit as the parameter uncertainty becomes vanishingly small.

P8 *Captures main elements of pre-posterior analysis.*

P9 *Solves optimal Bayesian classifier problem.*

The idea of pre-posterior analysis is to optimize a "what-if" thought experiment that calculates the Bayesian posterior resulting from "having already implemented" a given candidate experiment design (cf., [33, 42]). In the current paper, only the Bayesian optimal classifier (Bopt) satisfies (P9) since it minimizes Bayes Risk which is a property of the Bayesian posterior. However, MMopt minimizes an over-bound on the true Bayes Risk, which acts to indirectly reduce Bayes Risk. Hence MMopt satisfies (P8) and captures important elements of pre-posterior analysis.

It is worth noting that there is a separate literature on Bayesian optimal design [43]. Pre-posterior analysis methods from Bayesian optimal design have been applied in control theory [4, 5, 53] and in the context of PK problems in Merle and Mentre [37] and Schumitzky [45]. Interestingly, the second method discussed in [37] can be directly related to MMopt in that it defines informative experiments as those that minimize the covariance of the pre-posterior Bayesian distribution, while MMopt defines informative experiments as those that maximize its mode. Intuitively, both approaches are acting to "hone in" on the parameter values by making the pre-posterior Bayesian distribution "tall" (in the case of MMopt), and "thin" (in the case of [37]).

Generally speaking, Bayesian optimal approaches are promising and may find application to a broad class of PK problems in the future. However for the NP models considered here, such approaches are more complicated than MMopt, and generally require Monte Carlo simulation, multidimensional integration, and/or stochastic analysis which significantly adds to their computation.

Conclusions

MMopt is introduced as a novel experiment design method uniquely relevant to NP models. MMopt minimizes an overbound on the Bayes Risk, where the overbound in based on a key inequality first proved in [11]. The main idea is to think of individualized therapy with respect to an NP (i.e., discrete) prior, as a classification problem to find the support point that best matches the patient's behavior. This classification-based approach leads to the important notion of "response separation" as a guiding principle for understanding and optimizing the NP problem. It is interesting that Fisher-based designs completely ignore response separation, being motivated instead by arguments specifically relevant to parametric models. The issue of response separation underscores the essential difference between optimal designs intended for parametric models versus NP models. Despite this difference, the two approaches were shown to give identical results in the special case where the prior uncertainty becomes vanishingly small. This indicates that the Fisher-based designs may have some value in a restricted local sense when applied to NP models. Bayes Risk minimization has the additional advantage of causing the MAP Bayesian estimate to be more accurate, which is relevant to PK applications where MAP Bayesian control is applied for determining individualized drug doses. Using numerical examples, including a PK voriconazole study, and supporting theoretical analysis in an Appendix and Supplemental File, the usefulness of MMopt is demonstrated and discussed with respect to the ED, EID, and ELD classes of Fisher-information-based robust designs.

Future work will directed at comparing MMopt with alternative experimental design approaches, and developing a weighted version of MMopt that incorporates a matrix of misclassification costs (i.e., the cost of misclassifying support point i as j). This latter development will allow the clinician to tailor the experimental design to elicit specific information needed to support to a diverse range of clinical goals.

Appendix: Two-support-point problem with close parameters

Definition 1 (*Two-support-point problem*) A general class of two-support-point problems is defined of the form,

$$y(t_k) = \mu(t_k, a) + \sigma_k n_k, \tag{58}$$

where $a \in \Omega \overset{\Delta}{=} \{a_1, a_2\}$ is a random variable taking on values a_1 and a_2; $z(t, a_i)$, $i = 1, 2$ are the support point responses, each assumed to be continuous over the closed interval $t \in \mathcal{T} \overset{\Delta}{=} [t_A, t_B]$; $y(t_k)$ is the noisy measurement of $z(t, a)$ taken at discrete time t_k; $n_k \sim N(0, 1)$ is the measurement noise at time instant t_k; σ_k scales n_k to a desired level; and the prior support point probabilities are specified as p_1 for a_1 and p_2 for a_2 where $p_1 > 0$, $p_2 > 0$, and $p_1 + p_2 = 1$.

The experiment design problem is to find the set U of n optimal sampling times

$$U = \{t_1, \ldots, t_n\}, \tag{59}$$

on a specified time interval $t \in \mathcal{T} \overset{\Delta}{=} [t_A, t_B]$ that act to best identify the parameter vector $a \in \{a_1, a_2\}$.

Definition 2 (*Restricted two-support-point problem*) A restricted two-support-point problem is defined by Definition 1 under the additional conditions that the noise is independent of time, $\sigma_k = \sigma$; there is only one sample to be taken, $n = 1$, $U = \{t_1\}$; and the prior support point probabilities are uniformly distributed, i.e., $p_1 = p_2 = 0.5$. Furthermore, it is assumed that the function $\mu(t, a)$ is sufficiently smooth to admit a first-order Taylor expansion in the vicinity of parameter values a_1 and a_2 for all $t \in \mathcal{T}$.

In what follows, the asymptotic order notation $\mathcal{O}(\epsilon^k)$ is used to indicate a term that decreases to zero as ϵ^k. Specifically, $\lim_{\epsilon \to 0} \mathcal{O}(\epsilon^k) = 0$ and there exists a constant M such that [9],

$$\lim_{\epsilon \to 0} |\mathcal{O}(\epsilon^k)/\epsilon^k| = M < \infty. \tag{60}$$

Furthermore, the notational dependence of $\mu(t, a)$ on time t will be suppressed and denoted more simply as $\mu(a)$.

Theorem 2 *The costs associated with the ED, D [evaluated on the mean parameter value $\overline{a} = \frac{1}{2}(a_1 + a_2)$], EID and ELD optimal designs for the restricted two-support-point problem of Definition 2 can be written in terms of the response separation $r(t) = |\mu(a_2) - \mu(a_1)|$ for sufficiently small parameter uncertainty $\Delta a = a_2 - a_1$ as,*

$$J_{ED} = 0.5|M(a_1)| + 0.5|M(a_2)| = \frac{r(t)^2}{\sigma^2 \Delta a^2} + \mathcal{O}(\Delta a), \tag{61}$$

$$J_D = |M(\overline{a})| = \frac{r(t)^2}{\sigma^2 \Delta a^2} + \mathcal{O}(\Delta a), \tag{62}$$

$$J_{EID} = 0.5|M(a_1)|^{-1} + 0.5|M(a_2)|^{-1} = \frac{\sigma^2 \Delta a^2}{r(t)^2}(1 + \mathcal{O}(\Delta a)), \tag{63}$$

$$J_{ELD} = 0.5 \ln |M(a_1)| + 0.5 \ln |M(a_2)|$$
$$= 0.5 \ln \left(\frac{r(t)^4}{\sigma^4 \Delta a^4} + \mathcal{O}(\Delta a) \right). \tag{64}$$

Proof of Eq. (61): Consider the response $\mu(a)$ as a function of the parameter a. Let $\mu(a_2)$ be expressed in terms of the Taylor expansion about a_1

$$\mu(a_2) = \mu(a_1) + \frac{\partial \mu}{\partial a}\Big|_{a_1}(a_2 - a_1) + \mathcal{O}(a_2 - a_1)^2. \tag{65}$$

Letting $\Delta a \overset{\Delta}{=} a_2 - a_1$ yields upon rearranging,

$$\mu(a_2) - \mu(a_1) = \frac{\partial \mu}{\partial a}\Big|_{a_1} \Delta a + \mathcal{O}(\Delta a)^2. \tag{66}$$

Squaring both sides of (66) gives

$$(\mu(a_2) - \mu(a_1))^2 = (\Delta a)^2 \left(\frac{\partial \mu}{\partial a}\Big|_{a_1} \right)^2 + \mathcal{O}(\Delta a)^3. \tag{67}$$

Similarly, let $\mu(a_1)$ be expressed in terms of the Taylor expansion about a_2

$$\mu(a_1) = \mu(a_2) + \frac{\partial \mu}{\partial a}\Big|_{a_2}(a_1 - a_2) + \mathcal{O}(\Delta a)^2, \tag{68}$$

which can be rearranged as

$$\mu(a_2) - \mu(a_1) = \frac{\partial \mu}{\partial a}\Big|_{a_2} \Delta a + \mathcal{O}(\Delta a)^2. \tag{69}$$

Squaring both sides of (69) yields

$$(\mu(a_2) - \mu(a_1))^2 = (\Delta a)^2 \left(\frac{\partial \mu}{\partial a}\Big|_{a_2} \right)^2 + \mathcal{O}(\Delta a)^3. \tag{70}$$

Adding one half of (67) to one half of (70) gives

$$(\mu(a_2) - \mu(a_1))^2 = (\Delta a)^2 \left(\frac{1}{2}\left(\frac{\partial \mu}{\partial a}\Big|_{a_1}\right)^2 + \frac{1}{2}\left(\frac{\partial \mu}{\partial a}\Big|_{a_2}\right)^2 \right) + \mathcal{O}(\Delta a)^3. \tag{71}$$

Dividing both sides by $\sigma^2(\Delta a)^2$ gives,

$$\frac{1}{2\sigma^2}\left(\frac{\partial \mu}{\partial a}\Big|_{a_1}\right)^2 + \frac{1}{2\sigma^2}\left(\frac{\partial \mu}{\partial a}\Big|_{a_2}\right)^2 = \frac{(\mu(a_2) - \mu(a_1))^2}{\sigma^2 \Delta a^2} + \mathcal{O}(\Delta a). \tag{72}$$

Proof of Eq. (62): Combining relations $\overline{a} = \frac{1}{2}(a_1 + a_2)$ and $\Delta a = a_2 - a_1$, gives

$$\overline{a} = a_1 + \frac{1}{2}\Delta a, \tag{73}$$

$$\overline{a} = a_2 - \frac{1}{2}\Delta a. \tag{74}$$

Let $\mu(a_2)$ be expressed in terms of a Taylor expansion of $\mu(a)$ about $\overline{a}$

$$\mu(a_2) = \mu(\overline{a}) + \frac{\partial \mu}{\partial a}\Big|_{\overline{a}}(a_2 - \overline{a}) + \mathcal{O}(a_2 - \overline{a})^2. \tag{75}$$

Rearranging (74) gives $a_2 - \overline{a} = \frac{1}{2}\Delta a$ which is substituted into (75) to give

$$\mu(a_2) = \mu(\overline{a}) + \frac{\partial \mu}{\partial a}\Big|_{\overline{a}}\left(\frac{1}{2}\Delta a\right) + \mathcal{O}(\Delta a)^2 \tag{76}$$

Similarly, let $\mu(a_1)$ be expressed in terms of a Taylor expansion of $\mu(a)$ about $\overline{a}$

$$\mu(a_1) = \mu(\overline{a}) + \frac{\partial \mu}{\partial a}\Big|_{\overline{a}}(a_1 - \overline{a}) + \mathcal{O}(a_1 - \overline{a})^2. \tag{77}$$

Rearranging (73) gives $a_1 - \overline{a} = -\frac{1}{2}\Delta a$ which is substituted into (77) to give

$$\mu(a_1) = \mu(\overline{a}) - \frac{\partial \mu}{\partial a}\Big|_{\overline{a}}\left(\frac{1}{2}\Delta a\right) + \mathcal{O}(\Delta a)^2. \tag{78}$$

Subtracting (78) from (76) and rearranging gives

$$\mu(a_2) - \mu(a_1) = \frac{\partial \mu}{\partial a}\Big|_{\overline{a}} \Delta a + \mathcal{O}(\Delta a)^2. \tag{79}$$

Squaring both sides of (79) and dividing by $\sigma^2(\Delta a)^2$ gives upon rearranging,

$$\frac{1}{\sigma^2}\left(\frac{\partial \mu}{\partial a}\Big|_{\overline{a}}\right)^2 = \frac{(\mu(a_2) - \mu(a_1))^2}{\sigma^2 \Delta a^2} + \mathcal{O}(\Delta a). \tag{80}$$

Proof of Eq. (63): Reciprocating both sides of Eq. (67), using the relation $\frac{1}{1+\epsilon} = 1 - \epsilon + \mathcal{O}(\epsilon^2)$ for small ϵ, and rearranging gives,

$$\frac{\Delta a^2}{(\mu(a_2) - \mu(a_1))^2} = \left(\frac{\partial \mu}{\partial a}\Big|_{a_1}^2\right)^{-1}(1 - \mathcal{O}(\Delta a)). \tag{81}$$

Likewise, reciprocating both sides of Eq. (70) and rearranging gives,

$$\frac{\Delta a^2}{(\mu(a_2) - \mu(a_1))^2} = \left(\frac{\partial \mu}{\partial a}\Big|_{a_2}^2\right)^{-1}(1 - \mathcal{O}(\Delta a)). \tag{82}$$

Adding (81) and (82) and multiplying both sides by $\sigma^2/2$ gives upon rearranging

$$0.5\left(\frac{1}{\sigma^2}\frac{\partial \mu}{\partial a}\Big|_{a_1}^2\right)^{-1} + 0.5\left(\frac{1}{\sigma^2}\frac{\partial \mu}{\partial a}\Big|_{a_2}^2\right)^{-1}$$
$$= \frac{\sigma^2 \Delta a^2}{(\mu(a_2) - \mu(a_1))^2}(1 + \mathcal{O}(\Delta a)). \tag{83}$$

Proof of Eq. (64): Multiplying (67) and (70) yields,

$$(\mu(a_2) - \mu(a_1))^4 = \Delta a^4 \left(\frac{\partial \mu}{\partial a}\bigg|_{a_1}\right)^2 \left(\frac{\partial \mu}{\partial a}\bigg|_{a_2}\right)^2 + \mathcal{O}(\Delta a)^5. \tag{84}$$

Dividing both sides by $\sigma^4 \Delta a^4$ gives upon rearranging

$$\left(\frac{1}{\sigma^2}\frac{\partial \mu}{\partial a}\bigg|_{a_1}^2\right)\left(\frac{1}{\sigma^2}\frac{\partial \mu}{\partial a}\bigg|_{a_2}^2\right) = \frac{(\mu(a_2) - \mu(a_1))^4}{\sigma^4 \Delta a^4} + \mathcal{O}(\Delta a). \tag{85}$$

Taking $\frac{1}{2}\ln(\cdot)$ of both sides gives

$$0.5 \ln\left(\frac{1}{\sigma^2}\frac{\partial \mu}{\partial a}\bigg|_{a_1}^2\right) + 0.5 \ln\left(\frac{1}{\sigma^2}\frac{\partial \mu}{\partial a}\bigg|_{a_2}^2\right)$$
$$= 0.5 \ln\left(\frac{(\mu(a_2) - \mu(a_1))^4}{\sigma^4 \Delta a^4} + \mathcal{O}(\Delta a)\right). \tag{86}$$

Corollary 1 *As the parameter uncertainty $\Delta a = a_2 - a_1$ becomes small in the restricted two-support-point problem of Definition 2, the objective functions J_{ED}, J_D and J_{ELD} approach a monotonically increasing function of the response separation $r(t)$, while J_{EID} approaches a monotonically decreasing function of the response separation $r(t)$.*

Proof The result follows by the properties of relations (61), (62), (63), (64) as Δa becomes small, noting that functions $1/x$ and $\ln(x)$ are monotonically decreasing and increasing in x, respectively. $\qquad\square$

Corollary 1 indicates that as the parameter uncertainty $\Delta a = a_2 - a_1$ becomes small in the restricted two-support-point problem of Definition 2, the objectives of maximizing J_{ED}, J_D, J_{ELD} and minimizing J_{EID} asymptotically approach the single common objective of maximizing the response separation $r(t) = |\mu(a_2, t) - \mu(a_1, t)|$ over time $t \in \mathcal{T}$. These asymptotic properties are examined numerically in the "Two-support-point example" section.

References

1. Andes D, Marchillo K, Stamstad T, Conklin R (2003) In vivo pharmacokinetics and pharmacodynamics of a new triazole, voriconazole, in a murine candidiasis model. Antimicrob Agents Chemother 47(10):3165–3169
2. Atkinson AC, Donev AN, Tobias RD (2007) Optimum experimental designs, with SAS. Oxford University Press, New York
3. Baek Y (2006) An interior point approach to constrained nonparametric mixture models. PhD Thesis, University of Washington
4. Bayard DS (1991) A forward method for optimal stochastic nonlinear and adaptive control. IEEE Trans Autom Control 36(9):1046–1053
5. Bayard DS, Schumitzky A (2010) Implicit dual control based on particle filtering and forward dynamic programming. Int J Adapt Control Signal Process 24(3):155–177
6. Bayard DS, Jelliffe R, Schumitzky A, Milman M, Van Guilder M (1995) Precision drug dosage regimens using multiple model adaptive control: theory and application to simulated vancomycin therapy. In: Selected topics in mathematical physics, Prof. R. Vasudevan Memorial volume. World Scientific Publishing Co., Madras
7. Bayard DS, Jelliffe R, Neely M (2013) Bayes risk as an alternative to Fisher information in determining experimental designs for nonparametric models. In: PODE 2013 workshop: population optimum design of experiments, Lilly, UK, 15 June 2013
8. Bayard DS, Jelliffe R, Neely M (2014) Experiment design based on Bayes risk and weighted Bayes risk with application to pharmacokinetic systems. In: PODE 2014 workshop: population optimum design of experiments, Roche, Basel, 11 September 2014
9. Bender CM, Orszag SA (1978) Advanced mathematical methods for scientists and engineers. McGraw-Hill Book Company, New York
10. BestDose software package. http://www.lapk.org. Accessed 1 Jan 2016
11. Blackmore L, Rajamanoharan S, Williams BC (2008) Active estimation for jump Markov linear systems. IEEE Trans Autom Control 53(10):2223–2236
12. Box REP, Lucas HL (1959) Design of experiments in non-linear situations. Biometrika 46:77–90
13. Chaloner K, Verdinelli I (1995) Bayesian experimental design: a review. Stat Sci 10(3):273–304
14. D'Argenio DZ (1981) Optimal sampling times for pharmacokinetic experiments. J Pharmacokinet Biopharm 9(6):739–756
15. D'Argenio DZ (1990) Incorporating prior parameter uncertainty in the design of sampling schedules for pharmacokinetic parameter estimation experiments. Math Biosci 99:105–118
16. D'Argenio DZ, Schumitzky A, Wang X (2009) ADAPT 5 user's guide. Biomedical simulation resource. University of Southern California
17. Draper NR, Hunter WG (1967) The use of prior distributions in the design of experiments for parameter estimation in non-linear situations. Biometrika 54(1–2):147–153
18. Duda RO, Hart PE, Stork DG (2001) Pattern classification, 2nd edn. Wiley, New York
19. Fedorov VV (1972) Theory of optimal experiments. Academic, New York
20. Fedorov VV, Leonov SL (2013) Optimal design for nonlinear response models. CRC Press, New York
21. Gibaldi M, Perrier D (1982) Pharmacokinetics, 2nd edn. Marcel Dekker, Inc., New York
22. Goodwin GC, Payne RL (1977) Dynamic system identification: experiment design and data analysis. Academic, New York
23. Hennig S, Nyberg J, Fanta S, Backman JT, Hoppu K, Hooker AC, Karlsson MO (2012) Application of the optimal design approach to improve a pretransplant drug dose finding design for ciclosporin. J Clin Pharmacol 52(3):347–360. doi:10.1177/0091270010397731 (Epub 4 May 2011)
24. Hu C, Lovejoy WS, Shafer SL (1994) Comparison of some control strategies for three-compartment PK/PD models. J Pharmacokinet Biopharm 22(6):525–550
25. Huurneman LJ, Neely M, Veringa A, Prez FD, Ramos-Martin V, Tissing WJ, Alffenaar JWC, Hope W (2016) Pharmacodynamics of voriconazole in children: further steps along the path to true individualized therapy. Antimicrob Agents Chemother 60(4):2336–2342
26. Jelliffe R, Schumitzky A, Bayard D, Van Guilder M, Leary RH (2002) The USC*PACK programs for parametric and nonparametric population PK/PD modeling. Population Analysis Group in Europe, Paris

27. Karthaus M, Lehrnbecher T, Lipp HP, Kluge S, Buchheidt D (2015) Therapeutic drug monitoring in the treatment of invasive aspergillosis with voriconazole in cancer patients: an evidence-based approach. Ann Hematol 94(4):547–556

28. Katz D, D'Argenio DZ (1984) Discrete approximation of multivariate densities with application to Bayesian estimation. Comput Stat Data Anal 2(1):27–36

29. Kendall M, Stuart A (1979) The advanced theory of statistics, vol 2, 4th edn. Charles Griffin and Company, London

30. Kistos CP (2013) Optimal experimental design for non-linear models: theory and applications. Springer, New York

31. Landaw EM (1980) Optimal experiment design for biologic compartmental systems with applications to pharmacokinetics. PhD Dissertation, University of California, Los Angeles

32. Leary R, Jelliffe R, Schumitzky A, Van Guilder M (2001) An adaptive grid non- parametric approach to pharmacokinetic and dynamic (PK/PD) population models. In: Proceedings of the 14th IEEE symposium on computer-based medical systems, 2001, CBMS 2001. IEEE, p 389–394

33. Lindley DV (1972) Bayesian statistics, a review. Society for Industrial and Applied Mathematics (SIAM), Philadelphia

34. Lindsay B (1983) The geometry of mixture likelihoods: a general theory. Ann Stat 11:86–94

35. Luong ML, Al-Dabbagh M, Groll AH, Racil Z, Nannya Y, Mitsani D, Husain S (2016) Utility of voriconazole therapeutic drug monitoring: a meta-analysis. J Antimicrob Chemother 71(7):1786–1799

36. Mallet A (1986) A maximum likelihood estimation method for random coefficient regression models. Biometrika 73:645–656

37. Merle Y, Mentre F (1995) Bayesian design criteria: computation, comparison, and application to a pharmacokinetic and a pharmacodynamic model. J Pharmacokinet Biopharm 23(1):101–125

38. Neely M, Rushing T, Kovacs A, Jelliffe R, Hoffman J (2010) Voriconazole pharmacokinetics and pharmacodynamics in children. Clin Infect Dis 50(1):27–36

39. Neely M, Margol A, Fu X et al (2015) Achieving target voriconazole concentrations more accurately in children and adolescents. Antimicrob Agents Chemother 59:3090–3097

40. Pronzato L, Pazman A (2013) Design of experiments in nonlinear models: asymptotic normality, optimality criteria and small-sample properties. Springer, New York

41. Pronzato L, Walter E (1985) Robust experiment design via stochastic approximation. Math Biosci 75:103–120

42. Raiffa H, Schlaifer R (2000) Applied statistical decision theory. Wiley Classics Library, Wiley, New York

43. Ryan EG, Drovandi CC, McGree JM, Pettitt AN (2015) A review of modern computational algorithms for Bayesian optimal design. Int Stat Rev 84(1):128–154

44. Rodman JH, D'Argenio DZ, Peck CC (2004) Analysis of pharmacokinetic data for individualizing drug dosage regimens. In: Buron ME, Shaw LM, Schentag JJ, Evans WE (eds) Applied pharmacokinetics: principles of therapeutic drug monitoring, 4th edn. Applied Therapeutics, Spokane

45. Schumitzky A (1991a) Application of stochastic control theory to optimal design of dosage regimens. In: D'Argenio DZ (ed) Advanced methods of pharmacokinetic and pharmacodynamic systems analysis. Plenum Press, New York, pp 137–152

46. Schumitzky A (1991b) Nonparametric EM algorithms for estimating prior distributions. Appl Math Comput 45(2):143–157

47. Sheiner LB, Beal SL (1982) Bayesian individualization of pharmacokinetics: simple implementation and comparison with non-Bayesian methods. J Pharm Sci 71:1344–1348

48. Sheiner LB, Rosenberg B, Melmon KL (1972) Modeling of individual pharmacokinetics for computer-aided drug dosage. Comput Biomed Res 5:441–459

49. Sheiner LB, Halkin H, Peck C, Rosenberg B, Meimon KL (1975) Improved computer-assisted digoxin therapy: a method using feedback of measured serum digoxin concentrations. Ann Intern Med 82:619–627

50. Silvey SD (1980) Optimal design: an introduction to the theory for parameter estimation. Chapman and Hall, London

51. Tod M, Rocchisani J-M (1996) Implementation of OSPOP, an algorithm for the estimation of optimal sampling times in pharmacokinetics by the ED, EID and API criteria. Comput Methods Programs Biomed 50(1):13–22

52. Tod M, Rocchisani JM (1997) Comparison of ED, EID, and API criteria for the robust optimization of sampling times in pharmacokinetics. J Pharmacokinet Biopharm 25(4):515–537

53. Tse E, Bar-Shalom Y, Meier L III (1973) Wide-sense adaptive dual control for nonlinear stochastic systems. IEEE Trans Autom Control 18(2):98–108

54. Vajjah P, Duffull SB (2012) A generalisation of T-optimality for discriminating between competing models with an application to pharmacokinetic studies. Pharm Stat 11(6):503–510

55. Vozeh S, Steimer J-L (1985) Feedback control methods for drug dosage optimisation: concepts, classification and clinical application. Clin Pharmacokinet 10:457–476

56. Vozeh S, Berger M, Wenk M, Ritz R, Follath F (1984) Rapid prediction of individual dosage requirements for lignocaine. Clin Pharmacokinet 9:353–363

57. Walter E, Pronzato L (1987) Optimal experiment design for nonlinear models subject to large prior uncertainties. Am J Physiol Regul Integr Comp Physiol 253:R530–R534