Journal of Pharmacokinetics and Pharmacodynamics, Vol. 32, No. 1, February 2005 (© 2005)
DOI: 10.1007/s10928-005-2102-z

Robust Population Pharmacokinetic Experiment Design

Michael G. Dodds,[1,2] Andrew C. Hooker,[1,3] and Paolo Vicini[1,*]

Received July 10, 2003—Final February 23, 2005

The population approach to estimating mixed effects model parameters of interest in pharmacokinetic (PK) studies has been demonstrated to be an effective method in quantifying relevant population drug properties. The information available for each individual is usually sparse. As such, care should be taken to ensure that the information gained from each population experiment is as efficient as possible by designing the experiment optimally, according to some criterion. The classic approach to this problem is to design "good" sampling schedules, usually addressed by the D-optimality criterion. This method has the drawback of requiring exact advanced knowledge (expected values) of the parameters of interest. Often, this information is not available. Additionally, if such prior knowledge about the parameters is misspecified, this approach yields designs that may not be robust for parameter estimation. In order to incorporate uncertainty in the prior parameter specification, a number of criteria have been suggested. We focus on ED-optimality. This criterion leads to a difficult numerical problem, which is made tractable here by a novel approximation of the expectation integral usually solved by stochastic integration techniques. We present two case studies as evidence of the robustness of ED-optimal designs in the face of misspecified prior information. Estimates from replicate simulated population data show that such misspecified ED-optimal designs recover parameter estimates that are better than similarly misspecified D-optimal designs, and approach estimates gained from D-optimal designs where the parameters are correctly specified.

KEY WORDS: modeling; optimal experiment design; population kinetics; pharmacokinetics; simulation; ED-optimality; D-optimality.

[1]Resource Facility for Population Kinetics, Department of Bioengineering, University of Washington, Box 352255, Seattle, WA 98195-2255, USA.
[2]Pre-Clinical Development, ZymoGenetics, Inc, 1201 Eastlake Avenue, Seattle, WA 98102, USA.
[3]Division of Pharmacokinetics and Drug Therapy, Uppsala University, Biomedical Centre, Box 591, SE 751 24, Uppsala, Sweden.
*To whom correspondence should be addressed. E-mail: vicini@u.washington.edu

INTRODUCTION

Pharmacokinetic (PK) studies are carried out to understand the absorption, distribution, metabolism and excretion of compounds (either endogenous or exogenous in nature) in a living system. Inferences about the underlying behavior of the system studied are made by repeated measurements in numerous individuals, with the goal of quantifying the overall (population) response to a compound. Population experiments are often conducted in a very different manner than those carried out on a small number of individuals where rich data sets can be collected, since the resources required to execute an experiment that fully quantifies each individual are prohibitive.

The sparsity of data collected from each individual is a main feature of population studies, and represents a significant challenge in parameter identification (the inverse problem). As a result, mixed-effects models are often used when attempting to identify parameters from a population (1). These models partition sources of variability through the use of hierarchical statistical models. A key feature of mixed-effects models is the segregation of residual, unexplained variability (RUV) (measurement error, assay error and other Type-1 errors) and variability due to the differences between subjects in the study (2). This separation allows for a reduction in variance of the estimated population characteristics. Moreover, mixed-effects models alleviate the (often infeasible) need for the data to fully identify each individual.

Accuracy of parameter estimates in any modeling effort is highly dependent on experiment design. With the additional complication of sparse and costly data, a poorly designed population PK experiment can lead to inaccurate (biased) and unreliable (large variance) parameter estimates. In large-scale phase II and phase III clinical drug trials, this can lead to increased delay and cost in drug approval by the United States Food and Drug Administration (FDA) (3). As a result, optimal experimental design for population PK experiments, whose purpose is to maximize the information content in the data, is an important and relevant area of research.

Existing optimal designs for population PK studies have been investigated in the form of D-optimal design strategies by Duffull, Mallet, Mentré, Merle, Roccisani, Retout, Tod, these authors and others (4–6). However, very few software packages are available that implement these methods, principally *OSP-Fit* by Tod and Roccisani, *PFIM* by Retout, Mentré and Duffull, and *PopED* by Foracchia, Hooker, Vicini and Ruggeri (7–9). D-optimality seeks to reduce the scalarized (D-optimal: determinant) covariance of the estimated population parameters by selection of an experimental design. The invocation of the Cramer-Rao inequality provides

an asymptotic (lower) bound on the covariance of the parameter estimates as the inverse of the Fisher information matrix (FIM) (10). The FIM can be calculated given the model, prior information on the parameter values and a particular design. One criticism that has been leveled at this approach is that the requirement of exact prior information for the parameter values flies in the face of the very reason these experiments are being carried out. However, it has been demonstrated that D-optimal design quality is reasonably insensitive to minor (20–30%) prior misspecification (11,12).

D-optimality assumes nominal values for the parameters, usually gained from previous studies. If these values are subject to uncertainty, as is usually the case, or are misspecified, this design may not be optimal. Parameter values gained from studying small cohorts, such as in phase-I clinical trials, will almost certainly provide little reliable information about variability between individuals. Sequential approaches have been suggested, where small trials are carried out to further refine parameter estimates. More recently, the established D-optimality approach has been extended to incorporate uncertainty in the parameters of interest. The ED, EID and API criteria attempt to represent this uncertainty by assuming prior distributions for the PK parameters, rather than restricting them to a fixed value (7,13,14). These criteria are very appealing, but are not widely used due to, in part, the difficulties associated with optimizing the corresponding objective functions, which are generally posed as expected values of some functional of the FIM. Two packages are available, *PopED* and *OSP-Fit*, that treat this expection integral. However, they choose to treat the expectation integral via stochastic integration. We will choose a deterministic approximation in this approach.

The goal of this work is to alleviate the need for accurate prior information for the population PK parameters of interest. In clinical drug trials, parameter estimates from previous studies that provide these priors may be inappropriate (e.g. those gained from animal studies, or those gained from adult human studies when moving to pediatric therapies) or inaccurate (e.g. between-subject variability, BSV estimates gained from early trials where the number of subjects are small) but not irrelevant (15). Rather than discard this useful information, it would be advantageous to instead include the precision of that prior information as part of the design of future experiments. Here, we provide this ability by specifying the prior information on the parameters as distributions, rather than point-values, by employing the ED-optimality design criteria (ED-optimal: expectation of the determinant) (16,13). This adds a level of complexity to the problem, as evaluation of the expectation integral is generally intractable. However, we present a tractable approximation to the ED-optimal design criterion to address this issue.

BACKGROUND

Population Pharmacokinetic Modeling

We begin by developing a model for the i-th individual as part of an experiment performed on m individuals which is designed to recover a vector of p parameters (e.g. rate constants, volumes of distributions), β_i, for an underlying model, $f(x_i, \beta_i)$, thought to describe the measurement (e.g. concentration of a drug in serum), y_i, at n specified experimental variables (e.g. time) for that individual, x_i (17):

$$\begin{aligned} y_i &= f(x_i, \beta_i) + \epsilon_i \quad i = 1 \ldots m \\ \epsilon_i &\sim N(0, \boldsymbol{R}(x_i, \beta_i)) \\ x_i &= (x_{i,1}, \ldots, x_{i,n})^T \\ \beta_i &= (\beta_{i,1}, \ldots, \beta_{i,p})^T \end{aligned} \tag{1}$$

where the RUV, ϵ_i, is assumed independent and identically distributed, with mean zero and variance, $\boldsymbol{R}$, and may be heteroscedastic and dependent on the model parameters:

$$\boldsymbol{R}(\boldsymbol{x}_i, \boldsymbol{\beta}_i) = \begin{bmatrix} r_1(x_{i,1}, \boldsymbol{\beta}_i) & & 0 \\ & \ddots & \\ 0 & & r_n(x_{i,n}, \boldsymbol{\beta}_i) \end{bmatrix} \tag{2}$$

At this stage, the model is limited to a single level of statistical uncertainty, that of the RUV, and assumes that the parameter vector, β_i, is unique for each individual. This is an appropriate modeling framework for early PK experiments, where the number of individuals is small and the number of samples, n, that are taken from each individual is relatively high. While the parameter values are specific to each subject, no definitive information is available for the population at large. Averaging of the parameter estimates (two-stage analysis) is possible, but the assessment of variability in the population could be overestimated (18).

Later in drug development, investigators wish to identify both the average and the variation of each parameter across a large population. To accomplish this, we turn to mixed-effects models as a way to model individuals as samples from a larger population, thereby adding an additional statistical model level. We then extend Eqs. 1 and 2 for the i-th individual in the population by assuming that the individual parameter vector, β_i, is a function, $g(\cdot)$, of the population averages (fixed effects), β^{pop}, individual-specific anthropometric parameters (covariates), a_i, and parameters that

describe the difference between this individual and the population at large (random effects), b_i (18):

$$\begin{aligned} \beta_i &= g(\beta^{\text{pop}}, a_i, b_i), \quad b_i \sim N(0, \boldsymbol{D}) \\ \beta^{\text{pop}} &= (\beta_1^{\text{pop}}, \ldots, \beta_p^{\text{pop}})^T, \quad b_i = (b_{i,1}, \ldots, b_{i,q})^T \end{aligned} \tag{3}$$

where the q individual random-effects, b_i, are assumed independent of the RUV and are independently and identically distributed, with mean zero and variance, $\boldsymbol{D}$, that describes the spread of the parameter values across the population, or BSV:

$$\boldsymbol{D} = \begin{bmatrix} d_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & d_q \end{bmatrix} \tag{4}$$

Our population model thus includes two sources of variability, each appearing at different stages of the statistical model. Previously, specific parameter estimates were obtained for each subject. Now, each individual is considered to be a realization of the population at large. Most commonly, the so-called "mixed-effect model", $g(...)$, combines the population fixed-effects with the individual-specific random-effects in an additive or exponential manner. In the additive case, this gives rise to a normal distribution for the individual parameters, and in the exponential case, a lognormal distribution.

The task, then, is to design an experiment to estimate the parameters that do not change over the population (fixed-effects), β^{pop} (which describe the average model parameters over the entire population) and $\boldsymbol{D}$ (which describes the variance of the model parameters over the entire population):

$$\alpha = (\beta_1^{\text{pop}}, \ldots, \beta_p^{\text{pop}}, d_1, \ldots, d_q)^T \tag{5}$$

Additionally, the RUV, $\boldsymbol{R}$, could be included in the vector of fixed- effects. This, in essence, would optimize the design in such a way as to reduce the subsequent estimate variance for this parameter. Since the RUV is a confounding factor in estimating the BSV, this may be desirable in some instances (19).

Model Fitting

Parameter estimation (the inverse problem) by fitting the model to the data can be accomplished by numerous methods, many of which are based on the maximum likelihood (ML) estimator (18,20). We use $p(y_i|\ \alpha, b_i)$ to denote the probability of observing y_i for the i-th individual given that the fixed-effect vector is α and the random effect vector is b_i. The probability of y_i given α is the marginal:

$$p(y_i|\alpha) = \int p(y_i|\alpha, b_i)db_i \tag{6}$$

When M independent individuals are considered, the ML estimate, α^*, is the value that maximizes the joint log-likelihood function $L(\alpha)$:

$$\begin{aligned} L(\alpha) &= \sum_{i=1}^{M} \log p(y_i|\alpha) \\ \alpha^* &= \text{argmax}_\alpha[L(\alpha)] \end{aligned} \tag{7}$$

where argmax[·] indicates a maximization procedure that is performed on the argument, usually in the form of a gradient-based iterative optimization algorithm. It is important to note that we are only attempting to estimate the population characteristics, α, from the data and we also assume that the functional form of the model, y_i, and RUV, $\boldsymbol{R}$, are known and dependent only on the design variables and population parameters.

Once a ML estimator for α is obtained, the covariance of the estimated parameters can be evaluated. The Cramer-Rao inequality places an asymptotic (lower) bound on the covariance matrix through the FIM (10):

$$\begin{aligned} \text{Cov}[\alpha] &\geq F(\alpha, x)^{-1} \\ F(\alpha, x) &= E_y\left[\left(\frac{\partial}{\partial\alpha}L(\alpha, x)^T\right)\left(\frac{\partial}{\partial\alpha}L(\alpha, x)\right)\right] \end{aligned} \tag{8}$$

where Cov[·] indicates the covariance matrix of the argument, E_y[·] indicates the expectation of the argument with respect to the experimental observations, y_i, and "≥" indicates an inequality, in a matrix sense (21).

The computationally burdensome ML estimator must be simplified when, as in our case, the underlying model is nonlinear and complex. We assume that y_i given α and b_i is normally distributed and the second partial derivative of $p(y_i|\ \alpha,\ b_i)$ with respect to b_i is zero. With this assumption, the ML estimator is equivalent to the extended least squares (ELS) estimator (18,20,22,23):

$$\begin{aligned} L(\alpha, x)_{\text{ELS}} &= \sum_{i=1}^{M}\left[\frac{1}{2}\log[\det[2\pi \boldsymbol{V}(\alpha, x)]] + \frac{1}{2}[y_i - h(\alpha, x)]^T \boldsymbol{V}(\alpha, x)[y_i - h(\alpha, x)]\right] \\ h(\alpha, x) &= \text{E}_y[y_i] \\ \boldsymbol{V}(\alpha, x) &= \text{Cov}[y_i] \end{aligned} \tag{9}$$

where det[·] and log[·] indicate the determinant and natural logarithm, respectively, of the argument.

To ensure the ELS estimator assumptions are met, we linearize the model using a 1st-order Taylor series around the between-subject random-effects (20,24):

$$\begin{aligned} y_i &\approx f\left(x_i, g_i(\beta^{\text{pop}}, a_i, b_i)|_{b_i=0}\right) + \boldsymbol{B}(x_i, \beta^{\text{pop}}, a_i) \cdot b_i + \epsilon_i \\ \boldsymbol{B}(x_i, \beta^{\text{pop}}, a_i) &= \left(\frac{\partial}{\partial b_i} f(x_i, g_i(\beta^{\text{pop}}, a_i, b_i))\right)^T |_{b_i=0} \end{aligned} \tag{10}$$

which then assures us that the population model is independently (in ε and b) and identically distributed (normal), with mean and variance:

$$\begin{aligned} h(\alpha, x) &= E_y[y_i] \approx f\left(x_i, g_i(\beta^{\text{pop}}, a_i, b_i)|_{b_i=0}\right) + \boldsymbol{B}(x_i, \beta^{\text{pop}}, a_i) \cdot b_i \\ \boldsymbol{V}(\alpha, x) &= \text{Cov}[y_i] \approx \boldsymbol{R}\left(x_i, g_i(\beta^{\text{pop}}, a_i, b_i)|_{b_i=0}\right) \\ &\quad + \boldsymbol{B}(x_i, \beta^{\text{pop}}, a_i) \cdot \boldsymbol{D}(d_1, \ldots, d_q) \cdot \boldsymbol{B}(x_i, \beta^{\text{pop}}, a_i)^T \end{aligned} \tag{11}$$

Under these assumptions, the FIM becomes (25):

$$\begin{aligned} F(\alpha, x) = \frac{1}{2}\left[\frac{\partial}{\partial \alpha} \boldsymbol{V}(\alpha, x)\right]^T [\boldsymbol{V}^{-1}(\alpha, x) \otimes \boldsymbol{V}^{-1}(\alpha, x)] \frac{\partial}{\partial \alpha} \boldsymbol{V}(\alpha, x) \cdots \\ + \left[\frac{\partial}{\partial \alpha} h(\alpha, x)\right]^T \boldsymbol{V}^{-1}(\alpha, x) \frac{\partial}{\partial \alpha} h(\alpha, x) \end{aligned} \tag{12}$$

where "⊗" indicates the Kronecker product. This formulation differs only slightly from that proposed by Mentré and Duffull, in that the variance term, $\boldsymbol{V}$, retains dependencies on the fixed-effects that appear in the model (β^{pop}), through $\boldsymbol{R}$ retaining dependence on these variables and individual covariates (8, 9). Merlé and Tod studied the effects of assuming a constant variance term, $\boldsymbol{V}$, and noted only small discrepancies between the approximated FIM and a reference FIM in the between-subject parameters (26). However, they did note differences between the approximated FIM and the reference FIM when comparing between designs.

Optimal Experimental Design

D-optimality

The most widely used criterion is D-optimality, which consists of minimizing the determinant of the inverse FIM (3,24,27). This criterion has three features that make it appealing: computational feasibility, sound theoretical backing, and a geometric interpretation of minimizing the volume of the joint asymptotic confidence region for the parameters of

interest. Moreover, for our purposes, it will serve as a good benchmark in developing ED-optimality (13,28).

A design, x^{D}, is said to be D-optimal when it minimizes the negative determinant of the FIM for a given set of parameter values, α (21):

$$\begin{aligned} j^{\mathrm{D}}(x) &= \det[F(\alpha, x)] \\ x^{\mathrm{D}} &= \operatorname{argmin}_x\left[-j^{\mathrm{D}}(x)\right] \end{aligned} \tag{13}$$

ED-optimality

The ED-optimality criterion attempts to search for an experimental design that is best in the weighted average sense (16). That is, the parameters of interest are assigned a prior distribution, and an expectation is performed on the determinant of the FIM with respect to the joint probability of the prior. A design, x^{ED}, is said to be ED-optimal if it minimizes the negative expected (E_α) determinant of the FIM with respect to the parameter priors, α (21):

$$\begin{aligned} j^{\mathrm{ED}}(x) &= E_\alpha[\det[F(\alpha, x)]] \\ &= \int_{-\infty}^{+\infty} p(\alpha) \cdot \det[F(\alpha, x)] d\alpha \\ x^{ED} &= \operatorname{argmin}_x\left[-j^{\mathrm{ED}}(x)\right] \end{aligned} \tag{14}$$

where $E_\alpha[\cdot]$ represents the expectation of the argument which results in a multidimensional integral and $p(\alpha)$ is the probability density function that describes the prior information on the parameters.

This leads to a multidimensional integral to evaluate the expectation. The expectation will need to be evaluated many times, as the design variables are optimized. Numerical techniques can be used to make this evaluation, but the amount of computation required at each step is large, and grows with the dimension of α. Direct solution of the original integral can approximated by numerical integration (Gaussian quadrature) or, as recently proposed, Monte Carlo techniques (29). Neither of these techniques are suitable, at the current level of technology, for use in optimization. As such, an approximation to the expectation which may lead to more expedient calculations is desirable.

METHODS

Laplace's Approximation to Exponential Integrals

We borrow heavily from Bayesian methods to arrive at an accurate approximation of the expectation in our objective through application of

Laplace's Method. Originally published in 1774 by Laplace, and more recently applied to Bayesian posterior moments and marginal densities by Tierney and Kadane, this method replaces the expectation integral with an optimization step (30,31). Recently, this approximation has been described for marginal likelihood estimation for mixed effects models by Beal and Sheiner, which has been discussed in this context by Vonesh, Bell and others (25,22,32). This approximation potentially represents a large reduction in computational burden, especially as the dimension of α grows. Laplace's method for approximating exponential integrals can be written:

$$\begin{aligned} &\int_{-\infty}^{+\infty} e^{-k(\alpha,x)} d\alpha \approx \det\left[\partial_\alpha^2 k(\alpha^m, x)/(2\pi)\right]^{-1/2} \cdot e^{-k(\alpha^m,x)} \\ &\alpha^m = \operatorname{argmin}_\alpha [k(\alpha, x)] \end{aligned} \tag{15}$$

where α^m minimizes $k(\alpha,x)$ (equivalent to maximizing the negative function, as in the original formulation) with respect to α given x. This approximation is exact when the Hessian term $\partial_\alpha^2 k(\alpha^m, x)$ is positive definite, which is the case when $k(\alpha,x)$ is a strongly convex quadratic function of α. This approximation of our criterion (Eq. (14)) requires the calculation of the second derivatives of the transformed integrand, in contrast with other methods requiring computation of third- and perhaps higher-order terms (29).

We now recast Eq. (14) in these terms to arrive at an approximate optimality objective:

$$\begin{aligned} j^{\mathrm{ED}}(x) &= E_\alpha[\det[F(\alpha, x)]] \\ x^{\mathrm{ED}} &= \operatorname{argmin}_x \left[-j^{\mathrm{ED}}(x)\right] \\ &= \operatorname{argmin}_x \left[-\int_{-\infty}^{+\infty} p(\alpha) \cdot \det[F(\alpha, x)] d\alpha\right] \\ &= \operatorname{argmin}_x \left[-\int_{-\infty}^{+\infty} e^{-k(\alpha,x)} d\alpha\right] \\ k(\alpha, x) &\equiv -\log[p(\alpha) \cdot \det[F(\alpha, x)]] \\ x^{\mathrm{ED}} &\approx \operatorname{argmin}_x \left[-\det[\partial_\alpha^2 k(\alpha^m, x) \cdot (2\pi)^{-1}]^{-1/2} \cdot e^{-k(\alpha^m,x)}\right] \\ \alpha^m &= \operatorname{argmin}_\alpha [k(\alpha, x)] \end{aligned} \tag{16}$$

where $p(\alpha)$ is the prior information on the parameters of interest. The expectation integral has been replaced by a nested optimization problem in α, which generally can be expected to have a much lower dimension than the design variables (x^{ED}).

It is worthwhile to pause and consider the posed problem at a conceptual level. At each design set (the "outer" optimization problem) iteration, a vector of α^m is selected, again iteratively (the "inner" optimization problem), that maximizes the product of the prior information and scalarized FIM. So, the selection of α^m is based on both the information in the experiment and the prior.

We replace the second derivative, $\partial_\alpha^2 k(\alpha^m, x)$, with a finite difference approximation, which is consistent with the requirement of Laplace's approximation that $k(\alpha,x)$ be a strongly convex quadratic function of α (25).

Models

To test the robustness of our designs, we develop a simulation strategy for the designed experiments, then carry out parameter estimation on the simulated data and compare the estimates with the known values that drove the simulations. D-optimality is used as a baseline comparison, as this method is established in the literature, and reflects the current state-of-the-art in optimal experiment design

Population Model 1

Our first model is of the form commonly referred to as a "two-exponential" model, where the disappearance of a drug from blood plasma follows a bi-phasic exponential decay. The prior information we begin with when selecting an optimal design is loosely based on a previous work, which indicates a two-exponential model accurately captures the appearance and clearance of low-density lipoprotein (LDL) in blood plasma (33). We applied an additive mixed-effect model to the model parameters, resulting in four population average parameters ($\beta_1^{\text{pop}}, \ldots, \beta_4^{\text{pop}}$) and four subject-specific random-effect ($b_1, \ldots, b_4$) parameters. The between-subject parameters were modeled as samples from the overall population, mean zero and variance described by four BSV parameters ($d_1, \ldots, d_4$), as described in the Background of this work. The values of the BSV parameters were chosen arbitrarily. A log-transformation was applied to the BSV parameters. The RUV was constant throughout the study. Forty simulated subjects ($i = 1,\ldots,40$) were considered for each experiment, from which two samples (x_i) could be extracted. The time course of the study was from 0–14 days, thus providing bounds on the feasible sampling times for each individual (x_i). The parameters of interest in this study were the four central tendencies for the parameters and the related between-subject variance parameters, for a total of eight fixed effects parameters overall (α).

Mathematically, this experiment can be represented as:

$$
\begin{aligned}
y_i(\beta_i, x_i) &= \beta_{i,1} \cdot e^{-\beta_{i,2} \cdot x_i} + \beta_{i,3} \cdot e^{-\beta_{i,4} \cdot x_i} + \varepsilon_i \\
\beta_i &= (\beta_1^{\text{pop}} + b_{i,1}, \ldots, \beta_4^{\text{pop}} + b_{i,4}) \\
b_i &\sim N(0, \boldsymbol{D}) \\
\varepsilon_i &\sim N(0, \boldsymbol{R}) \\
\boldsymbol{D} &= \begin{bmatrix} e_1^d & & \cdots & 0 \\ & e^{d_2} & & \\ \vdots & & e^{d_3} & \\ 0 & \cdots & & e^{d_4} \end{bmatrix} \\
\boldsymbol{R} &= \begin{bmatrix} 250000 & & & 0 \\ & & \ddots & \\ & 0 & & 250000 \end{bmatrix} \\
\alpha &= [\beta_1^{\text{pop}}, \ldots, \beta_4^{\text{pop}}, \log(d_1), \ldots, \log(d_4)]
\end{aligned}
\tag{17}
$$

Population Model 2

Our second PK model comes from another commonly encountered PK model: the one-compartment, linear absorption model. Here, the model describes the delayed appearance of a drug in the plasma, often an oral dose, followed by a mono-phasic elimination from the plasma. The prior information we use is somewhat informed by a previous analysis of an anti-asthmatic drug, theophylline (22). We applied a exponential mixed-effect model to capture the three ($j = 1,\ldots,3$) population average parameters (β_j^{pop}) and three subject-specific random-effects (b_i). The between-subject parameters were modeled as samples from overall population, mean zero and variance described by three BSV parameters ($d_1, \ldots, d_3$). The BSV parameter values were chosen arbitrarily. A log-transformation was applied to both the population average and between-subject variance parameter. The RUV was proportional to the model output by a known amount. Thirty-six subjects ($n = 36$) were considered for each experiment, from each of which two samples (y_i) could be extracted. Each individual is additionally characterized by a covariate, dose over body weight (a_i). Covariates for 12 individuals were used (under the assumption of normality) to generate 24 additional synthetic covariates, bringing the number of subject in the simulated trial to 36. The time course of the study was from 0 to 24 hr, thus placing bounds on the

feasible sampling times for each individual (x_i). The parameters of interest in this study were the three population parameters and the related BSV parameters, totaling six parameters overall (α).

Mathematically, this experiment can be represented as:

$$
\begin{aligned}
&y_i(\beta_i, x_i) = f_i(\beta_i, x_i) + \varepsilon_i \\
&f_i(\beta_i, x_i) = \frac{a \cdot \beta_{i,1} \cdot \beta_{i,2}}{\beta_{i,3} \cdot (\beta_{i,1} - \beta_{i,2})} (e^{-\beta_{i,2} \cdot x_i} - e^{-\beta_{i,1} \cdot x_i}) \\
&a = \begin{bmatrix} 4.02, 4.44, 4.53, 4.4, 5.86, 4.0, 4.95, 4.53, 3.1, 5.5, 4.92, 5.3, 5.1, 4.18, \ldots \\ 5.04, 3.8, 4.69, 3.13, 4.26, 4.97, 4.39, 5.55, 4.15, 2.89, 3.71, 5.41, 4.54, \ldots \\ 4.91, 5.33, 3.04, 4.14, 4.1, 3.87, 4.49, 5.76, 4.6 \end{bmatrix} \\
&\beta_i = (e^{\beta_1^{\text{pop}}} \cdot e^{b_{i,1}}, \ldots, e^{\beta_3^{\text{pop}}} \cdot e^{b_{i,3}}) \\
&b_i \sim N(0, \boldsymbol{D}) \\
&\varepsilon_i \sim N(0, \boldsymbol{R}) \\
&\boldsymbol{D} = \begin{bmatrix} e^{d_1} \cdots 0 \\ \vdots \ e^{d_2} \ \vdots \\ 0 \cdots e_3^d \end{bmatrix} \\
&\boldsymbol{R} = \begin{bmatrix} (0.05\, f_1)^2 & 0 \\ & \ddots \\ 0 & (0.05\, f_n)^2 \end{bmatrix} \\
&\alpha = [\log(\beta_1^{\text{pop}}), \ldots, \log(\beta_3^{\text{pop}}), \log(d_1), \ldots, \log(d_3)]
\end{aligned} \tag{18}
$$

Optimal Designs

To demonstrate the utility of our ED-optimality design approach, three optimal designs were established for each model. Two D-optimal designs were created, one at the "true" values of the parameters (*Dt*) and one with severe misspecification of the parameters (*Dm*). Here, *we mean "true" in the sense that these values will be used to simulate population experiments for all design cases*. An experiment conducted using the *Dt* design should then be optimal by construction, since is assumes knowledge of the true parameter values. Conversely, an experiment conducted using the *Dm* design should perform very poorly when estimating the parameters. These two experimental designs should then provide the high and low benchmark, respectively, when examining the performance of ED-optimal designs.

Table I. Model 1 Case Study Prior Summary

Parameter	True value	Dt priors		Dm priors		ED priors		SD
		Mis %	μ	Mis %	μ	Mis %	μ	
β_1^{pop}	31000	0%	31000	150%	46500	150%	46500	7750
β_2^{pop}	0.61	0%	0.61	200%	1.22	200%	1.22	0.305
β_3^{pop}	14000	0%	14000	300%	42000	300%	42000	14000
β_4^{pop}	0.13	0%	0.13	400%	0.52	400%	0.52	0.195
$\log(d_1)$	18.28	0%	18.28	150%	18.68	150%	18.68	0.2027
$\log(d_2)$	−3.396	0%	−3.396	200%	−2.703	200%	−2.703	0.3466
$\log(d_3)$	16.69	0%	16.69	250%	17.6	250%	17.6	0.4581
$\log(d_4)$	−6.49	0%	−6.49	300%	−5.39	300%	−5.39	0.5493

The *Dt* priors (μ) were, by definition, assigned from the values used to simulate all data sets. The *Dt* priors, in normal space, were positively increased by the specified percentage (Mis %) to yield new parameter priors for the *Dm* case. The *ED* priors were defined as normal, mean equal to the *Dm* point-values, and standard-deviations (SD) equal to half of the difference between the *Dt* and *Dm* priors.

Table II. Model 2 Case Study Prior Summary

Parameter	True value	Dt priors		Dm priors		ED priors		SD
		Mis %	μ	Mis %	μ	Mis %	μ	
$\log(\beta_1^{pop})$	0.3038	0%	0.3038	150%	0.7093	150%	0.7093	0.4055
$\log(\beta_2^{pop})$	−2.573	0%	−2.573	200%	−1.880	200%	−1.880	0.6931
$\log(\beta_3^{pop})$	−3.289	0%	−3.289	125%	−3.066	125%	−3.066	0.2231
$\log(d_1)$	−0.9352	0%	−0.9352	200%	−0.2421	200%	−0.2421	0.6931
$\log(d_2)$	−3.990	0%	−3.990	250%	−3.074	250%	−3.074	0.9163
$\log(d_3)$	−3.154	0%	−3.154	300%	−2.055	300%	−2.055	1.099

The *Dt* priors (μ) were, by definition, assigned from the values used to simulate all data sets. The *Dt* priors, in normal space, were positively increased by the specified percentage (Mis %) to yield new parameter priors for the *Dm* case. The *ED* priors were defined as normal, mean equal to the *Dm* point-values, and standard-deviations (SD) equal to the difference between the *Dt* and *Dm* priors.

To construct the priors for the misspecified D-optimal design (*Dm*), we applied a positive percent increase to the *Dt* priors in normal-space. Tables I and II describe the exact values used for Model 1 and 2, respectively. D-optimal design has been noted to be insensitive to minor misspecification, and so this approach creates D-optimal designs (*Dm*) that should be poor. Transitioning from phase-I to phase-II trials, where the between-subject prior information underestimates the variability in the overall population, may create such misspecifications. Priors taken from animal studies and used to design phase-I trials may be likewise quite misspecified. It is, however, unlikely that every parameter in the system is

so misspecified. However, here, we wish to create D-optimal designs that should "fail", but will nonetheless serve as low benchmark.

ED-optimal designs were created (*ED*), each assuming a normal distribution for the parameters with means equal to the values used in the *Dm* design. *These priors should be interpreted as reflecting uncertainty, not population distribution*. In the case of Model 1, the BSV parameters were log-transformed, so the distribution is lognormal. In the case of Model 2, every parameter was log-transformed so as to be distributed lognormally. During the search of the integrand within Laplace's approximation, α^m minimizes $k(\alpha, x)$, care was necessary to assure that negative realizations of the parameters, particularly the variance parameters, were not allowed by the priors. Standard deviations for the priors were chosen so as to include the true values of the parameters within their distributions (in log-space), but with differing probability. The priors used in Model 1 included the true values at two standard deviation. The priors used in Model 2 included the true values at one standard deviations. *In this manner, the ED design has no advantage over the misspecified Dm design, except by defining our prior knowledge as a distribution instead of a point-value*.

D-Optimal Designs

The algorithm for finding D-optimal design values (x^{D}) is as follows. Initial values for the design points were selected to be balanced and evenly spaced across the admissible design domain. Point-values for the population parameters (α) were given as either the true values (*Dt*) or the misspecified values (*Dm*). The FIM was evaluated at this initial design. A combined trust-region/interior point method for optimization subroutine (Bell and Burke, unpublished results) was passed this design, objective information (model and derivatives required to evaluate the FIM), and convergence criterion (relative difference between each design value between iterations less than 10^{-3}). Upon termination, the design values were inspected to ensure that the design points fell into m clusters, where m is equal to the number of population average parameters (β_i^{pop}). This is a well known property of individual D-optimal designs where the model response is expected to be the same for each individual (no covariates) (34). However, this rule has only been demonstrated for exponential models in the context of optimizing individual experiments. If some design points fell outside the clusters, a quantitative comparison of the contribution to the FIM for the design points for those individuals was made against the FIM for individuals with design points within the clusters. In every circumstance, local minima were always the cause for this problem. If the design points were revealed to be sub-optimal, the design points

were replaced with points within the clusters. The interior point routine was reinitialized with the new design set, which then produced another design that was again inspected. Generally, only two of these inspections were required to produce designs that assigned design points from the m clusters to all individuals.'

ED-Optimal Designs

The algorithm for finding the ED-optimal design values (x^{ED}) is as follows. Initial values for the design points were selected to be equal to the design points found for the misspecified D-optimal design (*Dm*). Any change from the *Dm* design can be directly attributable to the specification of the prior as a density, rather than the point-values used in the D-optimal design. A combined trust-region/interior point method for optimization subroutine was passed this initial design, objective information (model, derivatives required to evaluate $k(\alpha,x)$ and convergence criterion). Within this outer optimization problem (x^{ED}), an inner optimization problem was solved (α^m that minimizes $k(\alpha,x)$ with respect to α given our current x^{ED}). Other investigators in this area have noted that as the prior distributions become broader, there appears to be less clustering of design points (28). If this design set, x^{Dm}, is "far" from the final design, x^{ED}, conditional on these priors, the optimization may fail because of numerical problems. Essentially, the inner optimization problem may select a value from the priors that leads to an ill-conditioned FIM (Laplace's approximation relies on a stable second derivative of the FIM). The solution here is to start with much narrower priors that more closely resemble the point-value priors used in the *Dm* case and change them incrementally until the target prior is achieved. Procedurally, we reduce the prior breadth by one-half until the problem stabilizes and a new design is found for the narrower priors. From here, the priors are broadened slightly in an iterative fashion until the target priors are achieved. This methodology is in alignment with the actual procedure an investigator would use when generating ED-optimal designs: first, create a D-optimal design using point-values for priors, then increase the prior breadth until they match the relevant uncertainty.

Design Quality Assessment

Experiment Simulation

For the purposes of our simulations, the true value for each parameter (α) was used. Samples were drawn for each between-subject random effect (b_i) and within-subject random-effect (ϵ_i), whose distributions were

defined as normal, mean zero, and variance given by $\boldsymbol{D}$ and $\boldsymbol{R}$, respectively. The seed value used to initialize the random sampling algorithm was fixed so as to provide the same individual's random effect across the three designs. Five hundred studies were simulated for each design, so as to compare the performance of each design in recovering the true value of the parameters. Data were censored to exclude realizations of parameters outside of physiological significance (negative β_i) for Model 1 and negative data realizations (y_i). In this case, the entire simulated population dataset was re-sampled.

Parameter Estimation

Estimation was performed using the first order approximation to the ML function within NONMEM V (UCSF and GloboMax) (22). Initial parameter values for the optimization were selected to be exactly the prior values used in each design case. For the *Dt* designs, these were the true values of the parameters. For the *Dm* and *ED* designs, the misspecified point-value priors and central tendencies of the priors were selected, respectively. In practice, an estimation step performed after a trial will likely use these values as initial guesses. *By choosing the misspecified values for the initial guesses, the estimation step can only recover the true values of the parameter by virtue of the quality of the data gained by the design*. Parameter estimates from each of the 500 studies for each of the three designs were obtained with no failures in estimation. However, in some cases, estimate precisions were not available. Most often, this was caused by a non-invertible Hessian of the likelihood. The frequency of this failure can be interpreted as the rate at which any one experiment using this design will provide unreliable estimates. For those estimates for which parameter precisions were available, we visualized the estimates and estimate coefficients of variation (%CV) for each design.

RESULTS

Computational Considerations

All of the designs, simulations and back-estimation problems were computed on a modern desktop computer (AMD Athlon XP, 2.2 Ghz). The D-optimal designs were delivered in on the order of tens of minutes. The ED-optimal designs were delivered in on the order of hours. This inflation is principally due to the difficult numerical problem that requires an additional two levels of partial derivatives to be computed. The simulated data sets were relatively trivial to compute, and were delivered in on the order of minutes. The parameter estimation step, as it involved

hundreds of population PK problems, was quite time-consuming, each taking approximately 2 min to complete. Each design for each problem required on the order of 12 hr of computing time. While this may seem a quite lengthly process, when it is considered in the overall context of clinical trials, the time required to carry out these simulation methods is relatively negligible.

Population Model 1

The parameters used in our *Dt* design and data simulation were $\beta^{pop} = [3.1\times10^4,\ 0.61,\ 1.4\times10^4,\ 0.13]^T$ and $\log(d) = [18.28,\ -3.396,\ 16.69,\ -6.49\ ]^T$. To construct the priors for the misspecified D-optimal design (*Dm*), we applied a positive percent increase to the *Dt* priors, resulting in a vector of priors whose values were $\beta^{pop} = [4.65\times10^4,\ 1.22,\ 4.2\times10^4,\ 0.52]^T$ and $\log(d) = [18.68,\ -2.703,\ 17.6,\ -5.39\ ]^T$. The ED-optimal design priors had the *Dm* values as means, and standard deviations defined relative to the degree of misspecification of the parameters (Table I). Figure 1 summarizes the prior values for each of the three design for each of the eight parameters in this case study.

D-optimal design for this model at the true values of the parameters (*Dt*) yields samples at 5×10^{-6}, 0.667, 6.17 and 14 days. A balanced design was selected, where half of the subjects are sampled at 5×10^{-6} and 0.667 days and the other at 6.17 and 14 days. Figure 2 shows an example set of a simulated data superimposed on the "typical" population prediction. The low value of 5×10^{-6} days here is an artifact of the numerical method used to solve this problem, as selection of zero for a sampling time yields an unstable objective due to central differencing, and should be interpreted as a directive to sample as early as possible. The high value selected is at the maximum allowable time point. These time points represent the optimal sampling times for this experiment, and will allow for the greatest possible precision in estimating the parameters, given the number of subject, available number of samples, and sampling time constraints. D-optimal design at the misspecified priors (*Dm*) yields sample times of 5×10^{-6}, 0.330, 2.30 and 4.92 days. This is to be expected, as the priors used for *Dm* indicate a more rapid decay of the exponential terms.

ED-optimal design for this model, centered at the misspecified values yield sampling times of 5×10^{-6}, 0.234, 7.07 and 14.0 days. Again, we note a sampling time at the allowed minimum for this experiment. The intermediate sampling times are one smaller (0.234 vs. 0.667) and the other larger (7.07 vs. 6.17) than the *Dt* design. The central tendency of the priors indicate faster-decaying exponentials, but the potential for slower decay rates is allowed by the priors. Recall that ED-optimality selects a

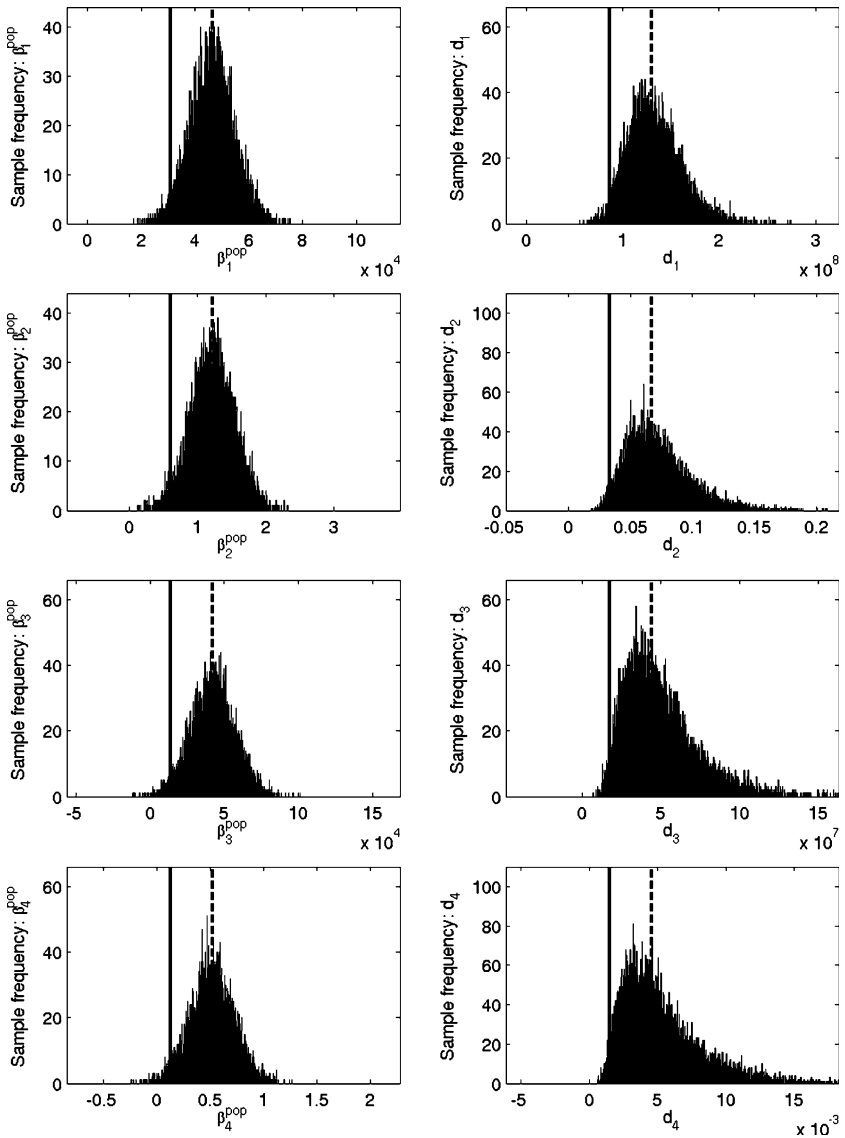

Fig. 1. Histogram of the parameter priors used in the Model 1 case study. The population average parameter priors appear in the left column, and the between-subject variance parameters appear in the right column. The true parameters are used as priors in the *Dt* case (solid vertical line), and a positive increase was applied to these values to create the *Dm* priors (dotted vertical line). The *ED* priors (histogram) have as their central tendencies the *Dm* priors, and are broad enough to include the true values of the parameters within two standard deviations of the central tendencies.

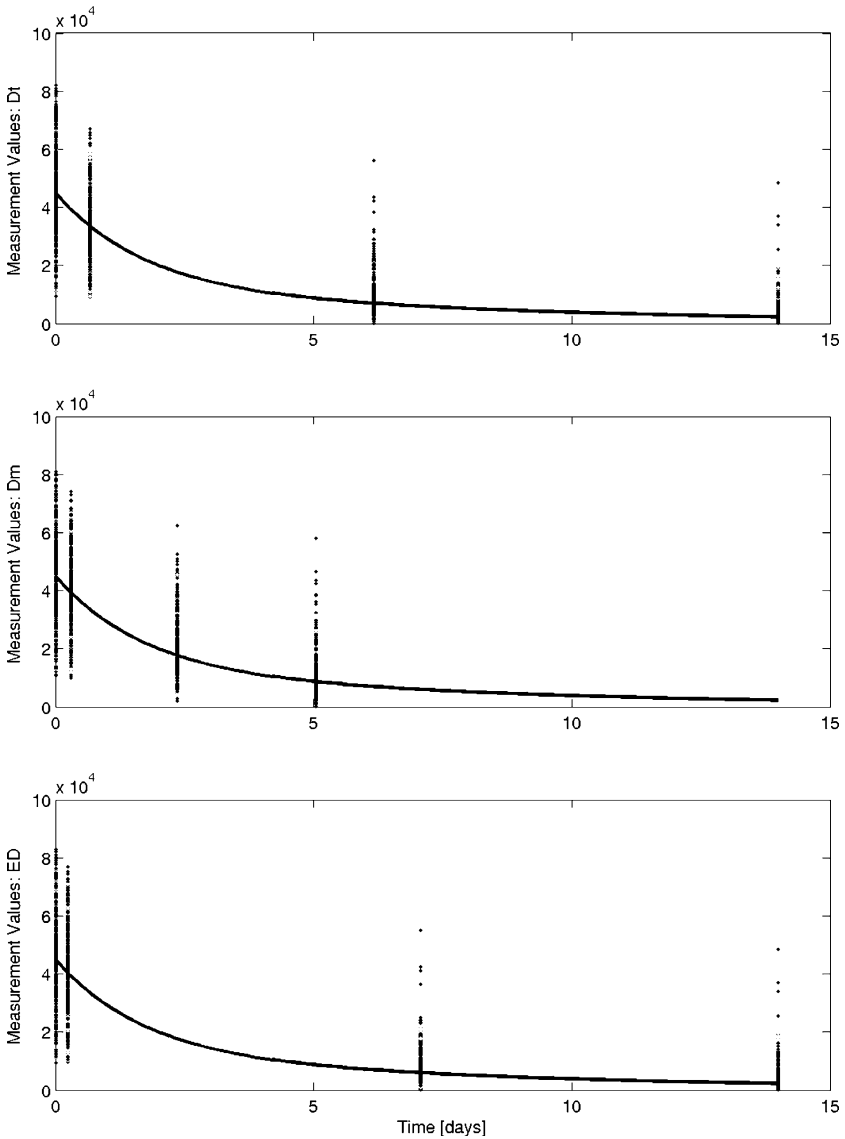

Fig. 2. Model 1 case study designs for (top panel to bottom panel) D-optimal designs (*Dt* and *Dm*) and ED-optimal design (*ED*). "Typical" population model responses (solid line) are plotted against time. Additionally, one of the 500 simulated population data sets used for parameter estimation are shown, to get a sense of the variability in the data (dots).

design that is best for the weighted average of the parameter priors. Thus, sampling at earlier times in the experiments is indicated because of the central tendencies of the priors, but some accounting for outliers in the priors must be made. The second time point is sooner than even the *Dm* design indicates, but the third time point is later than the *Dt* design would indicate.

Estimates for each of the eight parameters (β^{pop} and d) were available from the 500 replicates of each of the three experiments. The *Dt*, *Dm* and *ED* designs generated 495 (99%), 486 (97%) and 494 (99%) experiments, respectively, where estimate precisions could be calculated. Parameter estimates, Fig. 3, and%CV's, Fig. 4, are shown as box plots for those experiments where parameter precisions could be calculated. The parameter estimates (Figure 3) appear to be reasonably similar when comparing the *Dt* and *ED* designs. In contrast, the estimates derived from the *Dm* design appear markedly different. The amplitude term estimates (β_1^{pop} and β_3^{pop}) are asymmetrically distributed and negatively correlated (one higher, one lower than their true values). The slower of the two exponential decay rates (β_4^{pop}) is systematically underestimated. The%CV's (Fig. 4) show that the *Dm* design provides much less accurate estimates than either the *Dt* or *ED* design. Recall that the objective of D- and ED-optimal design is to reduce variance in our final parameter estimates.

Population Model 2

Design creation methodology parallels that used in the previous two-exponential case study. The *Dt* design priors correspond to the values used in the original publication, and were $\log(\beta^{pop}) = [0.3038, -2.573, -3.289]^T$ and $\log(d) = [-0.9352, -3.990, -3.154]^T$. To construct the priors for the *Dm* design, we applied positive percent increases to the *Dt* priors, whose values were $\log(\beta^{pop}) = [0.7093, -1.880, -3.066]^T$ and $\log(d) = [-0.2421, -3.074, -2.055]^T$. The ED-optimal prior distributions had the *Dm* values as their means, and standard deviations defined relative to each parameter's degree of misspecification (Table II). Figure 5 summarizes the prior information for each of the designs.

The inclusion of a covariate for each individual (dose/weight), would typically generate designs that are "tailored" to each individual. However, the measurement error model in this case is proportional to the model output. As the covariate appears as an amplitude term, and the model error (and therefore information available from sampling each individual) is proportional to the amplitude, these factors cancel out, yielding designs that are the same for every individual.

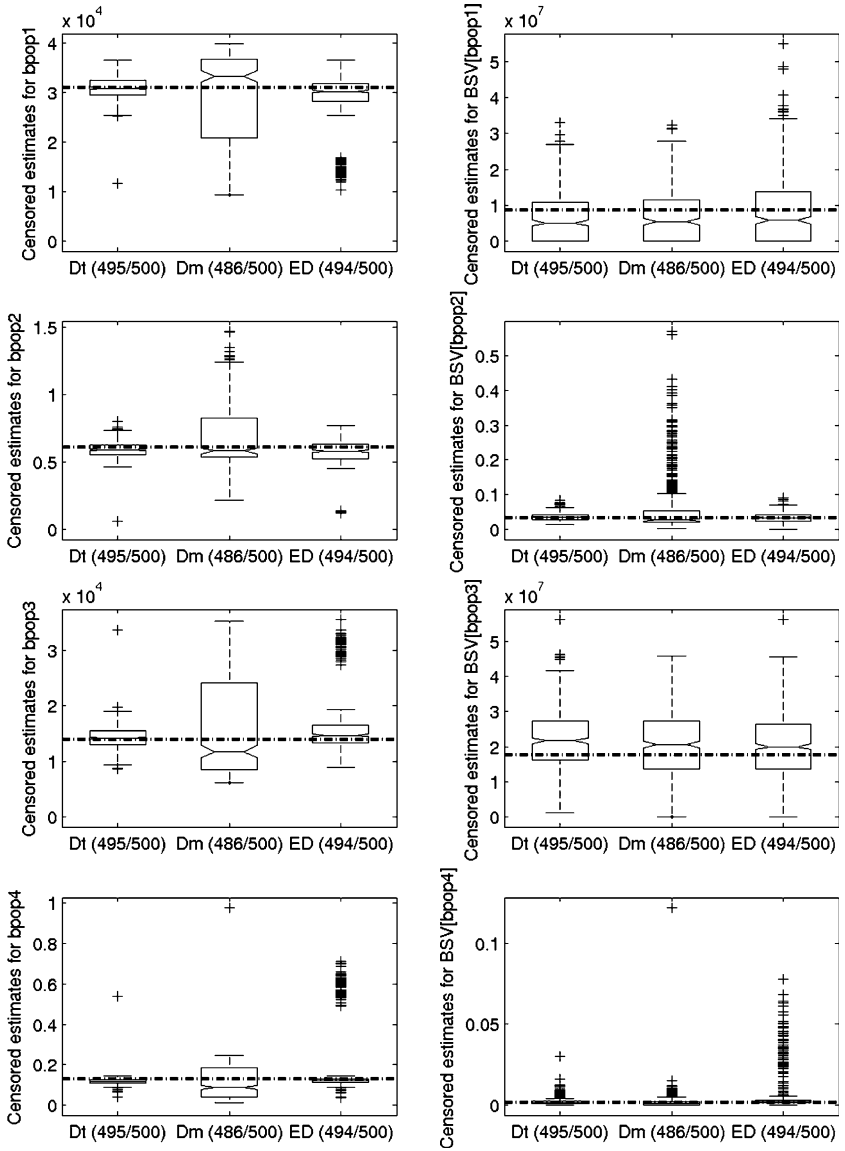

Fig. 3. Parameter estimate results for the subset of the 500 simulated experiments for the Model 1 case study for which estimate precisions were available. Population average parameters are arranged in the left column, between-subject variance in the right. The true parameter values are denoted as a dashed line, and the designs are ordered (left to right) *Dt*, *Dm*, and *ED*, with the number of estimates (out of 500) included in these (abscissa) figures. These box plots show the median of the estimates (notch), interquartile range (box extent), 1.5 times the 25% and 75% quartiles (whiskers), and any outlier estimates beyond 1.5 times the interquartile range (pluses).

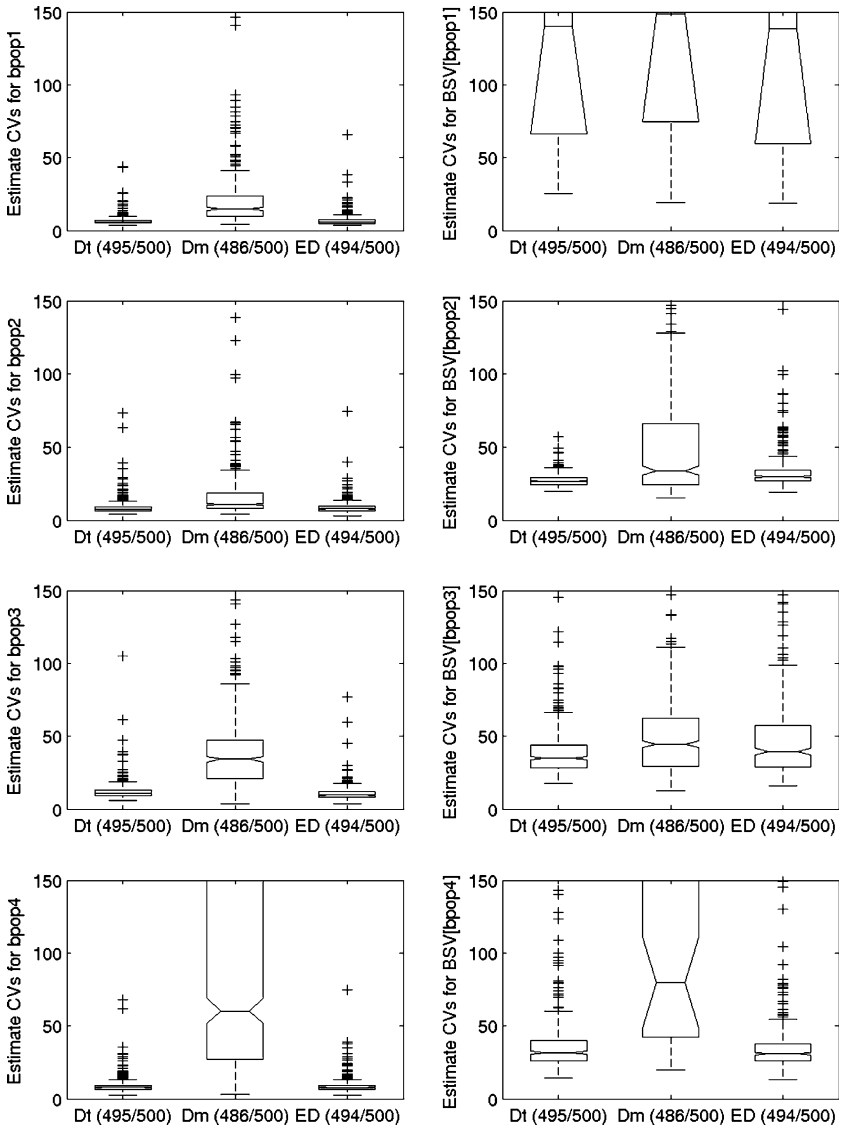

Fig. 4. Parameter estimate coefficients of variation (%CVs) for the subset of the 500 simulated experiments for the Model 1 case study for which estimate precisions were available. Population average parameter %CVs are arranged in the left column, between-subject variance %CVs in the right. The designs are ordered (left to right) *Dt*, *Dm*, and *ED*, with the number of estimates (out of 500) included in these (abscissa) figures. These box plots show the median of the estimates (notch), interquartile range (box extent), 1.5 times the 25% and 75% quartiles (whiskers), and any outlier estimates beyond 1.5 times the interquartile range (pluses).

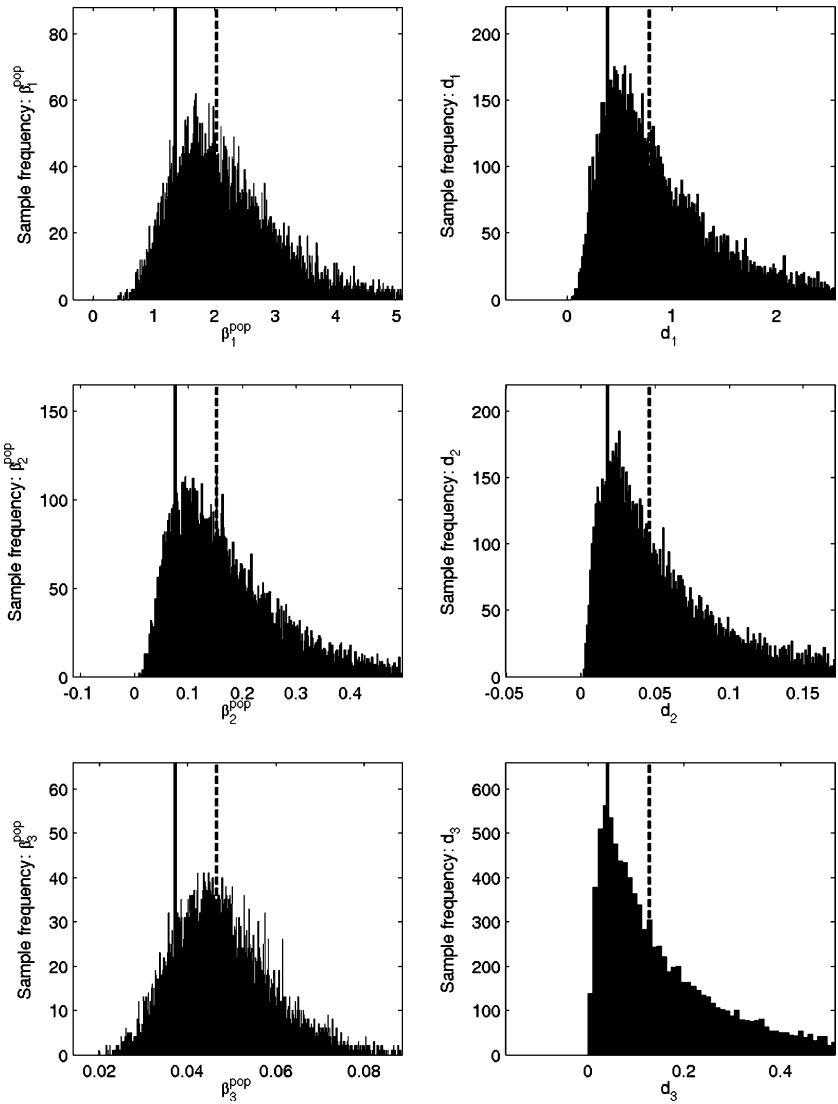

Fig. 5. Histogram of the parameter priors used in the Model 2 case study. The population average parameter priors appear in the left column, and the between-subject variance parameters appear in the right column. The true parameters are used as priors in the *Dt* case (solid vertical line), and a positive increase was applied to these values to create the *Dm* priors (dotted vertical line). The *ED* priors (histogram) have as their central tendencies the *Dm* priors, and are broad enough to include the true values of the parameters within one standard deviations of the central tendencies.

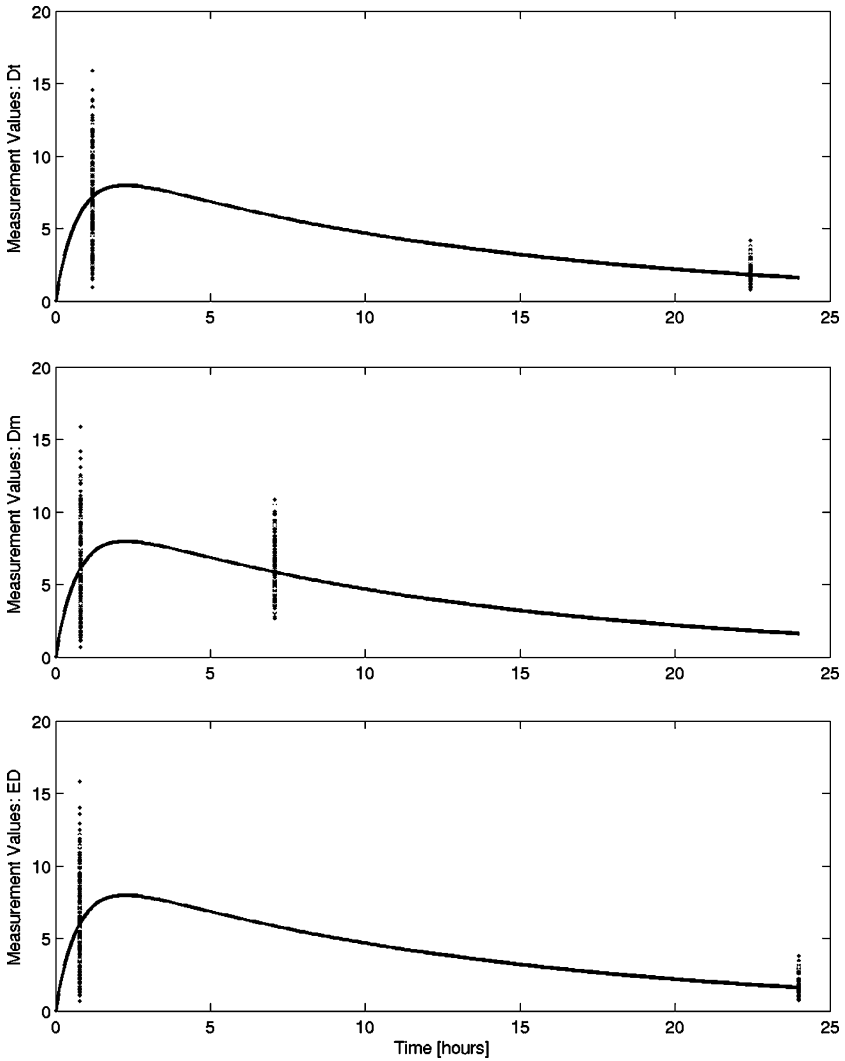

Fig. 6. Model 2 case study designs for (top panel to bottom panel) D-optimal designs (*Dt* and *Dm*) and ED-optimal design (*ED*). "Typical" population model responses (solid line) are plotted against time. Additionally, one of the 500 simulated population data sets used for parameter estimation are shown, to get a sense of the variability in the data (dots).

D-optimal design for this model at the true values of the parameters (*Dt*) yields samples at 5×10^{-6}, 1.187 and 22.42 hr. Figure 6 shows an example simulated set of data superimposed on the "typical" population prediction. The low value of 5×10^{-6} hr here is an artifact of the

numerical method used to solve this problem, as selection of zero for a sampling time yields an unstable objective due to central differencing. The high value selected is close to the maximum allowable time point. These time points represent the optimal sampling times for this experiment, and will allow for the greatest possible accuracy in estimating the parameters. D-optimal design at the misspecified priors (*Dm*) yields sample times of 5×10^{-6}, 0.817 and 7.07 hr.

ED-optimal design for this model, centered at the misspecified values yield sampling times of 5×10^{-6}, 0.769 and 24.0 hr. The intermediate sampling times are smaller (0.769 vs. 1.187) and larger (24.0 vs. 22.42) than the *Dt* design.

Estimates for each of the six parameters (β^{pop} and d) were available from the 500 replicates of each of the three experiments. The *Dt*, *Dm* and *ED* designs generated 500 (100%) experiments, where estimate precisions could be calculated. Parameter estimates, Fig. 7, and %CV's, Fig. 8, are shown as box plots for those experiments where parameter precisions could be calculated. The parameter estimates (Fig. 7) appear to be very similar when comparing the *Dt* and *ED* designs, but markedly different when examining the *Dm* design. The elimination rate constant and clearance estimates (β_2^{pop},β_3^{pop}) for the *Dm* case appear to have a larger spread, as do the estimates for the related BSV parameters. The %CV's (Fig. 8) follow a similar pattern. Precise estimates for all parameters appear more difficult in the Dm case, except for the BSV parameter related to the absorption rate (d_1), where the estimates bias (Fig. 7) is quite high.

DISCUSSION

The failure rate of the *Dm* designs was only slightly higher compared to the *Dt* designs, 3% vs. 1% for Model 1, and 0% vs. 0% for Model 2, respectively. However, the *ED* designs had a failure rate of 1% and 0% for Models 1 and 2, which is a good result. Figures 3 and 7 show larger spreads for the *Dm* design estimates. Figure 3 reveals a potential problem for the *Dm* design in obtaining unbiased estimates for the amplitude terms in Model 1. Figure 7 reveals a potential problem for the *Dm* design in obtaining unbiased estimate for the clearance term in Model 2. The parameter estimate precisions are also strong evidence for establishing that the *ED* designs perform better than the *Dm* designs. For Model 1 (Fig. 4), the precisions are significantly poorer for the *Dm* design than either the *ED* or *Dt* designs. For Model 2 (Fig. 8), same pattern is repeated. The *Dm* designs provide estimates that are more biased, have larger spreads and are less precise than either the *ED* or *Dt* designs. The *ED* designs, by comparison, compare quite well to the *Dt* design, being

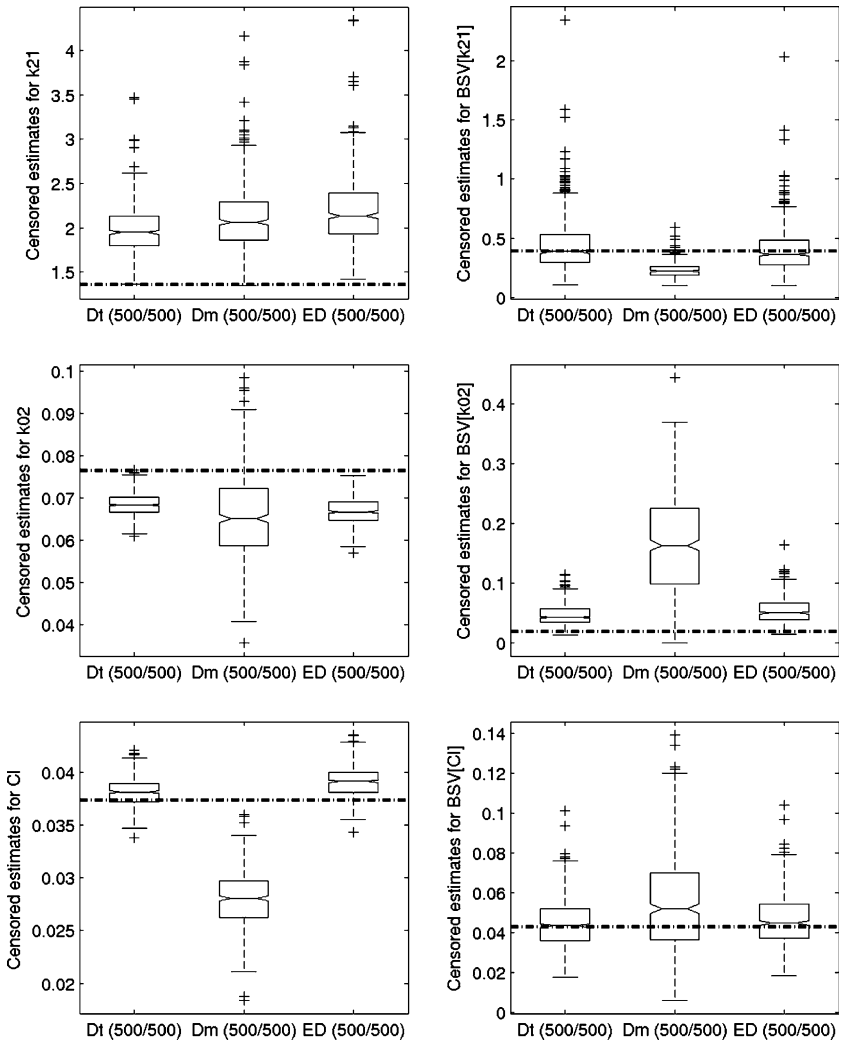

Fig. 7. Parameter estimate results for the subset of the 500 simulated experiments for the Model 2 case study for which estimate precisions were available. Population average parameters are arranged in the left column, between-subject variance in the right. The true parameter values are denoted as a dashed line, and the designs are ordered (left to right) *Dt*, *Dm*, and *ED*, with the number of estimates (out of 500) included in this (abscissa) figure. These box plots show the median of the estimates (notch), interquartile range (box extent), 1.5 times the 25% and 75% quartiles (whiskers), and any outlier estimates beyond 1.5 times the interquartile range (pluses).

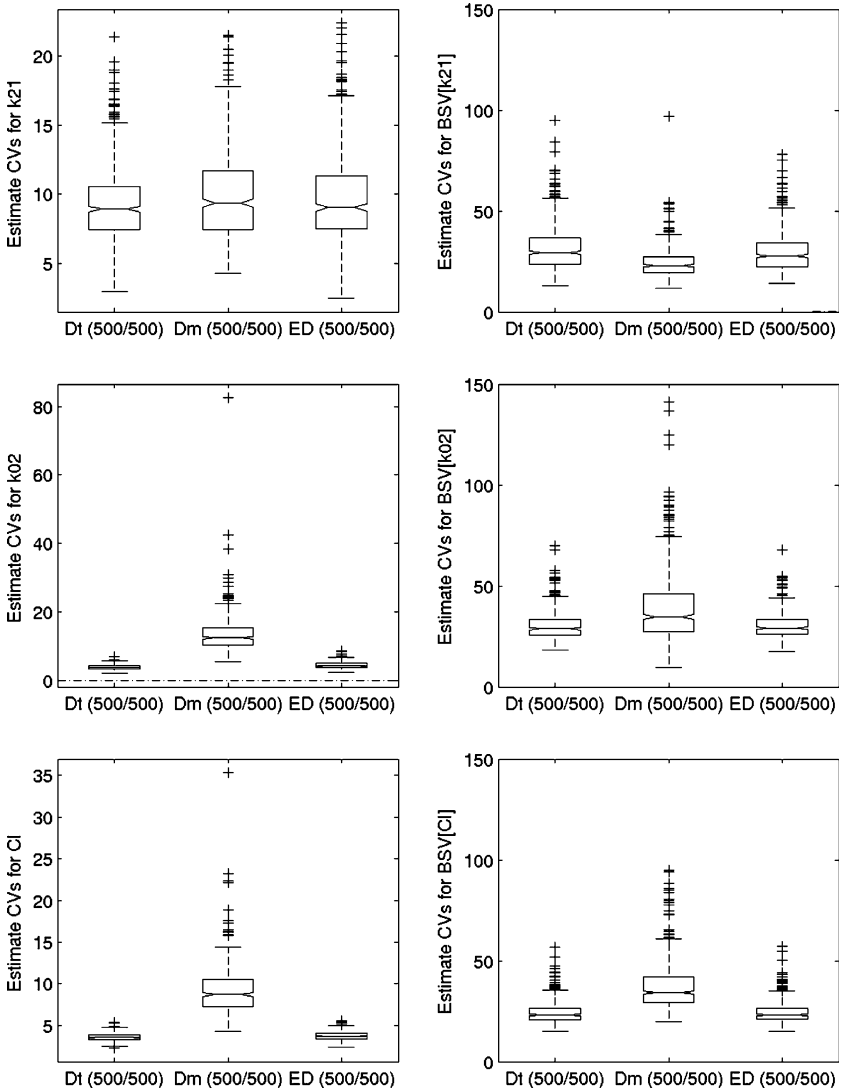

Fig. 8. Parameter estimate coefficients of variation (%CVs) for the subset of the 500 simulated experiments for the Model 2 case study for which estimate precisions were available. Population average parameter %CVs are arranged in the left column, between-subject variance %CVs in the right. The designs are ordered (left to right) *Dt*, *Dm*, and *ED*, with the number of estimates (out of 500) included in this (abscissa) figure. These box plots show the median of the estimates (notch), interquartile range (box extent), 1.5 times the 25% and 75% quartiles (whiskers), and any outlier estimates beyond 1.5 times the interquartile range (pluses).

only slightly more biased, diffuse and imprecise than the *Dt* design. Recall that the *Dm* and *ED* estimation steps were initialized from the misspecified prior values. Therefore, *these metrics describe how the quality of the data collected from these designs is sufficient to shift the estimates from the misspecified values towards the true values of the parameters.*

Comparing the designs graphically (Figs. 2 and 6) provides insight into these behaviors. The *ED* designs are only slightly different than the *Dt*, and so we can expect only slight differences in results derived from these experiments. The *Dm* designs for both models, differ substantially from the *Dt* designs. The *ED* design, whose prior central tendencies are informed by the *Dm* priors, demonstrates that the averaging process does adjust the design values substantially.

Insight Into the ED-optimal Sampling Choices

At our ED-optimal design, the method has selected a set of sampling times (x) that maximize the expectation of the product of the prior probability density function and the Fisher information. Additionally, this result relies upon the selection of a set of parameters (α^m) at the optimal design that maximize the product of the prior probability density function and the Fisher information. Recall that the iterative, nested procedure is:

- Find the design (x^{ED}) that minimizes our approximation of the objective, conditional on the parameters (α^m).
- Find the parameters (α^m) that minimizes the product of the priors and determinant of the FIM, conditional on the current design (x^{ED}).

A radical departure of α^m from the central tendencies of the priors will result in a *ED* design that is significantly different from the *Dm* design. Recall, the *ED* priors have as their central tendencies the *Dm* priors. It may also be telling to examine the α^m's relative to the *Dt* and *Dm* designs, as this may explain why the *ED* designs perform better than the *Dm* designs.

For Model 1, these values were $\alpha^m = [\ 48422,\ 1.147,\ 47819.3,\ 0.078,\ 18.66,\ -2.72,\ 17.16,\ -5.19]$. The true values for the parameters, used in the *Dt* design, were $\alpha^{Dt} = [\ 31000,\ 0.61,\ 14000,\ 0.13,\ 18.28,\ -3.40,\ 16.69,\ -6.49]$. The misspecified values, used in the *Dm* design and as central tendencies for the *ED* design, were $\alpha^{Dm} = [\ 46500,\ 1.22,\ 42000,\ 0.52,\ 18.68,\ -2.70,\ 17.60,\ -5.39]$. The key difference between the *ED* and the *Dm* designs can be explained by examining the *ED* selection of the exponential decay rates, the first being 93% of the *Dm* prior and 171% of the *Dt* prior, and the second being 64% of the *Dm* prior and 95% of the *Dt*

prior. The design is selected, in a sense, for the case where the exponential decay rates are slower than the *Dm* prior would indicate. In fact, the *ED* design considers the possibility that the second decay rate may be even slower than the *Dt* prior. Slower exponential decay indicates later sampling may be required. Hence, we find a sampling time even later than the *Dt* design (7.06 vs. 6.17 days). Here, we can see how the averaging process is making a substantial impact on the design choices, and the choice of the parameter values that maximize the expectation integrand explains this difference.

For Model 2 these values were $\alpha^{m} = [\ 0.676, -1.712, -3.066, 0.254, -3.894, -3.56]$. The true values for the parameters, used in the *Dt* design, were $\alpha^{Dt} = [\ 0.3038, -2.573, -3.289, -0.9352, -3.990, -3.154]$. The misspecified values, used in the *Dm* design and as central tendencies for the *ED* design, were $\alpha^{Dm} = [0.7093, -1.880, -3.066, -0.2421, -3.074, -2.055]$. This model has a proportional error structure, so there is a preference for later sampling times as these data will have less error associated with them. When compared to the D-optimal priors, the first three *ED* parameter choices are 97%, 118% and 100% of the *Dm* priors. One interesting feature of the *ED* parameter choices is that the BSV related to the oral absorption rate is 164% of the *Dm* prior and 328% of the *Dt* prior. With a very large BSV for the oral absorption rate, the absorption rate for a specific individual could fall between 32% and 311% of the typical population value (one standard deviation of the BSV). This may explain why the middle sampling time is sooner than even the *Dm* design would indicate: this amounts to a recommendation to sample earlier to get a precise estimate for those individuals with a very fast absorption. As there is a benefit to sampling late where there is little error, we see a sampling time selected as late as allowed, possibly to estimate the elimination rate in the case where an individual's absorption rate is very slow compared to the overall population.

Summary

The two case-studies examined preliminarily demonstrate the feasibility and utility of ED-optimal design in settings where parameter prior information available for optimal design is misspecified. In both cases the performance, in terms of reduced estimate bias and reduced estimate coefficients of variation, of ED-optimal designs approaches that of D-optimal designs whose priors are exactly the parameters used to simulate the data, despite the fact that the central tendencies of the ED-optimal designs were severely misspecified (50–400%). Certainly, the ED-optimal designs are

superior to the misspecified D-optimal designs in the context of prior misspecification.

Two approaches have previously been suggested to overcome the additional complications of the ED-optimality criterion. Pronzato and Walter initially suggested stochastic approximation techniques, and D'Argenio later suggested an approach relying on adaptive random search suitable for discrete priors (13,35). Both of these methods have been applied in the available packages, *OSP-Fit* and *PopED*. In this work, we present an additional option for approximating the expectation integral with our specific objective, ED-optimality. However, the same methodology could in principle be applied to EID- and API-optimality.

It is our opinion that three major obstacles prevent widespread acceptance of optimal experimental design in the population clinical trial context: the lack of available software tools, the limitation of point-valued priors, and the computational burden of models with sufficient complexity to answer today's problems. ED-optimality, while alleviating the second limitation, presents a difficult numerical problem, which we have made tractable by replacing the expectation integral with a nested optimization problem.

ACKNOWLEDGMENTS

This work was funded by NIH grants RR-12609, EB-001975 "Resource Facility for Population Kinetics" and GM-60021. The authors would like to thank an anonymous reviewer and Dr. Lewis Sheiner for their helpful comments, and Dr. Bradley Bell for useful discussions.

REFERENCES

1. Center for Drug Evaluation and Research (CDER). Guidance for Industry: Population Pharmacokinetics (1999).
2. N. H. Holford, H. C. Kimko, J. P. Monteleone, and C. C. Peck. Simulation of clinical trials. *Annu. Rev. Pharmacol. Toxicol.* **40**:209–234 (2000).
3. D. Z. D'Argenio. Optimal sampling times for pharmacokinetic experiments. *J. Pharmacokinet. Biopharm.* **9**:739–756 (1981).
4. A. C. Hooker, M. Foracchia, M. G. Dodds, and P. Vicini. An evaluation of population D-optimal designs via pharmacokinetic simulations. *Ann. Biomed. Eng.* **31**:98–111 (2003).
5. S. B. Duffull, S. Retout, and F. Mentré. The use of simulated annealing for finding optimal population designs. *Comp. Meth. Prog. Biomed.* **69**:25–35 (2002).
6. S. B. Duffull, F. Mentré, and L. Aaron. Optimal design of a population pharmacokinetic experiment for ivabradine. *Pharm. Res.* **18**(1):83–89 (2001).

7. M. Tod and J.-M. Rocchisani. Implementation of OSPOP, an algorithm for the estimation of optimal sampling times in pharmacokinetics by the ED, EID and API criteria. *Comp. Meth. Prog. Biomed.* **50**:13–22 (1996).
8. S. Retout, S. Duffull, and F. Metre. Development and implementation of the population Fisher information matrix for the evaluation of population pharmacokinetic designs. *Comp. Method. Prog. Biomed.* **65**:141–151 (2001).
9. M. Foracchia, A. Hooker, P. Vicini, and A. Ruggeri. POPED, a software for optimal experiment design in population kinetics. *Comput. Method. Prog. Biomed.* **74**(1):29–46 (2004).
10. R. J. Larsen and M. L. Marx. *An Introduction to Mathematical Statistics and its Application*, 2. Prentice-Hall, Englewood Cliffs, NJ; 1986.
11. R. R. Sitter. Robust designs for binary data. *Biometrics* **48**:1145–1155 (1992).
12. E. M. Landaw. Robust sampling designs for compartmental models under large prior eigenvalue uncertainties. In: *Mathematics and Computers in Biomedical Applications*, (J. Eisendfeld and C. DeLisi Eds.), Elsevier Science, North Holland p. 181–187 1985.
13. L. Pronzato and E. Walter. Robust experiment design via stochastic approximation. *Math. Biosci.* **75**:103–120 (1985).
14. M. Tod and J.-M. Rocchisani. Comparison of ED, EID and API criteria for the robust optimization of sampling times in pharmacokinetics. *J. Pharmacokin. Biopharm.* **25**:515–537 (1997).
15. F. Mentré, C. Dubruc, and J. P. Thenot. Population pharmacokinetic analysis and optimization of the experimental design for mizolastine solution in children. *J. Pharmacokinet. Pharmacodyn.* **28**(3):299–319 (2001).
16. V. V. Fedorov. *Theory of Optimal Experiments*, Academic Press, New York, NY, 1972.
17. N. Laird, N. Lange, and D. Stram. Maximum likelihood computations with repeated measures: application of the EM algorithm. *J. Am. Stat. Assoc.* **82**:97–105 (1987).
18. M. Davidian and D. M. Giltinan. *Nonlinear Models for Repeated Measurement Data*. Chapman and Hall, New York; 1995.
19. Y. Merle and F. Mentré. Bayesian design criteria: Computation, comparison, and application to a pharmacokinetic and a pharmacodynamic model. *J. Pharmacokinet. Biopharm.* **1**:101–125 (1995).
20. E. F. Vonesh and V. M. Chinchilli. *Linear and Nonlinear Models for the Analysis of Repeated Measurements*, Marcel Dekker, New York; 1997.
21. V. V. Fedorov and P. Hackl. *Model Oriented Design of Experiments*, Springer, New York; 1997.
22. S. L. Beal and L. B. Sheiner. NONMEM user's guide. Technical report, University of California, San Francisco (1992).
23. C. C. Peck, S. L. Beal, L. B. Sheiner, and A. I. Nichols. Extended least squares nonlinear regression: A possible solution to the "choice of weights" problem in analysis of individual pharmacokinetic parameters. *J. Pharmacokin. Biopharm.* **5**:545–557 (1984).
24. F. Mentré, A. Mallet, and D. Baccar. Optimal design in random-effects regression models. *Biometrika* **84**:429–442 (1997).
25. B. M. Bell. Approximating the marginal likelihood estimate for models with random parameters. *Appl. Math. Comput.* **119**:57–75 (2001).
26. Y. Merlé and M. Tod. Impact of pharmacokinetic-pharmacodynamic model linerization on the accuracy of population information matrix and optimal designs. *JPKPD* **28**(4):363–387 (2001).
27. G. E. P. Box and H. L. Lucas. Design of experiments in nonlinear situations. *Biometrika* **46**:77–90 (1959).
28. M. Tod, F. Mentré, Y. Merle, and A. Mallet. Robust optimal design for the estimation of hyperparameters in population pharmacokinetics. *J. Pharmacokinet. Biopharm.* **26**(6):689–716 (1998).
29. B. P. Carlin and T. A. Louis. *Bayes and Empirical Bayes Methods for Data Analysis*, Chapman & Hall/CRC, New York; 2000.
30. P. S. Laplace. Memoir on the probability of the causes of events (English translation of the 1774 French original by Stigler SM). *Stat. Sci.* **1**:364–378 (1986).

31. L. Tierney and J. B. Kadane. Accurate approximations for posterior moments and marginal densities. *J. Am. Stat. Assoc.* **81**(393):82–86 (1986).
32. E. F. Vonesh. A note on the use of Laplace's approximation for nonlinear mixed-effects models. *Biometrika* :447–452 (1996).
33. A. Chait, D. Foster, D. G. Miller, and E. L. Bierman. Acceleration of low-density lipoprotein catabolism in man by total parenteral nutrition. *Proc. Soc. Exp. Biol. Med.* **168**(1):97–104 (1981).
34. E. M. Landaw. *Optimal design for individual parameter estimation in pharmacokinetics*, Raven Press, New York; 1985.
35. D. Z. D'Argenio. Incorporating prior parameter uncertainty in the design of sampling schedules for pharmacokinetic parameter estimation experiments. *Math. Biosci.* **99**(1): 105–118 (1990).