J Pharmacokinet Pharmacodyn (2009) 36:125–145
DOI 10.1007/s10928-009-9114-z

Simultaneous optimal experimental design on dose and sample times

Joakim Nyberg · Mats O. Karlsson · Andrew C. Hooker

Received: 25 July 2008 / Accepted: 10 March 2009 / Published online: 25 March 2009

Abstract Optimal experimental design can be used for optimizing new pharmacokinetic (PK)–pharmacodynamic (PD) studies to increase the parameter precision. Several methods for optimizing non-linear mixed effect models has been proposed previously but the impact of optimizing other continuous design parameters, e.g. the dose, has not been investigated to a large extent. Moreover, the optimization method (sequential or simultaneous) for optimizing several continuous design parameters can have an impact on the optimal design. In the sequential approach the time and dose where optimized in sequence and in the simultaneous approach the dose and time points where optimized at the same time. To investigate the sequential approach and the simultaneous approach; three different PK–PD models where considered. In most of the cases the optimization method did change the optimal design and furthermore the precision was improved with the simultaneous approach.

Keywords Optimal design · Simultaneous optimization · Dose optimization · Pharmacometrics

Introduction

Population pharmacokinetic (PK) models are becoming increasingly important in the drug development field. These models can be used to describe experiments already performed, and perhaps more importantly, these models can be used to predict what will happen in new trials. The models can be used to predict concentrations for new individuals, predict outcomes based on a pharmacodynamic (PD) model, etc. Population models can also be used as prior information to optimal experimental

J. Nyberg (✉) · M. O. Karlsson · A. C. Hooker
Division of Pharmacokinetics and Drug Therapy, Department of Pharmaceutical Biosciences, Uppsala University, Box 591, 751 24 Uppsala, Sweden
e-mail: Joakim.nyberg@farmbio.uu.se

designs which can optimize the designs for new studies based on a prior model. These designs optimize the study design parameters to provide as much information as possible on the parameters of the model. The methods for computing optimal experimental designs using non-linear mixed effects models was first elaborated by Mentré et al. [1, 2], and since then the method has mostly been used to optimize the measurement times of an experiment [3] or the number of individuals in a study [4].

However, optimal design theory does not limit the type of design variables we can optimize to improve the information content of our experiment. In theory any design variable could be optimized. Examples of other optimization parameters are dose, infusion duration of a drug, wash-out period length for disease progression studies, start and stop time of studies, titration schemes, etc. In fact, non-time type design variables are often more crucial to select appropriately. Ideally both sample times and other design variables ought to be optimized. Most of the current population optimal design software does not support optimization on design parameters other than time and therefore little research about optimizing on these parameters has been done [5, 6]. In this paper the two design variables we choose to optimize on are dose and sample times. This joint optimization could be applicable very early on in drug development as well as for late phase studies of a drug. Examples are: screening experiments where optimal doses (concentrations) and sampling times for the screening of the drug candidates could be optimized, disease progression studies where a dose and sample time could be optimized to give a certain power. Further, in highly standardized tests like the glucose tolerance tests, optimization of dose and sample times could be applicable and increase the efficiency of the test.

Optimization on design parameters other than sampling times could be of great importance, especially when optimizing on PK–PD models or when the elimination of the PK model is non-linear. In PK–PD or only PD models the dose could be crucial to reach the desired effect. For example, with a typical Emax dose-concentration-effect model a certain dose has to be given to achieve good knowledge about the maximum effect. A less trivial optimization problem is to select a dose that gives good estimates of the concentration at 50% of the maximal effect (EC50) and in this situation optimal design theory could be very helpful.

As mentioned above, the correlation between measurement times and other design parameters (such as dose) could play a crucial role in designing an experiment. Considering the dose-concentration-effect Emax model, the optimal times to measure concentrations and effects so as to gain information about the EC50 parameter will obviously be dependent on the dose given. Similar to the estimation problem with PK/PD models where PK and PD parameters can be estimated sequentially and simultaneously [7], one can envisage a range of simultaneous and sequential approaches to optimize on different design parameters (e.g. sample times and dose).

The aim of this paper is to investigate some of the methods available in optimal design for optimizing over other design variables in addition to measurement times. To do this we investigate various examples where dose and measurement times can be optimized and the optimization of dose is not trivial (mostly non-linear models). We then approach the optimal design calculations with both simultaneous and sequential methods and explore the resulting optimal designs.

Methods

The population model

The *i*th individual responses from a non-linear mixed effects model (e.g. a population PK–PD model) can be described as

$$\vec{y}_i = \mathrm{f}\left(\vec{t}_i, \vec{\beta}_i\right) + \mathrm{h}\left(\vec{t}_i, \vec{\beta}_i, \vec{\varepsilon}_i\right) = \begin{cases} \mathrm{f}_1\left(\vec{t}_{i1}, \vec{\beta}_{i1}\right) + \mathrm{h}_1\left(\vec{t}_{i1}, \vec{\beta}_{i1}, \vec{\varepsilon}_{i1}\right) \\ \vdots \\ \mathrm{f}_j\left(\vec{t}_{ij}, \vec{\beta}_{ij}\right) + \mathrm{h}_j\left(\vec{t}_{ij}, \vec{\beta}_{ij}, \vec{\varepsilon}_{ij}\right) \end{cases} \quad (1)$$

where $\mathrm{f}_j(\,)$ is a function that describes how the response j change with the experimental variables, $\vec{t}_{ij}$ is the *i*th individual vector (arrow indicates vector) of design variables belonging to response j (e.g. PK sampling times). The model parameters for the *i*th individual are functions of the typical values in the model $\vec{\theta}$, the between subject variability (BSV) of the *i*th individual $\vec{\eta}_i$, and the covariates (e.g. dose, weight, etc.) for that individual $\vec{a}_i$, i.e. $\vec{\beta}_i = \mathrm{g}\left(\vec{\theta}, \vec{\eta}_i, \vec{a}_i\right)$. The function $\mathrm{h}_j(\,)$ represents the residual error model for response j, $\vec{\varepsilon}_{ij}$ is the vector of residual error terms for the *i*th individual response j. The residual error is normally distributed with mean zero and a variance of $\mathbf{\Sigma}$, i.e. $\vec{\varepsilon}_{ij} \sim \mathbf{N}(0, \mathbf{\Sigma})$. It is also assumed that the measurement errors are independent (even though that is not necessary as in Gueorguieva et al. [8]) but they are functions of the experimental variables and the model parameters. Hence the variance matrix for the measurements can be defined as

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_n^2 \end{pmatrix} \quad (2)$$

where n is the number of residual error parameters in all of the responses. The BSV is assumed to be normally distributed with a mean zero and a variance of $\mathbf{\Omega}$ ($\eta_i \sim \mathbf{N}(0, \mathbf{\Omega})$). In this work we assume that the $\vec{\eta}_i$ are independent from each other (although this restriction is not required, see [8] for an example) and independent from the residual variability:

$$\mathbf{\Omega} = \begin{pmatrix} \omega_1^2 & & 0 \\ & \ddots & \\ 0 & & \omega_k^2 \end{pmatrix} \quad (3)$$

where k is the number of BSV parameters in all responses. Note also that it is possible within this framework that some responses share the same residual variability (e.g. $\sigma_{pk} = \sigma_{pd}$) and also that the design variables for some responses are coupled in the optimization (e.g. $\vec{t}_{pk,i} = \vec{t}_{pd,i}$), however this is not the case in this paper.

Fisher information matrix

In general, optimal design theory is based on the Cramer-Rao inequality which states that the inverse of the Fisher information matrix is a lower bound for the variance-covariance matrix of the parameters in a model. The population Fisher information matrix (FIM) for population models was first proposed by Mentré et al. [1] and further developed and improved by Retout et al. [2, 9]. In this work we compute the FIM using the approach described by Forracchia et al. [5] with some modifications (described below). The Fisher information matrix is defined as the expectation value of the second derivative of the negative joint log likelihood function with respect to the estimated parameters:

$$\mathbf{FIM} = -E\left[\frac{\partial^2 \mathrm{L}\left(\vec{\Theta}\right)}{\partial\vec{\Theta}\partial\vec{\Theta}^T}\right] \tag{4}$$

where $\vec{\Theta} = \left[\vec{\theta}, \vec{\omega}^2, \vec{\sigma}^2\right]$ is the vector of the parameters to estimate in the population model. To guarantee that the random effects are normally distributed we must linearize both f( ) and h( ) with respect to the random effects (η and ε). In this treatment we linearize about the mean values ($\eta = 0$ and $\varepsilon = 0$) as in the FO method from NONMEM [10, 11]. It has been shown [12] that the prediction from the FO approximation in optimal design calculations correlates well to the standard errors computed by NONMEM [11] using the FOCE method or Monolix [13] using the SAEM algorithm. Linearizing around $\vec{\eta}_i = 0,\ \vec{\varepsilon}_i = 0$ we get:

$$\begin{aligned}\vec{y}_i &= \mathrm{f}\left(\vec{t}_i, \mathrm{g}\left(\vec{\theta}, \vec{\eta}_i, \vec{a}_i\right)\right) + \mathrm{h}\left(\vec{t}_i, \mathrm{g}\left(\vec{\theta}, \vec{\eta}_i, \vec{a}_i\right), \vec{\varepsilon}_i\right) \\ &\approx \mathrm{f}\left(\vec{t}_i, \mathrm{g}\left(\vec{\theta}, \vec{\eta}_i = 0, \vec{a}_i\right)\right) + \mathbf{L}_i\left(\vec{t}_i, \mathrm{g}(.)\right) \cdot \vec{\eta}_i + \mathbf{H}_i\left(\vec{t}_i, \mathrm{g}(.), \vec{\varepsilon}_i\right) \cdot \vec{\varepsilon}_i\end{aligned} \tag{5}$$

where

$$\mathbf{L}_i\left(\vec{t}_i, \mathrm{g}\left(\vec{\theta}, \vec{\eta}_i, \vec{a}_i\right)\right) \equiv \frac{\partial \mathrm{f}}{\partial \vec{\eta}_i}\left(\vec{t}_i, \mathrm{g}\left(\vec{\theta}, \vec{\eta}_i, \vec{a}_i\right)^T\right)\Big|_{\vec{\eta}_i = 0} \tag{6}$$

and

$$\mathbf{H}_i\left(\vec{t}_i, \mathrm{g}\left(\vec{\theta}, \vec{\eta}_i, \vec{a}_i\right), \vec{\varepsilon}_i\right) \equiv \frac{\partial h}{\partial \vec{\varepsilon}_i}\left(\vec{t}_i, \mathrm{g}\left(\vec{\theta}, \vec{\eta}_i, \vec{a}_i\right)^T, \vec{\varepsilon}_i\right)\Big|_{\vec{\eta}_i = 0, \vec{\varepsilon}_i = 0} \tag{7}$$

In this linearization we allow for any residual variability error model and follow the linearization performed by Beal [11]. This is an extension of the previous FIM calculations [3, 5, 8, 9] which only allow for additive and proportional error structures. Now the mean and the variance of the linearized model can be defined

$$E_{\vec{y}_i}[\vec{y}_i] \approx \mathrm{f}\left(\vec{t}_i, \mathrm{g}\left(\vec{\theta}, \vec{\eta}_i = 0, \vec{a}_i\right)\right) \tag{8}$$

$$\mathbf{Var}(\vec{y}_i) \approx \mathbf{L} \cdot \mathbf{\Omega} \cdot \mathbf{L}^T + diag\left(\mathbf{H} \cdot \mathbf{\Sigma} \cdot \mathbf{H}^T\right) \tag{9}$$

The individual Fisher information matrix depending on the parameters $\vec{\Theta}$ and the design variables $\vec{x}_i = \left[\vec{t}_i, \vec{a}_i\right]$ may now be written as [5]

$$\mathbf{FIM}_i\left(x_{i,}\vec{\Theta}\right) = \begin{pmatrix} \mathbf{M}_{1i} & \mathbf{0} \\ \mathbf{M}_{2i} & \mathbf{M}_{3i} \end{pmatrix}^T \begin{pmatrix} \mathbf{Var}(\vec{y}_i)^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_{4i}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{M}_{1i} & \mathbf{0} \\ \mathbf{M}_{2i} & \mathbf{M}_{3i} \end{pmatrix} \tag{10}$$

where

$$\mathbf{M}_{1i} = \frac{\partial \mathrm{f}\left(\vec{\mathrm{t}}_{\mathrm{i}}, \mathrm{g}\left(\vec{\theta}, 0, \vec{\mathrm{a}}_{\mathrm{i}}\right)\right)}{\partial \vec{\theta}} \tag{11}$$

$$\mathbf{M}_{2i} = \frac{\partial vec(\mathbf{Var}(\vec{y}_i))}{\partial \vec{\theta}} \tag{12}$$

$$\mathbf{M}_{3i} = \left[\frac{\partial vec(\mathbf{Var}(\vec{y}_i))}{\partial \vec{\omega}^2}, \frac{\partial vec(\mathbf{Var}(\vec{y}_i))}{\partial \vec{\sigma}^2}\right] \tag{13}$$

$$\mathbf{M}_{4i} = 2 \cdot \mathbf{Var}(\vec{y}_i) \otimes \mathbf{Var}(\vec{y}_i) \tag{14}$$

The population FIM is the sum of the individual matrices in an experiment. If we assume that certain groups of individuals in a study will all have the same design (and the same covariates) then all these individuals will have the same FIM and hence the FIM for a number of groups may be defined as:

$$\mathbf{FIM}\left(\vec{x}, \vec{\Theta}\right) = \sum_{i=1}^{m} \mathbf{FIM}_i\left(\vec{x}_i, \vec{\Theta}\right) = \sum_{i=1}^{N_g} g_i \cdot \mathbf{FIM}_i\left(\vec{x}_i, \vec{\Theta}\right) \tag{15}$$

where N_g is the number of groups, g_i is the number of individuals in group i and $m = \sum_{i=1}^{N_g} g_i$.

D-optimal design

In d-optimal design the determinant of the Fisher information matrix, $\left|\mathbf{FIM}\left(\vec{x}, \vec{\Theta}\right)\right|$, is maximized, and hence, by the Cramer-Rao inequality the expected variance-covariance matrix will have the lowest possible asymptotic lower bound when estimating parameters. By optimizing with the determinant all possible perturbations of the elements in a matrix, i.e. the off-diagonals in the FIM, are considered.

Comparing designs

When comparing different designs the efficiency is the most common metric. The efficiency for d-optimal design is defined as in Atkinson–Donovev [14],

$$D_{Eff} = \left(\frac{\left|\mathbf{FIM}\left(\vec{x}_1, \vec{\Theta}\right)\right|}{\left|\mathbf{FIM}\left(\vec{x}_2, \vec{\Theta}\right)\right|}\right)^{1/p} \tag{16}$$

where $\vec{x}_1$ and $\vec{x}_2$ are two different designs, $\vec{\Theta}$ are the model parameter values and p are the number of parameters in the designs. This is a good measurement, however it does not reflect the number of parameters in the model except as an average. For

example an efficiency of 95% when having a few parameters in the model might be quite similar to the optimal design but the same efficiency could reflect a larger imprecision of the model in general if the model has a large number of parameters. In this work only designs from the same model with the same parameter values $\vec{\Theta}$ are compared, therefore the ratio of the $|FIM|$ for different designs will mainly be used as a measurement.

$$FIM_{ratio} = D_{Eff}^{p} = \frac{\left|\mathbf{FIM}\left(\vec{x}_1, \vec{\Theta}\right)\right|}{\left|\mathbf{FIM}\left(\vec{x}_2, \vec{\Theta}\right)\right|} \tag{17}$$

Simultaneous and sequential optimization

Optimization of design variables can be done one at a time, i.e. sequentially, where for example optimization on dose and sample times could be performed by optimizing on dose first, then sample times or by optimizing on sample times first and then doses. By optimizing sequentially we reduce the solution space, but also reduce the chance of finding the true optimum.

In the simultaneous approach, similar to the methods presented by Hooker et al. and Zhang et al. [7, 15], all design parameters are optimized at the same time, i.e. within the search algorithms. This, in some circumstances discussed later, allows for a more global search. For example the gradient of the FIM used in the optimization search will be pointing in both the dose and the sampling times steepest descent at the same time.

Three different optimization approaches were considered in this work for the optimization of doses and sample times:

1. T|D: Fix the doses, optimize on sample times, then fix the optimal times and optimize on doses.
2. D|T: Fix the sample times, optimize on doses, then fix the optimal doses and optimize on sample times.
3. T,D: Optimize on doses and sample times at the same time.

The difference between the approaches can be seen in Fig. 1 where the peak corresponding to the true optimum will only be guaranteed to be chosen in the simultaneous approach, if a global optimization is assured, but not in the sequential approaches.

The optimization algorithms used were first a random search with adaptive narrowing (RS) with 300 iterations (narrowing after 50 iterations) followed by a Steepest Descent (SD) optimization with 150 iterations and finally a line search (LS) method with 50 iterations for each design variable to be optimized. The LS algorithm is performed by dividing the design space in each dimension in 50 equal sizes and the algorithm finds the best (e.g. highest determinant) criterion out of those 50 values. This is done sequentially for each design variable. The optimal design was said to be found if the LS did not change the optimal design given by the SD, otherwise the method re-started with the RS, SD and LS until convergence. These methods are presented in more details in Forrachia et al. [5] and are all standard settings in PopED [5].

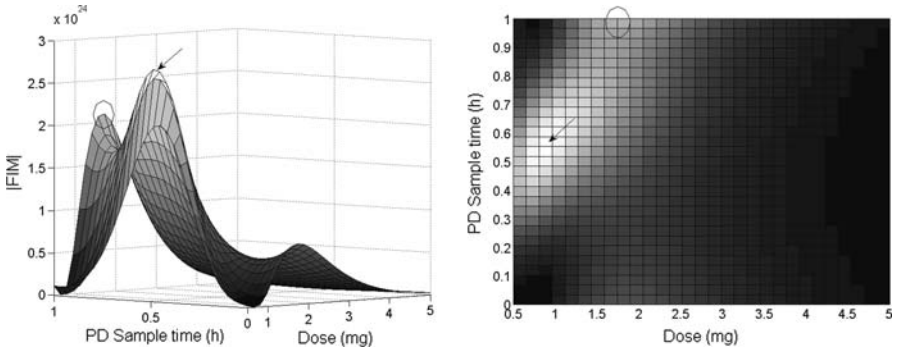

Fig. 1 The surface of the determinant of the FIM for an Emax dose-concentration-effect model. The left panel shows the determinant of FIM as a surface with the corresponding view from above in the right panel. It can be seen that a T|D approach with an initial dose of 2.5 mg will get 1 h as the optimal time and hence get the peak marked with a ring as the optimal determinant. The T,D approach will instead get the true optimum at time 0.6 (h) with a dose of 1 mg marked with an arrow

For the SD method the simultaneous approach (T,D) will point in both the dose and time derivative direction but the sequential approaches will point in one of these directions at each iteration, depending on time-first or dose-first. Similarly; for the RS and LS methods T,D will search in both the time and the dose dimension while the sequential methods will only search in either the time dimension or the dose dimension within an iteration.

Examples

Three examples were evaluated to cover different models and different approaches for the optimization. In all of the examples the residual variability terms were assumed fixed and known (as previously assumed by Forrachia et al. [5]). In general; it is quite easy to have precise parameter estimates of residual variability parameters compared to e.g. between subject variability parameters and are therefore excluded as parameters in the Fisher information matrix. All the optimization was performed in the Matlab version of PopED [5], available free of charge at http://poped.sourceforge.net. The different design setups in the examples

Table 1 Design setups

Design setup	Model 1	Model 2	Model 3
(a) Sample times different between groups, no placebo dose	×	×	×
(b) Sample times the same between groups, no placebo dose	×	×	–
(c) Sample times different between groups, placebo doses for PD measurements	×	×	–
(d) Sampling times same between groups, placebo doses for PD measurements	×	–	–

All design setups (a, b, c and d) had the initial sampling times spread evenly within a group and the initial doses spread evenly between different groups. For each setup, designs were calculated for 1–5 groups

are presented in Table 1. The number of groups in the optimization differed between one group and five groups. All design setups had the initial sampling times spread evenly within a group and the initial doses spread evenly between different groups. To minimize the risk that a local minimum was found the setup was optimized five times for each of the sequential and simultaneous approaches.

Three specific (widely used) models were used and they were chosen to have a non-obvious dose (non-linear in the dose–response). As described above numerous numbers of groups, doses, sample times and grouping of samples and doses was investigated to cover as large design range as possible.

Model 1: intravenous PK, E-max PD model

In this example a one-compartment PK model with bolus input followed by a direct Emax PD-model is used. This model was also investigated by Hashimoto and Sheiner [16]. The PK model in mg/l for the *i*th subject is a one compartment intravenous bolus dose model with a proportional residual variability according to

$$\begin{aligned} \mathrm{f}_{pk}\left(\vec{t}_{pk,i},\vec{\beta}_i\right) &= \frac{Dose_i}{V_i}\cdot e^{-\frac{CL_i}{V_i}\cdot\vec{t}_{pk,i}} \\ \mathrm{h}_{pk}\left(\vec{t}_{pk,i},\vec{\beta}_i,\vec{\varepsilon}_{pk,i}\right) &= \mathrm{f}_{pk}\left(\vec{t}_{pk,i},\vec{\beta}_i\right)\cdot\vec{\varepsilon}_{pk,i}\,, \\ \vec{\varepsilon}_{pk,i} &\sim \mathbf{N}\left(0,\sigma^2_{pk}\right) \end{aligned} \tag{18}$$

where $\vec{t}_{pk,i}$ is a vector with the PK sample times for subject i and $\vec{\beta}_i$ is a vector with the ith subjects model parameters and covariates.

The ith subjects parameters for clearance (CL_i) and volume of distribution (V_i) has an additive normally distributed BSV term.

The pharmacodynamic (PD) model was characterized by an Emax model with an additive residual variability and its connection to the PK model for the ith subject is

$$\begin{aligned} y_i &= \begin{cases} \mathrm{f}_{pk}\left(\vec{t}_{pk,i},\vec{\beta}_i\right)+\mathrm{h}_{pk}\left(\vec{t}_{pk,i},\vec{\beta}_i,\vec{\varepsilon}_{pk,i}\right) \\ \mathrm{f}_{pd}\left(\vec{t}_{pd,i},\vec{\beta}_i\right)+\mathrm{h}_{pd}\left(\vec{t}_{pk,i},\vec{\beta}_i,\vec{\varepsilon}_{pd,i}\right) \end{cases} \\ \mathrm{f}_{pd}\left(\vec{t}_{pd,i},\vec{\beta}_i\right) &= E_{0,i}+\frac{Emax_i\cdot\mathrm{f}_{pk}\left(\vec{t}_{pd,i},\vec{\beta}_i\right)}{EC_{50,i}+\mathrm{f}_{pk}\left(\vec{t}_{pd,i},\vec{\beta}_i\right)}, \\ \mathrm{h}_{pd}\left(\vec{t}_{pd,i},\vec{\beta}_i,\vec{\varepsilon}_{pd,i}\right) &= \vec{\varepsilon}_{pd,i} \\ \vec{\varepsilon}_{pd,i} &\sim \mathbf{N}\left(0,\sigma^2_{pd}\right) \end{aligned} \tag{19}$$

The typical values for the PK model drove the concentration input to the PD model (i.e. in a first order (FO) way). Similar to the PK part of the model the parameters for the PD model have an additive BSV. $Emax_i$ represents the maximal effect for subject i, $EC_{50,i}$ is the concentration at 50% of the effect and $E_{0,i}$ is the baseline of the effect. The parameter values for the model can be seen in Table 2.

Table 2 The model parameter values for Model 1

Variable	Value
θ_{CL}	0.50 (l/h)
θ_{V}	0.20 (l)
θ_{Emax}	1.00
$\theta_{EC_{50}}$	1.00 (mg/l)
θ_{E_0}	1.00
ω^2_{CL}	0.01 $(l/h)^2$
ω^2_{V}	0.0016 $(l)^2$
ω^2_{Emax}	0.01
$\omega^2_{EC_{50}}$	0.09 $(mg/l)^2$
$\omega^2_{E_0}$	0.09
σ^2_{pk}	0.15
σ^2_{pd}	0.15

The random parameters, the between subject variability ω^2 and the residual variability σ^2, are entered in variance units

The doses limits in the optimization were 0.5–5 mg for setup a and b, and 0–5 mg for setup c and d. The PK and PD sample times were limited to be taken within 0–1 h.

Four different design setups were evaluated for this model. For each of the four setups the optimal design was calculated for 1–5 groups of 20 individuals. The number of samples per group was always five, where two of the samples belonged only to the PK model and three were dedicated to the PD model. To ensure that numerical approximations did not affect the result each setup was calculated with analytic derivatives instead of numerical differences.

Model 2: oral PK, E-max PD model

A one-compartment model for oral administration of theophylline [17] is used as the 2nd example. The PK model is followed by a direct linked Emax PD model as in the previous example (Eq. 19).

The PK model has an additive and a proportional residual variability term. The *i*th subject has a PK-response according to

$$\begin{aligned}
&\mathrm{f}_{pk}\left(\vec{t}_{pk,i},\vec{\beta}_i\right) = \frac{Dose_i \cdot k_{a,i} \cdot k_{e,i}}{CL_i \cdot \left(k_{a,i}-k_{e,i}\right)} \cdot \left(e^{-k_{e,i}\cdot\vec{t}_{pk,i}} - e^{-k_{a,i}\cdot\vec{t}_{pk,i}}\right)\\
&\mathrm{h}_{pk}\left(\vec{t}_{pk,i},\vec{\beta}_i,\vec{\varepsilon}_{pk,i}\right) = \mathrm{f}_{pk}\left(\vec{t}_{pk,i},\vec{\beta}_i\right)\cdot\vec{\varepsilon}_{pk_p,i} + \vec{\varepsilon}_{pk_a,i}\\
&\vec{\varepsilon}_{pk_a,i} \sim \mathbf{N}\left(0,\sigma^2_{pk_a}\right)\\
&\vec{\varepsilon}_{pk_p,i} \sim \mathbf{N}\left(0,\sigma^2_{pk_p}\right)
\end{aligned} \tag{20}$$

where $\vec{t}_{pk,i}$ is a vector with the PK sample times for subject i and $\vec{\beta}_i$ is a vector with the model parameters and the covariates for the ith subject. As in Model 1 the link between the PK model and the PD model was a first order link. CL_i is the clearance for subject i, $k_{a,i}$ and $k_{e,i}$ are the rate constants from the absorption compartment into

the central compartment and the elimination rate constant from central compartment. All the parameters in the PK model have exponential BSV.

The pharmacodynamic (PD) model was characterized by an Emax model with an additive residual variability and its connection to the PK model for the *i*th subject is

$$\mathrm{f}_{pd}\left(\vec{t}_{pd,i},\vec{\beta}_i\right) = E_{0,i} + \frac{Emax_i \cdot \mathrm{f}_{pk}\left(\vec{t}_{pd,i},\vec{\beta}_i\right)}{EC_{50,i} + \mathrm{f}_{pk}\left(\vec{t}_{pd,i},\vec{\beta}_i\right)}, \qquad (21)$$

$$\mathrm{h}_{pd}\left(\vec{t}_{pd,i},\vec{\beta}_i,\vec{\varepsilon}_{pd,i}\right) = \vec{\varepsilon}_{pd,i} \sim \mathbf{N}\left(0,\sigma^2_{pd}\right)$$

The PD model has an additive BSV on all parameters. The parameter values are shown in Table 3.

Three different setups were evaluated for this model (a, b and c, see Table 1). The doses were the same for the different setups, i.e. a single dose of 1–10 mg for all groups. The number of individuals in a group was fixed to 4. The time points (three PK samples and three PD samples) were restricted to be in the range of 0–25 h for both the PK and the PD model. In setup (c) only two PK and two PD samples were allowed, the samples times for the PD measurements in this setup were $\vec{t}_{pd} = [0, 120]$.

Model 3: intravenous PK with Michaelis Menten elimination

The one-compartment PK model with Michaelis Menten elimination and a combined additive and proportional residual variability term is used in this example. The PK model is defined by the equation

Table 3 The model parameter values for Model 2

Variable	Value
θ_{CL}	0.0373 (l/h)
θ_{k_a}	2.71 (1/h)
θ_{k_e}	0.0763 (1/h)
θ_{Emax}	1.00
$\theta_{EC_{50}}$	1.00 (mg/l)
θ_{E_0}	1.00
ω^2_{CL}	0.0238
$\omega^2_{k_a}$	0.7840
$\omega^2_{k_e}$	0.0185
ω^2_{Emax}	0.09
$\omega^2_{EC_{50}}$	0.09 (mg/l)^2
$\omega^2_{E_0}$	0.01
$\sigma^2_{pk_p}$	0.15
$\sigma^2_{pk_a}$	0.419 (mg/l)^2
σ^2_{pd}	0.15

The random parameters, the between subject variability ω^2 and the residual variability σ^2 are entered in variance units

$$
\begin{aligned}
\mathrm{y}_i\left(\vec{t}_i,\vec{\beta}_i,\vec{\varepsilon}_i\right) &= \mathrm{f}_{pk}\left(\vec{t}_i,\vec{\beta}_i\right) + \mathrm{h}_{pk}\left(\vec{t}_i,\vec{\beta}_i,\vec{\varepsilon}_i\right)\\
\mathrm{f}_{pk}\left(\vec{t}_i,\vec{\beta}_i\right) &= \frac{\mathrm{A}_{1,i}\left(\vec{t}_i,\vec{\beta}_i\right)}{V_i}\\
\mathrm{h}_{pk}\left(\vec{t}_i,\vec{\beta}_i,\vec{\varepsilon}_i\right) &= \mathrm{f}_{pk}\left(\vec{t}_i,\vec{\beta}_i\right)\cdot\vec{\varepsilon}_{pk_p,i} + \vec{\varepsilon}_{pk_a,i}\\
\frac{\mathrm{dA}_{1,i}\left(\vec{t}_i,\vec{\beta}_i\right)}{\mathrm{d}t} &= -\frac{Vmax_i\cdot\mathrm{A}_{1,i}}{Km_i+\mathrm{A}_{1,i}}\\
\vec{\varepsilon}_{pk_a,i} &\sim \mathbf{N}\left(0,\sigma^2_{pk_a}\right)\\
\vec{\varepsilon}_{pk_p,i} &\sim \mathbf{N}\left(0,\sigma^2_{pk_p}\right)
\end{aligned}
\tag{22}
$$

with the closed form solution found by Beal [17]

$$
\mathrm{A}_{1,i}\left(\vec{t}_i,\vec{\beta}_i\right) = Km_i\cdot\mathrm{F}\left(\mathrm{X}\left(\vec{t}_i,\vec{\beta}_i\right)\right) \tag{23}
$$

where the function F satisfies the equation $y\cdot e^y = e^x$ and

$$
\mathrm{X}\left(\vec{t}_i,\vec{\beta}_i\right) = \ln\frac{\mathrm{A}_{1,i}\left(0,\vec{\beta}_i\right)}{Km_i} + \frac{\mathrm{A}_{1,\mathrm{i}}\left(0,\vec{\beta}_\mathrm{i}\right)}{Km_i} - \frac{Vmax_i}{Km_i}\cdot\vec{t}_i \tag{24}
$$

where the initial amount in compartment one at time zero is $\mathrm{A}_{1,i}\left(0,\vec{\beta}_i\right) = Dose_i$ (units). The concentration of subject i is $\mathrm{f}_{pk}\left(\vec{t}_i,\vec{\beta}_i\right)$ (units/l). The maximal rate constant is given by $Vmax_i$ (units/h), the Michaelis Menten constant is represented by Km_i (units) and the volume of distribution is denoted V_i (l). All parameters in the model have an exponential BSV term. The function F was solved in an iterative way using Newton's method with a precision of $\varepsilon \sim 10^{-9}$. The parameter values used for this model are presented in Table 4.

Only one setup was investigated with this model (a, see Table 1). The dose in this setup was $Dose_i \in [10, 500]$ (units) with one dose per group. Three time points for each group were considered and were constrained $\vec{t}_i \in [0, 2]$ (h). There were 20 individuals in each group.

Table 4 The model parameter values for Model 3

Variable	Value
θ_{Vmax}	182 (units/h)
θ_{Km}	73 (units)
θ_V	6 (l)
ω^2_{Vmax}	0.01
ω^2_{Km}	0.01
ω^2_V	0.01
$\sigma^2_{pk_p}$	0.01
$\sigma^2_{pk_a}$	1 (units/l)^2

The random parameters are shown in variance units

Results

The five optimizations for each approach and group certified that an optimal design was found since it did not change the determinant of the FIM more than $\sim 1.5\%$ for Models 1 and 3 on average. For Model 2 the average change in the optimization was 6%. In all cases the largest determinant of the Fisher information matrix was chosen as the optimal design.

Model 1

The optimal dose and sampling schedule for the 5 group designs in setup 1a are shown in Fig. 2. In all cases the PK sampling times were the same but the optimal PD sampling schedule and the optimal doses to give were different.

The ratio $Deff^p$ between the sequential and simultaneous $|FIM|$ values for the three different optimization approaches is shown in Fig. 3. It can be seen that it is always at least as informative and often more informative to optimize with the simultaneous approach. The time first approach (T|D) seems, in general, to be more similar to the simultaneous approach (T,D) than the dose first (D|T). The results from setup 1c and 1d is not presented in the figure to make the figure easier to overview but these setups showed similar trends as the setup 1a and 1b (see Appendix 3).

Model 2

As seen in Fig. 4 the time first-approach (T|D) does not seem to suffer anything in efficiency compared to the simultaneous approach (T,D) for the setup 2b, while the setup 2a, time-first (T|D), with only two groups, is much more inefficient than the simultaneous approach. The dose-first (D|T) approach is always less efficient than the simultaneous approach (T,D). The result from setup 2c is excluded in the figure

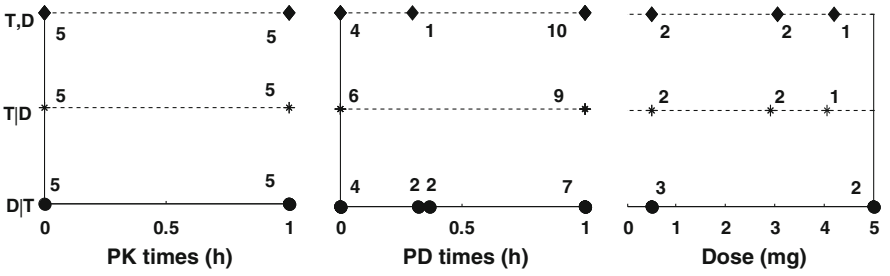

Fig. 2 The optimal sampling schedule and dose for design setup 1a with five groups. The PK optimal sampling times are seen to the left. The PD sampling times are seen in the middle and the optimal doses are to the right. The optimal sampling times and the doses are clustered together for all the groups. That is, a 6 means that six samples are taken at the same or very close time point. Two PK samples were available per group which gives a total of ten PK samples per approach. Similarly; three PD samples were taken per group which gives a total of 15 PD samples per approach. Finally; one dose was given per group which gives a total of 5 doses per approach

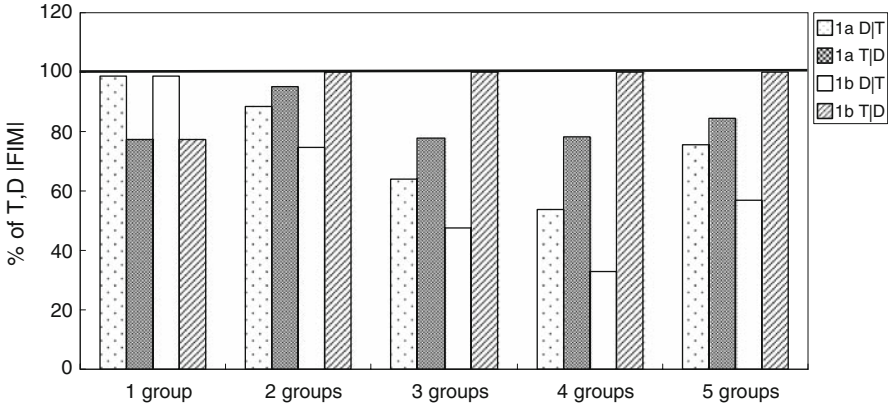

Fig. 3 Fraction of the simultaneous (T,D) D-optimal FIM for Model 1. The T|D approach for setup 1b is similar to the T,D approach, i.e. ~100%. The black horizontal line represents the T,D approach

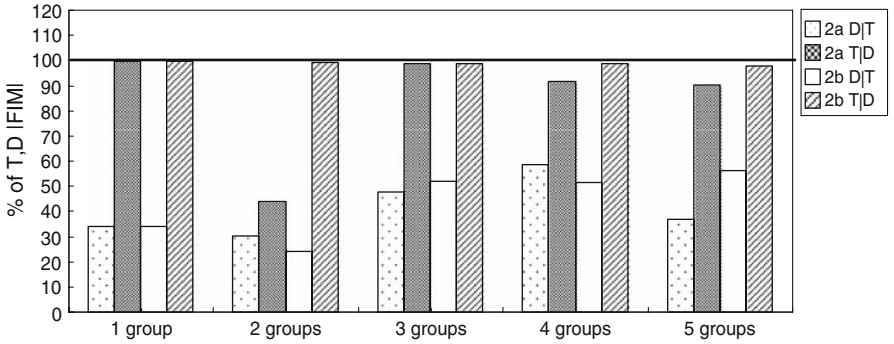

Fig. 4 Fraction of the simultaneous (T,D) D-optimal FIM for Model 2. The black horizontal line represents the T,D approach

to get an easy overview of the findings. However this setup showed similar trends as setup 2a and 2b (see Appendix 3).

Model 3

Figure 5 shows the ratio $Deff^p$ between the sequential and simultaneous $|FIM|$ values for the different optimization approaches. In all cases the sequential approaches are worse than the simultaneous approach. The worst approach seen in Fig. 5 is the sequential approach with dose first (D|T), where with one group the ratio $Deff^p$ is 0.5 indicating that there is 50% less information in this approach compared to the simultaneous approach (T,D).

In this example we also examined the expected coefficients of variation (CV) of the parameters in the different approaches. This can be seen in Fig. 6. Again, the worst scenario compared to the simultaneous approach (T,D) is clearly to optimize

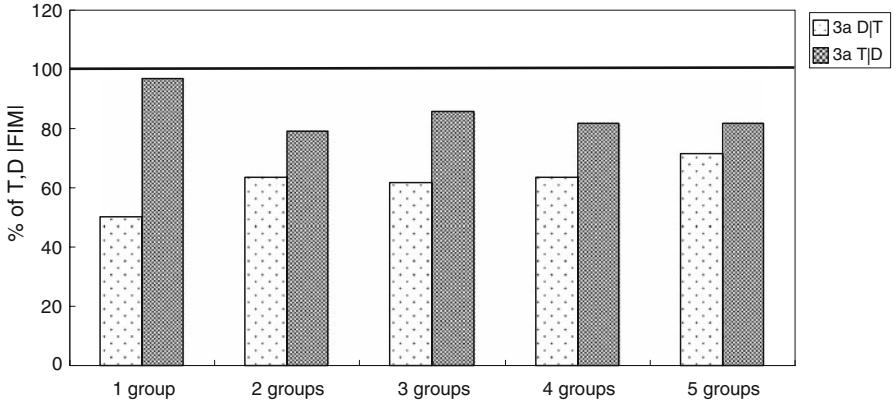

Fig. 5 The |FIM| ratio between the D|T, T|D and the T,D approach for setup 3a. The black horizontal line represents the T,D approach

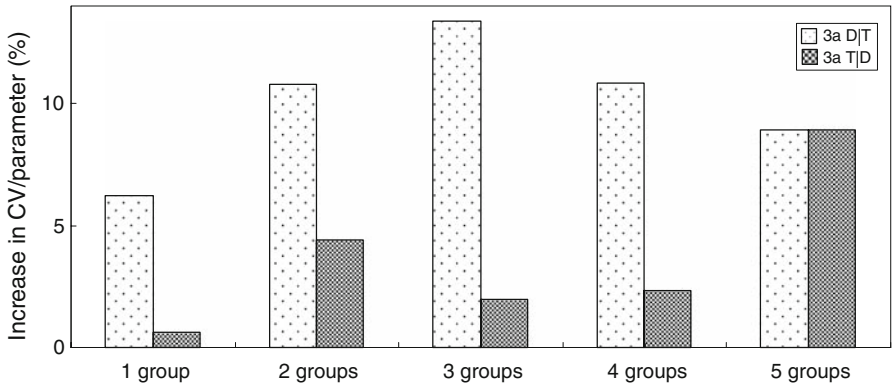

Fig. 6 The increase of the coefficient of variation (CV) in percent per parameter of the sequential D|T, T|D approaches compared to the simultaneous (T,D) approach for setup 3a

dose first (D|T). However with Fig. 6 it is clear that the worst scenario occurs with 3 groups (instead of 1 group as seen in Fig. 5). One possible explanation for this divergence is that the off-diagonal elements in the FIM (i.e. the correlations between the expected variances of the parameter estimates) are considered in the ratio shown in Fig. 5 but not in the CVs in Fig. 6.

Discussion

We have demonstrated that the optimization of design variables other than time is not only possible (as shown previously by Forrachia et al. [5]) but that the method used for the optimization is important. Our results clearly show that the approach

used when optimizing affects the resulting optimal design. For example, in Fig. 2 it can be seen that the PD-sampling times and doses differ between the different approaches. These differences in design also result in significant differences in the information content of the experiments. In some examples the ratio between the determinant of the Fisher information matrix of the sequential and simultaneous approaches could be as small as $\sim 25\%$ (see Fig. 3), indicating that the T,D approach is about four times as informative as the sequential approach.

In our results, the best sequential approach is clearly the T|D approach. In some cases no difference was found between the T|D approach and the T,D approach to optimization. One possible reason for this result is the fact that the sequential approaches are more dependent on the initial values of the design parameters. That is, when optimizing dose first and then sample times the initial placement of sample times (at the beginning of the optimization) will affect the chosen optimal doses. In the examples considered in this work the initial doses were closer to the (simultaneous) optimal doses than the initial time points were. Thus the optimization of time-points first resulted in better optimal designs in these examples, but no generalization as to which sequential strategy is better can be provided.

In this work we have focused on optimizing two design variables (dose and sample times). However, the results can be seen in a more general light—optimization on all design variables to be determined could be done, if possible, simultaneously. Examples of other design variables that could be optimized are: times for changing treatment in titration or disease progression studies, dose schedule in oncology trials, run-in and wash out period lengths for drug–drug interaction studies, provocation size or duration in for example glucose tolerance studies, doses in combination studies for determination of synergy/antagonism, etc.

In this study a continuous dose has been used. However, often doses are restricted to, for example, different tablet sizes. Discrete doses are a special case of the continuous dose optimization. To handle this special case, an optimization on time could be done on all of the separate discrete doses (combination of doses) and after optimization the best dose-time combination would be chosen. This approach is not sequential in the sense used previously in this work and will not suffer from the problems presented here, but it could be quite costly in computer run-time if the number of discrete doses is large. Further; if the number of continuous design variables to optimize over is greater than one (for example, sample times and the stop time of an experiment) the simultaneous approach should be used regardless of if the doses are discrete or not.

The models used in the examples have had quite smooth surfaces of the determinant of the FIM (see Fig. 1). If more than one major peak is present in the search space then optimizing sequentially may greatly reduce your chance of finding the global optimum. Thus, it is realistic to believe that with a spikier surface of the determinant of the Fisher information matrix even more discrepancy will be found between the simultaneous and sequential approaches.

In this exercise no parameter uncertainty was introduced, i.e. ED-optimal design. ED-optimal design has been elaborated and discussed in [18–21]. We believe that parameter uncertainty in general is both useful and realistic because the fact that the parameters are not perfectly known in general. However a sequential ED-optimal

design approach will still suffer from the disadvantages shown from sequential D-optimal design; in some cases even more due to a spikier Fisher Information Surface (see Appendix 2). It is again difficult to provide some rule on when the sequential ED-optimal design approach could be expected to perform worse than the simultaneous ED-optimal but a general rule could be that, similar to the D-optimal approach, one should always use a simultaneous approach.

Furthermore, in this exercise, no model misspecification was accounted for except for the residual error. Our belief is that the same type of differences, seen in the different approaches presented here, will be seen when using other types of criterions (different from D-optimal design) such as optimizing over competing models. These types of criterions have been used for optimizing population models [3], however we believe that more investigation is needed to see the impact of these types of criterions on sequential and simultaneous optimization of different design variables.

To our knowledge there exists four ways to perform an optimal design search of two or more continuous design variables:

1) The simultaneous approach (T,D) that, given a good optimizer, will (in general) be the best method of finding a global optimum. This method will also in general need the least amount of computational power.
2) The sequential approaches (T|D), (D|T), optimizing each continuous variable once. T|D and D|T are compared in this paper to T,D. All of these approaches have similar run times (in general). However, as we can see from these results, T|D and D|T will often perform worse than T,D and will never perform better. T|D and D|T will also perform worse than the iterative sequential approach described below.
3) The iterative sequential approach, optimizing each continuous variable iteratively until an iteration does not change the "optimal value". This approach is in general more time consuming than the T,D and will not always find the optimal design even if a global optimizer is present. This is true simply because of the restriction in the search space. Even though we believe that this approach will produce results more close to the simultaneous approach than the sequential methods described above, this method will come with a significantly higher computational cost.

   An empirical example was investigated to see how this approach differs from the T,D approach in a simple example. The optimization was performed in only two dimensions of the surface presented in Fig. 1, i.e. optimization over a PD sample and a dose. The T,D approach was repeated five times and with different (randomly chosen) initial values and the true optimum was found every time. For the iterative T|D approach an initial dose of >1.625 mg (regardless of the initial sample time) always converged to the local optima. Similarly; for the D|T approach, an initial sample time of $0.91 \text{ h} < t_{init} < 0.17 \text{ h}$ (regardless of the initial dose) will give the local optimum. These results indicate that the probability of finding the true optima is zero for $\sim 75\%$ of the iterative T|D search space and $\sim 26\%$ for the iterative D|T search space and we conclude that

it is important to consider a simultaneous approach even when an iterative sequential approach is possible (see Appendix 1 for more details).

4) Divide all the continuous variables except one into discrete variables and optimize on each combination of discrete values and the continuous variable separately. This approach will be very time consuming and might miss the optimal design due to too wide discretizations of the design variables. However this method will (under the assumption that a fine enough grid is defined and that the optimizer will find a global optimum in the single continuous dimension) find the optimal design and it is a global optimizer per se.

In this work we only compared the two run-time equivalent methods (1) and (2) but we expect the third method to always be less or equal to the simultaneous approach in terms of efficiency and in general more time consuming. Regarding method (4); this method is applicable when we assume that the optimization surface is quite flat between the discretization of the design space and also that we have few dimensions (otherwise the run-times will be huge).

To conclude; it is as good or better to optimize simultaneously compared too sequentially. This is simply because optimizing sequentially puts restrictions on the search space of possible design values. These restrictions make it harder, and sometimes impossible, for the sequential optimal design to find the global optimal design in a given search space. These methods for simultaneous optimization of multiple types of design variables are available in the current version of PopED, freely available on the web at http://poped.sourceforge.net.

Appendix 1: simultaneous versus iterative sequential optimization

A repeated sequential iterative optimization on the setup presented in Fig. 1 was performed where optimization was done on the PD sample and the Dose. This was repeated numerous times to assure that the highest determinant was found. Regardless of the initial values, the simultaneous approach always found the global optimum. This was repeated with 5 different combinations of initial values for the dose and sample time (randomly chosen).

For the iterative sequential approach different initial values were also tested. After the first iteration of Time first (T|D) or Dose first (D|T) optimization initial values were updated for the next iteration to the optimal value from the previous iteration. The search was considered to have converged when the design did not change in an iteration.

The results show that regardless of the initial values for the time and an initial value of the Dose > 1.625 the local optima in the T|D approach would be found (the circles area in Fig. 1).

Similarly, for the D|T approach, regardless of the initial value of the dose and with an initial sample time of >0.91 or <0.17 the local optimum would be found. Further, as an example, with an initial value of 0.2 the global optimum will only be found after ~ 15 iterations.

These results show that the iterative sequential approach does not give the same results as the simultaneous approach. Further, in this simple case assuming a randomly distributed initial value over the design space, it can be shown that in $\sim$75% of the cases with the T|D approach the probability of finding the global optima is 0. Similarly the D|T approach will in $\sim$26% of the cases also have a probability of 0 finding the true optima while the simultaneous approach will have a probability > 0 finding the global optimum.

Appendix 2: simultaneous ED-optimal design versus sequential ED-optimal design

An ED-optimal design was tested with the simultaneous approach against the iterative sequential approach.

Model 1 was used (see Fig. 1) with an ED-uncertainty assigned to the model. The magnitude of the ED-uncertainty for the effect versus time plot is presented in Fig. 7. This effect is presented as a visual predictive check with 1000 simulations given the model (and the uncertainty). The ED-FIM surface was calculated with 90

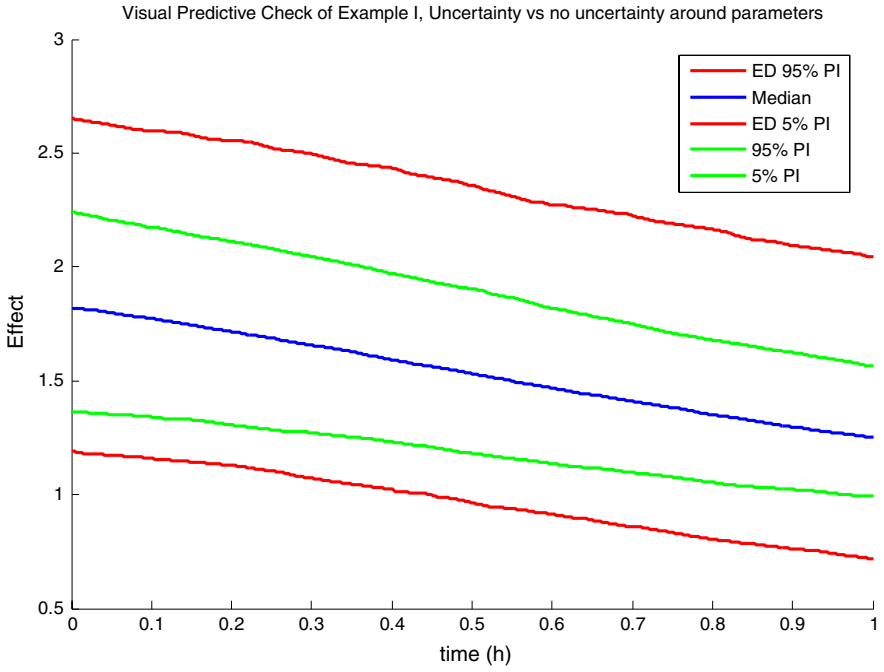

Fig. 7 The magnitude of the ED-uncertainty for the predicted effect. The outermost lines are the 95% and 5% prediction interval for the effect when having ED-uncertainty assigned to the model. The middle line shows the median of the predicted effect and the lines outside the median shows the 95% and 5% prediction interval without ED-uncertainty

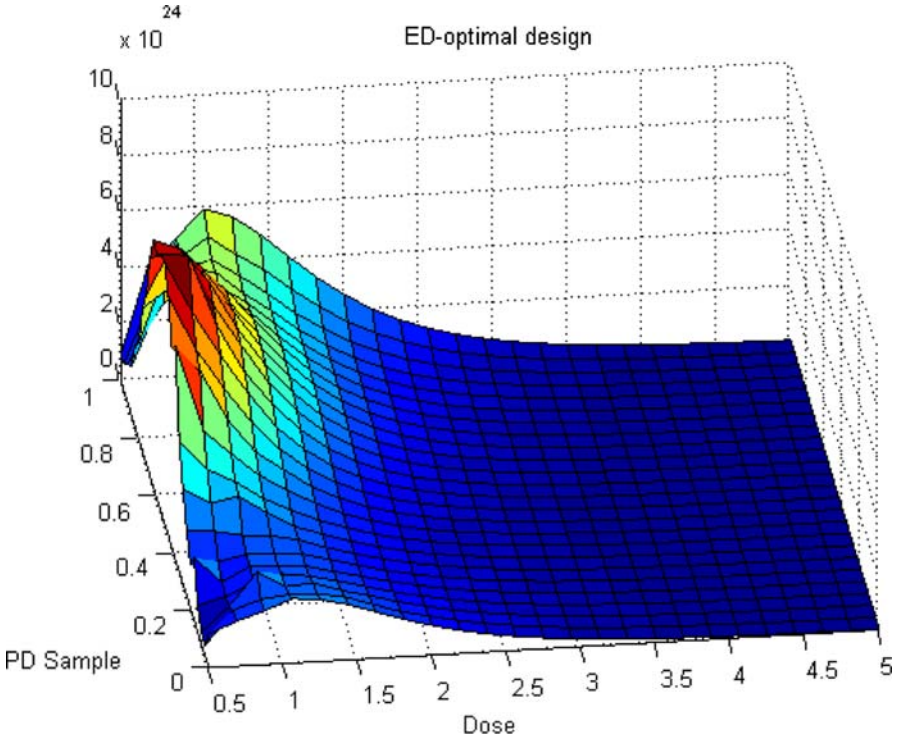

Fig. 8 The surface of the expected determinant of FIM against two of the design variables (sample time and dose) when ED-uncertainty was assigned to the model

Latin Hyper Cube samples from the distribution around the parameters with Monte Carlo Integration.

Figure 8 shows the FIM surface for this ED optimal design. Although not tested because of runtime issues, it is believed that with an initial dose >1 and using the time first approach (T|D) the probability of finding the global optimum will be 0 as in Appendix 1. In this example the advantage of the simultaneous approach is even more pronounced than for D-optimal design because of the spikier surface of the ED-optimal FIM. However, a general rule as to when an ED-optimal design will suffer more than a D-optimal design when optimizing sequentially is not presented here.

Appendix 3

See Fig. 9.

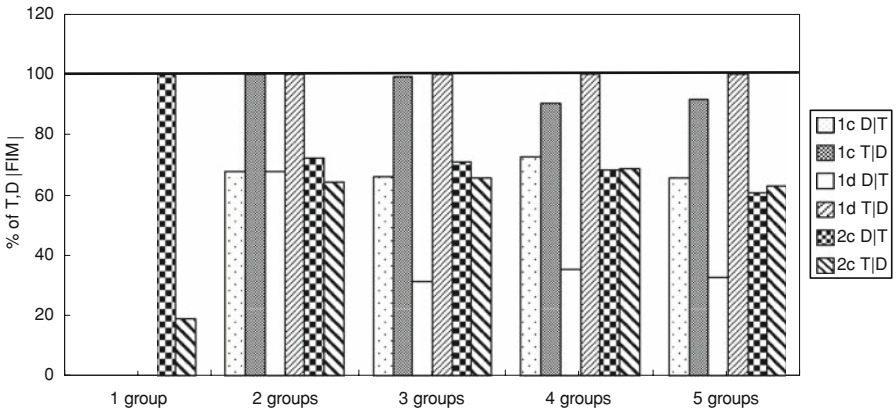

Fig. 9 Fraction of the simultaneous (T,D) D-optimal FIM for setup 1c, 1d and 2c. The black horizontal line represents the T,D approach

References

1. Mentré F, Mallet A, Baccar D (1997) Optimal design in random-effects regression models. Biometrika 84(2):429–442. doi:10.1093/biomet/84.2.429
2. Retout S, Duffull S, Mentre F (2001) Development and implementation of the population Fisher information matrix for the evaluation of population pharmacokinetic designs. Comput Methods Prog Biomed 65(2):141–151. doi:10.1016/S0169-2607(00)00117-6
3. Waterhouse TH et al (2005) Optimal design for model discrimination and parameter estimation for itraconazole population pharmacokinetics in cystic fibrosis patients. J Pharmacokinet Pharmacodyn 32(3–4):521–545. doi:10.1007/s10928-005-0026-2
4. Duffull SB, Mentre F, Aarons L (2001) Optimal design of a population pharmacodynamic experiment for ivabradine. Pharm Res 18(1):83–89. doi:10.1023/A:1011035028755
5. Foracchia M et al (2004) POPED, a software for optimal experiment design in population kinetics. Comput Methods Prog Biomed 74(1):29–46. doi:10.1016/S0169-2607(03)00073-7
6. Mould DR, Denman NG, Duffull S (2007) Using disease progression models as a tool to detect drug effect. Clin Pharmacol Ther 82(1):81–86. doi:10.1038/sj.clpt.6100228
7. Zhang L, Beal SL, Sheiner LB (2003) Simultaneous vs. sequential analysis for population PK/PD data I: best-case performance. J Pharmacokinet Pharmacodyn 30(6):387–404. doi:10.1023/B:JOPA.0000012998.04442.1f
8. Gueorguieva I et al (2007) A program for individual and population optimal design for univariate and multivariate response pharmacokinetic-pharmacodynamic models. Comput Methods Prog Biomed 86(1):51–61. doi:10.1016/j.cmpb.2007.01.004
9. Retout S, Mentré F (2003) Further developments of the Fisher information matrix in nonlinear mixed effects models with evaluation in population pharmacokinetics. J Biopharm Stat 13(2):209–227. doi:10.1081/BIP-120019267
10. Wang Y (2007) Derivation of various NONMEM estimation methods. J Pharmacokinet Pharmacodyn 34(5):575–593. doi:10.1007/s10928-007-9060-6
11. Beal SL, Sheiner LB (1998) NONMEM User's Guide. Technical report. University of California, San Francisco
12. Brazolli C, Retout S, Mentré F (2007) Population design in nonlinear mixed effects multiple responses models: extension of PFIM and evaluation by simulation with NONMEM and MONOLIX, p 16, Abstr 1176 [www.page-meeting.org/?abstract=1176]
13. MONOLIX: version 2.3 (2007) http://www.monolix.org
14. Atkinson AC, Donev AN (1992) Optimum experimental designs. Oxford University Press, Oxford
15. Hooker A, Vicini P (2005) Simultaneous population optimal design for pharmacokinetic-pharmacodynamic experiments. AAPS J. 7(4):E759–E785. doi:10.1208/aapsj070476

16. Hashimoto Y, Sheiner LB (1991) Designs for population pharmacodynamics: value of pharmacokinetic data and population analysis. J Pharmacokinet Biopharm 19(3):333–353
17. Beal SL (1983) Computation of the explicit solution to the Michaelis-Menten equation. J Pharmacokinet Biopharm 11(6):641–657. doi:10.1007/BF01059062
18. Pronzato L, Walter E (1985) Robust experiment design via stochastic approximation. Math Biosci 75:103–120. doi:10.1016/0025-5564(85)90068-9
19. Tod M et al (1998) Robust optimal design for the estimation of hyperparameters in population pharmacokinetics. J Pharmacokinet Biopharm 26(6):689–716. doi:10.1023/A:1020703007613
20. Dodds MG, Hooker AC, Vicini P (2005) Robust population pharmacokinetic experiment design. J Pharmacokinet Pharmacodyn 32(1):33–64. doi:10.1007/s10928-005-2102-z
21. Dokoumetzidis A, Aarons L (2007) Bayesian optimal designs for pharmacokinetic models: sensitivity to uncertainty. J Biopharm Stat 17(5):851–867. doi:10.1080/10543400701514007