

OPTIMAL DESIGN FOR POPULATION PK/PD MODELS

SERGEI LEONOV — ALEXANDER ALIEV

ABSTRACT. We provide some details of the implementation of optimal design algorithm in the PkStaMp library which is intended for constructing optimal sampling schemes for pharmacokinetic (PK) and pharmacodynamic (PD) studies. We discuss different types of approximation of individual Fisher information matrix and describe a user-defined option of the library.

1. Introduction

Optimal design of experiments for population PK/PD studies received a considerable attention in statistical literature and software development over the last decade. Discussions of the theory of optimal experimental design for nonlinear mixed effects models and its applications in drug development got a fresh start with the creation in 2006 of the annual PODE workshop (Population Optimum Design of Experiments); see

<http://www.maths.qmul.ac.uk/bb/PODE/PODE.html>.

The discussion of different software tools for population PK/PD optimal designs started at PODE 2007 and continued at every workshop since then. The discussed software packages include PFIM (see Bazzoli et al. [6]), PkStaMp (see Aliev et al. [2]), PopDes (see Gueorguieva et al. [13]), PopED (see Nyberg et al. [20]), and WinPOPT (see Duffull et al. [7]).

In this paper we continue the discussion of software tools for optimization of sampling schemes for PK/PD models. In Aliev et al. [1], [2] we described the PkStaMp library for constructing locally D -optimal designs for population compartmental PK and PK/PD models. The library is written in Matlab and compiled as a single executable file which does not require a Matlab license.

© 2012 Mathematical Institute, Slovak Academy of Sciences.

2010 Mathematics Subject Classification: Primary: 62K05; Secondary: 62P10.

Keywords: optimal model-based design; Fisher information matrix; D -optimality, PK/PD models, compartmental models.

The main focus of this paper is on different types of approximation of individual Fisher information matrices. We also describe a user-defined option as implemented in the PkStaMp library.

2. Model

The standard model of observations that we use in the PkStaMp library is as follows:

$$y_{ij} = \eta(x_{ij}, \gamma_i) + v_{ij}, \quad i = 1, \dots, N, \quad j = 1, \dots, k_i, \quad (1)$$

where x_{ij} are times of taking PK or PD measurements y_{ij} for patient i ; k_i is the number of measurements for patient i ; N is the total number of patients in the study; $\eta(x, \gamma)$ is the response function; γ_i is a vector of individual parameters of patient i ; v_{ij} are residual errors which have additive and proportional components of variability,

$$v_{ij} = \varepsilon_{1,ij} + \varepsilon_{2,ij}\eta(x_{ij}, \gamma_i), \quad (2)$$

where $\varepsilon_{1,ij}$, $\varepsilon_{2,ij}$ are random variables with zero mean, such that vectors $\varepsilon_{1,i} = (\varepsilon_{1,i}, \dots, \varepsilon_{1,ik_i})^T$ and $\varepsilon_{2,i'} = (\varepsilon_{2,i'}, \dots, \varepsilon_{2,i'k_{i'}})^T$ are mutually independent for all i, i' , and

$$\mathbf{E}(\varepsilon_{1,i}\varepsilon_{1,i}^T) = \sigma_A^2 \mathbf{I}_{k_i}, \quad \mathbf{E}(\varepsilon_{2,i}\varepsilon_{2,i}^T) = \sigma_P^2 \mathbf{I}_{k_i},$$

where \mathbf{I}_k denotes a $(k \times k)$ -identity matrix.

In compartmental modeling, the amounts of drug at different compartments satisfy the system of ordinary differential equations. For example, for a one-compartment model with first-order absorption and linear elimination, with a single dose D administered at time $x = 0$,

$$\begin{cases} \dot{g}_0(x) = -K_a g_0(x), & g_0(0) = D, \\ \dot{g}_1(x) = K_a g_0(x) - K_e g_1(x), & g_1(0) = 0, \end{cases} \quad (3)$$

where $g_0(x)$ is the amount of drug at the site of administration at time x , $g_1(x)$ is the amount of drug in the central compartment, K_a and K_e are absorption and elimination rate constants, respectively, and $\eta(x, \gamma) = g_1(x)/V$ is the drug concentration in the central compartment

$$\eta(x, \gamma) = \frac{DK_a}{V(K_a - K_e)} \left(e^{-K_e x} - e^{-K_a x} \right), \quad (4)$$

where $\gamma = (K_a, K_e, V)^T$ is the vector of response parameters, and V is the volume of distribution. It is assumed that the individual response parameters γ_i are independently sampled from a given population, either normal,

$$\gamma_i \sim \mathcal{N}(\gamma^0, \mathbf{\Omega}), \quad (5)$$

or log-normal,

$$\gamma_{il} = \gamma_l^0 e^{\zeta_{il}}, \quad \boldsymbol{\gamma}_0 = \left(\gamma_1^0, \dots, \gamma_{m_\gamma}^0\right)^T, \quad \boldsymbol{\zeta}_i = (\zeta_{i1}, \dots, \zeta_{im_\gamma})^T \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}), \quad (6)$$

where m_γ is the dimension of the vector of response parameters; $l = 1, \dots, m_\gamma$; $\boldsymbol{\gamma}^0$ is an $(m_\gamma \times 1)$ -vector of population means, and $\boldsymbol{\Omega}$ is an $(m_\gamma \times m_\gamma)$ variance-covariance matrix of i.i.d. random vectors $\boldsymbol{\gamma}_i$ in (5) or $\boldsymbol{\zeta}_i$ in (6). The vector $\boldsymbol{\gamma}^0$ and the matrix $\boldsymbol{\Omega}$ are often referred to as population parameters. By $\boldsymbol{\theta} = (\boldsymbol{\gamma}^0, \boldsymbol{\Omega}; \sigma_A^2, \sigma_P^2)$ we denote the combined vector of model parameters to be estimated, and by m its length.

3. Fisher information matrix and optimal designs

The key object in estimation and optimal design for repeated measures models is the Fisher information matrix $\boldsymbol{\mu}(\mathbf{x}, \boldsymbol{\theta})$ of a k -dimensional point \mathbf{x} which in PK/PD applications corresponds to a $(k \times 1)$ -sequence of sampling times. Once the information matrix $\boldsymbol{\mu}(\mathbf{x}, \boldsymbol{\theta})$ for any candidate sequence \mathbf{x} is defined, then one can calculate the normalized Fisher information matrix

$$\mathbf{M}(\boldsymbol{\xi}, \boldsymbol{\theta}) = \sum_u w_u \boldsymbol{\mu}(\mathbf{x}_u, \boldsymbol{\theta}) \quad (7)$$

for any continuous design $\boldsymbol{\xi} = \{(\mathbf{x}_u, w_u)\}$, where $\sum_u w_u = 1$ and weights w_u correspond to relative frequencies of sequences \mathbf{x}_u , $u = 1, \dots, n$. In the PkStaMp library we minimize the D -optimality criterion,

$$\boldsymbol{\xi}^* = \arg \min_{\boldsymbol{\xi}} |\mathbf{M}^{-1}(\boldsymbol{\xi}, \boldsymbol{\theta})|, \quad (8)$$

where the optimization is performed with respect to continuous designs $\boldsymbol{\xi}$, and sequences \mathbf{x}_u in (7) belong to a pre-specified design region \mathfrak{X} . We implement the first-order optimization algorithm with forward and backward steps which goes back to the publications of Wynn [24], Fedorov [8], Atwood [4]. See also Atkinson and Donev [3], Fedorov and Hackl [10].

3.1. Numerical procedure

Let $\boldsymbol{\xi}_s$ be the current design on step s . First, a sequence \mathbf{x}_s^+ is found such that

$$\mathbf{x}_s^+ = \arg \max_{\mathbf{x} \in \mathfrak{X}} d(\mathbf{x}, \boldsymbol{\xi}_s, \boldsymbol{\theta}) \quad (\text{forward step}), \quad (9)$$

where $d(\mathbf{x}, \boldsymbol{\xi}, \boldsymbol{\theta}) = \text{tr} [\mathbf{M}^{-1}(\boldsymbol{\xi}, \boldsymbol{\theta}) \boldsymbol{\mu}(\mathbf{x}, \boldsymbol{\theta})]$ is the sensitivity function of the D -criterion, and the updated design $\boldsymbol{\xi}_s^+$ is obtained according to

$$\boldsymbol{\xi}_s^+ = (1 - \alpha_s) \boldsymbol{\xi}_s + \alpha_s \boldsymbol{\xi}(\mathbf{x}_s), \quad 0 < \alpha_s < 1, \quad (10)$$

where $\xi(\mathbf{x})$ is the unit measure atomized at \mathbf{x} , i.e., the design supported on the single sequence \mathbf{x} . Then the worst sequence in the current design is found according to

$$\mathbf{x}_s^- = \arg \min_{\mathbf{x} \in \mathfrak{X}_s^+} d(\mathbf{x}, \xi_s^+, \boldsymbol{\theta}) \quad (\text{backward step}), \quad (11)$$

where \mathfrak{X}_s^+ is the support set of the design ξ_s^+ , and the design ξ_{s+1} is obtained from

$$\xi_{s+1} = (1 - \alpha'_s)\xi_s^+ + \alpha'_s\xi(\mathbf{x}_s^-), \quad \alpha'_s = -\min\left(\alpha_s, \frac{p_s}{1 - p_s}\right), \quad (12)$$

where p_s is the weight of sequence \mathbf{x}_s^- in the design ξ_s^+ . The sequence of positive numbers α_s must satisfy standard conditions $\sum_s \alpha_s = \infty$, $\alpha_s \rightarrow 0$ as $s \rightarrow \infty$, for example, $\alpha_s \sim 1/s$.

As described in Aliev et al. [2], the design region \mathfrak{X} in the PkStaMp library can be defined in two possible ways which we consider rather practical.

- (1) The user selects the number k of samples and a finite set of candidate sampling times (say, every hour, every half-hour etc. over the specified time window), and then the program enumerates all possible sequences of length k from this set and uses these sequences as the design region.
- (2) The user saves arbitrary candidate sequences in a file, and then the program uses these sequences as the design region.

It is worthwhile to remark that if the information matrix $\boldsymbol{\mu}(\mathbf{x}, \boldsymbol{\theta})$ can be calculated (approximated) for any candidate ‘‘observational unit’’ \mathbf{x} in the design region \mathfrak{X} , then the construction of locally optimal designs follows a rather simple routine for either relatively simple models (fixed effects, one measurement per observational unit), or more complex models where several measurements are taken for a one or more dependent variables, as in our examples of serial sampling in PK studies. Indeed, no matter how complex the underlying model is, if the individual information matrix $\boldsymbol{\mu}(\mathbf{x}, \boldsymbol{\theta})$ is defined for all $\mathbf{x} \in \mathfrak{X}$, then the problem of finding the locally optimal design ξ^* is reduced to the optimization problem (8) in the space of information matrices $\{\mathbf{M}(\xi, \boldsymbol{\theta})\}$. Moreover, by construction, the design region \mathfrak{X} is finite for all models in our library. Therefore, the forward step (9) is reduced to optimization over the finite number of candidate sequences which allows us to calculate individual information matrices prior to running the optimal design algorithm. For more details on the numerical procedure and its settings, see Fedorov et al. [9, Section 7.1.5], Aliev et al. [2].

3.2. Approximation of the information matrix

If a vector of observations \mathbf{Y} has mean $E(\mathbf{Y}|\mathbf{x}) = \boldsymbol{\eta}(\mathbf{x}, \boldsymbol{\theta})$ and variance $\text{Var}(\mathbf{Y}|\mathbf{x}) = \mathbf{S}(\mathbf{x}, \boldsymbol{\theta})$, then for normally distributed \mathbf{Y} there exists a closed-form

expression for the information matrix $\boldsymbol{\mu}(\mathbf{x}, \boldsymbol{\theta})$, see Magnus and Neudecker [14]

$$\boldsymbol{\mu}_{\alpha\beta}(\mathbf{x}, \boldsymbol{\theta}) = \frac{\partial \boldsymbol{\eta}^T}{\partial \theta_\alpha} \mathbf{S}^{-1} \frac{\partial \boldsymbol{\eta}}{\partial \theta_\beta} + \frac{1}{2} \text{tr} \left[\mathbf{S}^{-1} \frac{\partial \mathbf{S}}{\partial \theta_\alpha} \mathbf{S}^{-1} \frac{\partial \mathbf{S}}{\partial \theta_\beta} \right], \quad (13)$$

$$\boldsymbol{\eta} = \boldsymbol{\eta}(\mathbf{x}, \boldsymbol{\theta}), \quad \mathbf{S} = \mathbf{S}(\mathbf{x}, \boldsymbol{\theta}),$$

where $\alpha, \beta = 1, \dots, m$. Let $\boldsymbol{\eta}(\mathbf{x}, \boldsymbol{\theta}) = [\eta(x_1, \boldsymbol{\theta}), \dots, \eta(x_k, \boldsymbol{\theta})]^T$, and let $\mathbf{F} = \mathbf{F}(\mathbf{x}, \boldsymbol{\gamma}^0) = [\partial \boldsymbol{\eta}(\mathbf{x}, \boldsymbol{\theta}) / \partial \gamma_\alpha]_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^0}$ be a $(k \times m_\gamma)$ matrix of partial derivatives of $\boldsymbol{\eta}(\mathbf{x}, \boldsymbol{\theta})$ with respect to response parameters $\boldsymbol{\gamma}^0$. Using the first-order approximation (first-order Taylor expansion) together with (2), for normally distributed γ_i one gets

$$\mathbf{S}(\mathbf{x}, \boldsymbol{\theta}) \simeq \mathbf{F} \boldsymbol{\Omega} \mathbf{F}^T + \sigma_P^2 \text{Diag}[\boldsymbol{\eta}(\mathbf{x}, \boldsymbol{\theta}) \boldsymbol{\eta}^T(\mathbf{x}, \boldsymbol{\theta}) + \mathbf{F} \boldsymbol{\Omega} \mathbf{F}^T] + \sigma_A^2 \mathbf{I}_k. \quad (14)$$

For log-normally distributed γ_i as in (6), the matrix $\boldsymbol{\Omega}$ on the right-hand side of (14) has to be replaced with

$$\tilde{\boldsymbol{\Omega}} = \text{Diag}(\boldsymbol{\gamma}^0) \boldsymbol{\Omega} \text{Diag}(\boldsymbol{\gamma}^0). \quad (15)$$

In the above formulae, $\text{Diag}(\mathbf{a})$ denotes a diagonal matrix with diagonal elements equal to either a_{ll} when \mathbf{a} is a square matrix, or a_l when \mathbf{a} is a vector. For the derivation of (14)–(15), see the Appendix; cf. Gagnon and Leonov [12]. The formulae (14)–(15) are used in (13) to approximate the individual information matrix $\boldsymbol{\mu}(\mathbf{x}, \boldsymbol{\theta})$. Note also that while in (13) we use notation $\boldsymbol{\eta}(\mathbf{x}, \boldsymbol{\theta})$, in fact, when using the first-order approximation, the function $\boldsymbol{\eta}$ depends only on response parameters $\boldsymbol{\gamma}^0$ as in (4).

4. Software comparison

In 2009–2010 participants of PODE workshop performed the comparison of different software tools using the one-compartment model (3) as an example. The settings were proposed by Nick Holford and France Mentré and were based on data from the earlier warfarine study. The model was parameterized via clearance CL , so that $K_e = CL/V$. It was assumed that the individual response parameters $\boldsymbol{\gamma}_i = (K_{ai}, CL_i, V_i)$ follow the log-normal distribution (6) with the mean vector $\boldsymbol{\gamma}^0 = (1, 0.15, 8)$ and the diagonal covariance matrix $\boldsymbol{\Omega} = \text{Diag}(\omega_r^2) = \text{Diag}(0.6, 0.07, 0.02)$. It was also assumed that the additive component of residual variance vanishes, $\sigma_A^2 = 0$, and that $\sigma_P^2 = 0.01$. So the combined vector of parameters for this example was

$$\boldsymbol{\theta} = (k_a^0, CL^0, V^0; \omega_{k_a}^2, \omega_{CL}^2, \omega_V^2; \sigma_P^2)^T = (1, 0.15, 8; 0.6, 0.07, 0.02; 0.01). \quad (16)$$

The goal was to compare the Fisher information matrix for the 8-sample sequence

$$\mathbf{x} = (0.5, 1, 2, 6, 24, 36, 72, 120) \text{ hours.} \quad (17)$$

All software developers reported coefficients of variation

$$CV_s = \sqrt{[\boldsymbol{\mu}^{-1}(\mathbf{x}, \boldsymbol{\theta})]_{ss} / N / \theta_s}, \quad s = 1, 2, \dots, 7,$$

where $N = 32$ was the number of patients in the actual warfarine study. Thus values CV_s can be interpreted as coefficients of variation obtained from the design which uses the sequence \mathbf{x} from (17) for all 32 patients in the study.

Monte Carlo simulation studies were also performed in NONMEM (by Joakim Nyberg) and MONOLIX, see [18] (by Caroline Bazzoli), and sample estimates of the coefficients of variation were obtained: for a single run, data were generated according to the model (1)–(4), (6) with parameters $\boldsymbol{\theta}$ from (16) and the sampling sequence \mathbf{x} from (17) for 32 patients, and parameters were estimated using the nonlinear mixed effects models estimation algorithm from either NONMEM or MONOLIX. Then after 1000 runs, sample statistics and coefficients of variations were calculated.

After comparing the outputs, it turned out that all population design tools produced similar coefficients of variation for all model parameters except the absorption rate K_a : $CV(K_a) = 0.052$ for PkStaMp and PopDes while $CV(K_a) = 0.139$ for other tools. Simulations in both NONMEM and MONOLIX resulted in estimates $CV(K_a) = 0.12 \div 0.13$. The discrepancy between the outputs suggested to look closer at how calculations were implemented by different software developers.

The matrix $\boldsymbol{\mu}$ in (13) may be written in the block-diagonal form; e.g., see Retout and Mentré [23]

$$\boldsymbol{\mu}(\mathbf{x}, \boldsymbol{\theta}) = \left\{ \begin{array}{cc} \mathbf{A} & \mathbf{C} \\ \mathbf{C}^T & \mathbf{B} \end{array} \right\}, \quad \mathbf{A} = \mathbf{A}_1 + \mathbf{A}_2, \quad \mathbf{A}_1 = \mathbf{F}^T \mathbf{S}^{-1} \mathbf{F}, \quad (18)$$

where for our specific example

$$\mathbf{A}_{2,\alpha\beta} = \frac{1}{2} \text{tr} \left[\mathbf{S}^{-1} \frac{\partial \mathbf{S}}{\partial \theta_\alpha} \mathbf{S}^{-1} \frac{\partial \mathbf{S}}{\partial \theta_\beta} \right], \quad \alpha, \beta = 1, 2, 3; \quad (19)$$

$$\mathbf{C}_{\alpha\beta} = \frac{1}{2} \text{tr} \left[\mathbf{S}^{-1} \frac{\partial \mathbf{S}}{\partial \theta_\alpha} \mathbf{S}^{-1} \frac{\partial \mathbf{S}}{\partial \theta_\beta} \right], \quad \alpha = 1, 2, 3, \quad \beta = 4, \dots, 7; \quad (20)$$

$$\mathbf{B}_{\alpha\beta} = \frac{1}{2} \text{tr} \left[\mathbf{S}^{-1} \frac{\partial \mathbf{S}}{\partial \theta_\alpha} \mathbf{S}^{-1} \frac{\partial \mathbf{S}}{\partial \theta_\beta} \right], \quad \alpha, \beta = 4, \dots, 7.$$

So \mathbf{A} is the (3×3) -matrix which contains partial derivatives with respect to response parameters γ_α ; \mathbf{C} is the (3×4) -matrix which contains mixed derivatives with respect to response parameters γ_α and variance parameters $(\omega_\beta^2, \sigma_P^2)$; and

\mathbf{B} is the (4×4) -matrix which contains derivatives with respect to all variance parameters $(\omega_\beta^2, \sigma_P^2)$.

In PkStaMp we used the first-order approximation (14), (15) and the full matrix $\boldsymbol{\mu}(\mathbf{x}, \boldsymbol{\theta})$ in (18). It turned out that the other software tools (PFIM, PopED, WinPOPT) used the following settings:

- “Exclude” the matrix \mathbf{A}_2 in the calculation of the matrix \mathbf{A} in (18) and use $\mathbf{A}_2 = \mathbf{0}$.
- Exclude the matrix \mathbf{C} in the calculation of the matrix $\boldsymbol{\mu}$ in (18) and use $\mathbf{C} = \mathbf{0}$ instead.
- Also exclude the term $\mathbf{F}\boldsymbol{\Omega}\mathbf{F}^T$ in the square brackets on the right-hand side of (14).

These differences led to quite visible differences in the elements of the information matrix $\boldsymbol{\mu}$ which correspond to the absorption rate K_a . Once we made the initial settings identical, the output results coincided, too. Still several questions remained, in particular (a) what are the consequences of the first-order approximation in (14) and (15), and (b) which option is preferable, the “full” where the matrices \mathbf{A}_2, \mathbf{C} are preserved, or the “reduced” where $\mathbf{A}_2 = \mathbf{C} = \mathbf{0}$?

4.1. Linearization for log-normal population distribution

Suppose that ζ is a normal random variable, $\zeta \sim \mathcal{N}(0, \omega^2)$, and that $\gamma = e^\zeta$. Then the first-order approximation leads to

$$\gamma \approx 1 + \zeta, \quad \mathbb{E}\gamma \approx 1, \quad \text{Var}(\gamma) \approx \mathbb{E}\zeta^2 = \omega^2. \quad (21)$$

On the other hand, the exact moments of the log-normally distributed random variable γ are

$$\mathbb{E}\gamma = e^{\omega^2/2}, \quad \text{Var}(\gamma) = e^{\omega^2} (e^{\omega^2} - 1), \quad (22)$$

and, therefore, when ω^2 is not too small, the first-order approximation may lead to substantial distortion of the distribution in general, and moments in particular. In our example $\omega_{K_a}^2 = 0.6$, and the analogs of (21) and (22) are

$$\mathbb{E}K_{ai} \approx 1, \quad \text{Var}(K_{ai}) \approx 0.6, \quad (23)$$

and

$$\mathbb{E}K_{ai} = 1.35, \quad \text{Var}(K_{ai}) = 1.5, \quad (24)$$

respectively. For more discussion on linearization options, see Mielke and Schwabe [17].

4.2. Using $\mathbf{A}_2 = \mathbf{0}$ in (18)

To get an idea about the effect of setting $\mathbf{A}_2 = \mathbf{0}$ in (18), consider a single-response fixed effects model (1), i.e., $\boldsymbol{\Omega} = \mathbf{0}$. Let $\{\varepsilon\}$ in (2) be normally distributed, with known residual variances $\sigma_A^2 = 0$ and σ_P^2 . In this case

$\text{Var}(y_{ij}) = \sigma_P^2 \eta^2(x_{ij}, \gamma^0)$, so the term \mathbf{A}_2 in (19) reduces to $\mathbf{A}_2 = 2 \frac{\mathbf{F}^T \mathbf{F}}{\eta^2}$, and the formula (13) for the Fisher information matrix becomes

$$\boldsymbol{\mu}_F(x, \boldsymbol{\theta}) = \frac{1}{\sigma_P^2} \frac{\mathbf{F}^T \mathbf{F}}{\eta^2} + 2 \frac{\mathbf{F}^T \mathbf{F}}{\eta^2} = \left(\frac{1}{\sigma_P^2} + 2 \right) \frac{\mathbf{F}^T \mathbf{F}}{\eta^2}. \quad (25)$$

When the term \mathbf{A}_2 is not used, then the “reduced” information matrix is

$$\boldsymbol{\mu}_R(x, \boldsymbol{\theta}) = \frac{1}{\sigma_P^2} \frac{\mathbf{F}^T \mathbf{F}}{\eta^2}. \quad (26)$$

To evaluate the effect of the missing term \mathbf{A}_2 on the coefficient of variation, one can check the ratio

$$\sqrt{\frac{\mu_{\alpha\beta,F}}{\mu_{\alpha\beta,R}}} = \sqrt{\frac{2 + 1/\sigma_P^2}{1/\sigma_P^2}} = \sqrt{1 + 2\sigma_P^2} \sim 1 + \sigma_P^2 \quad \text{for small } \sigma_P^2. \quad (27)$$

So (27) suggests that for our example with $\sigma_P^2 = 0.01$, the effect of dropping the term \mathbf{A}_2 may be minimal. Once the proportional residual variance σ_P^2 becomes larger, then the effect of the missing term \mathbf{A}_2 may be more pronounced.

Note also that in the considered example the Fisher information matrix $\boldsymbol{\mu}_F(x, \boldsymbol{\theta})$ coincides, up to the coefficient of proportionality, with the information (moment) matrix $\boldsymbol{\mu}_R(x, \boldsymbol{\theta})$ which corresponds to the iteratively reweighted least squares estimator $\hat{\boldsymbol{\theta}}_{\text{IRLS}}$; see Fedorov and Leonov [11, Section 5.4]. While the variance-covariance matrix of $\hat{\boldsymbol{\theta}}_{\text{IRLS}}$ is larger than the one of the maximum likelihood estimator $\hat{\boldsymbol{\theta}}_{\text{MLE}}$, the optimal designs for the two methods are the same.

4.3. Linearization and using $\mathbf{C} = \mathbf{0}$ in (18)

We do not have a good explanation of why the reduced version of the information matrix $\boldsymbol{\mu}(\mathbf{x}, \boldsymbol{\theta})$ with $\mathbf{C} = \mathbf{0}$ led to a “better” approximation that was closer to the simulations in NONMEM and MONOLIX. A possible heuristic explanation is that setting $\mathbf{C} = \mathbf{0}$ helped in some way to balance the effect of the distortion due to the first-order approximation; see (23), (24). From a practical point of view, taking $\mathbf{C} = \mathbf{0}$ and, therefore, using the simpler covariance structure may simplify the estimation algorithm. However, in general, we do not see any solid reasons for using $\mathbf{C} = \mathbf{0}$ instead of considering the full matrix with \mathbf{C} defined in (20).

4.4. Other types of approximation

If one uses the second-order approximation of the response function $\eta(x, \gamma_i)$ in the vicinity of γ^0 , then it follows from (5) that the expectation of $\eta(x, \gamma_i)$ with respect to the distribution of parameters γ_i can be approximated as

$$\mathbb{E}[\eta(x, \gamma_i)] \approx \eta(x, \gamma^0) + \frac{1}{2} \text{tr} [\mathbf{H}(x, \gamma^0) \boldsymbol{\Omega}], \quad (28)$$

where $\mathbf{H}(x, \boldsymbol{\gamma}^0)$ is the matrix of second-order partial derivatives of the response function,

$$\mathbf{H}(x, \boldsymbol{\gamma}^0) = \left[\frac{\partial^2 \eta(x, \boldsymbol{\gamma})}{\partial \gamma_\alpha \partial \gamma_\beta} \right] \Big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^0},$$

e.g., see Fedorov and Leonov [11, Section 5.5.3]. As noted in Aliev et al. [2], the formula (14) for the variance matrix \mathbf{S} utilizes first derivatives \mathbf{F} of the response $\boldsymbol{\eta}$, and, therefore, calculation of the derivatives of \mathbf{S} in (13) requires second-order derivatives of $\boldsymbol{\eta}$. Thus, with the second-order approximation (28), one will require fourth-order derivatives of the response function $\boldsymbol{\eta}$ which numerically will be rather tedious.

One of potential ways of improving the calculation of the information matrix $\boldsymbol{\mu}(\mathbf{x}, \boldsymbol{\theta})$ and avoiding numerical approximation as in (14), (15) or (28), is to calculate the mean $\boldsymbol{\eta}(\mathbf{x}, \boldsymbol{\theta})$ and variance $\mathbf{S}(\mathbf{x}, \boldsymbol{\theta})$ via Monte Carlo simulations at each candidate sampling time x_j :

- Generate L independent realizations of response parameters $\boldsymbol{\gamma}_i$ from a given distribution (5) or (6), $i = 1, \dots, L$.
- Generate values $\mathbf{Y}_i = \{y_{ij}\}$ according to (1) and (2), with $x_{ij} \equiv x_j$ for all i .
- Calculate empirical mean and variance

$$\begin{aligned} \hat{\boldsymbol{\eta}} &= \hat{\boldsymbol{\eta}}(\mathbf{x}, \boldsymbol{\theta}) = \hat{\mathbf{E}}_{\boldsymbol{\theta}}(\mathbf{Y}) = \frac{1}{L} \sum_{i=1}^L \mathbf{Y}_i, \\ \hat{\mathbf{S}} &= \widehat{\text{Var}}_{\boldsymbol{\theta}}(\mathbf{Y}) = \frac{1}{L-1} \sum_{i=1}^L (\mathbf{Y}_i - \hat{\boldsymbol{\eta}}) (\mathbf{Y}_i - \hat{\boldsymbol{\eta}})^T \end{aligned} \quad (29)$$

- Use the formula (13) to calculate $\boldsymbol{\mu}(\mathbf{x}, \boldsymbol{\theta})$ with values $\{\hat{\boldsymbol{\eta}}, \hat{\mathbf{S}}\}$ from (29).

Note that the described Monte Carlo approach will eliminate the need to calculate second- and higher-order derivatives of the response function since the formula (14) or its analogs will not be used. The limitation of this approach is that it still relies on the normal approximation (13).

Figure 1 illustrates three different types of approximation of the response function η . The solid line presents $\eta(x, \boldsymbol{\gamma}^0)$, i.e., the first-order approximation; see (21). The dashed line shows $\eta(x, \boldsymbol{\gamma}_{\log}^0)$ where $\boldsymbol{\gamma}_{\log}^0$ is the true mean of the log-normal distribution; see (22). The dotted line presents $\hat{\eta}(x)$ which is obtained via the Monte Carlo approach as in (29). The differences between the three curves are mostly pronounced during the absorption phase and at the beginning of the elimination phase (before and after the peak concentration, respectively). These differences can lead to the differences in the computation of the information matrix $\boldsymbol{\mu}(\mathbf{x}, \boldsymbol{\theta})$.

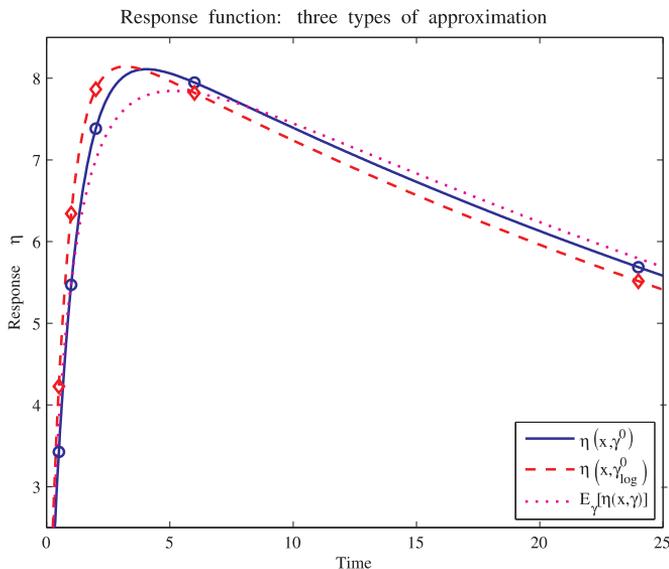


FIGURE 1. Mean response curves. Solid – the 1st order approximation, see (21); dashed – computed at mean values of log-normal distribution, see (24); dotted – Monte Carlo average as in (29). Locations of circles and diamonds correspond to sampling times from the sequence \mathbf{x} in (17).

5. User-defined option

A new benchmark example was proposed by France Mentré for PODE 2011 meeting. The example is based on hepatitis C viral dynamics model (HCV); see Neumann et al. [19] The model includes two components. The PK component is a one-compartment model which is similar to (3), but with continuous drug infusion of dose $D = 180$ mg for the duration of one day, repeated every week. Thus the first equation in (3) has to be replaced with

$$\dot{g}_0(x) = -K_a g_0(x) + r(x), \quad r(x) = \{D, \text{ if } x \in [x_l, x_l + 1], \text{ or } 0 \text{ otherwise}\},$$

where x_l are starting times of infusion, $x_l = 0, 7, 14, \dots$ (days). The PD model describes the dynamics of the number of “target cells” T , infected cells I and the viral load v

$$\begin{cases} \dot{T}(x) = -C_1 T(x) - C_2 T(x)v(x) + C_3, \\ \dot{I}(x) = -\delta I(x) + C_2 T(x)v(x), \\ \dot{v}(x) = -C_4 \left\{ 1 - \frac{1}{1 + [EC_{50}/\eta_1(x)]^n} \right\} I(x) - cv(x), \end{cases} \quad (30)$$

where $\eta_1(x) = \eta_1(x, \gamma) = g_1(x)/V$ is the drug concentration in the central compartment. The measured PD endpoint is $\eta_2(x) = \eta_2(x, \gamma) = \log_{10} v(x)$.

OPTIMAL DESIGN FOR POPULATION PK/PD MODELS

For the purposes of software comparison, it was assumed that the measurement error model (2) had the additive component only, and $\sigma_P^2 = 0$ for both PK and PD responses. It was also assumed that C_1, C_2, C_3, C_4 were fixed constants. In total, there were seven response parameters: three PK parameters (K_a, K_e, V) and four PD parameters (δ, EC_{50}, n, c). The log-parameterization was used for response parameters, i.e.,

$$\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_7) = (\ln K_a, \dots, \ln c),$$

with the normal distribution of parameters γ_i with the diagonal covariance matrix $\boldsymbol{\Omega} = \omega^2 \mathbf{I}_7$, $\omega^2 = 0.25$. The goal was to compare the Fisher information matrix for the 12-sample sequence

$$\mathbf{x} = (0, 0.25, 0.5, 1, 2, 3, 4, 7, 10, 14, 21, 28) \text{ days}, \quad (31)$$

with the simultaneous measurement of PK and PD responses η_1 and η_2 . For more details on model settings, see Mentré et al. [15].

To run the HCV example, we implemented a “user-defined” option in the PkStaMp library. This option allows the user to perform the following actions:

- Similar to built-in models, input model parameters $\boldsymbol{\gamma}^0$ and $\boldsymbol{\Omega}$; cf. Fig. 1 from Aliev et al. [2] and Fig. 2, top panel. In the HCV example, $K_a = 0.8$, so because of log-parameterization we used

$$\theta_1 = \log(0.8) = -0.223143 \quad \text{etc.}$$

- Input different types of dosing. For example, continuous infusion can be specified by starting times and durations; see Fig. 2, bottom right corner of the top panel.
- Similar to built-in models, specify residual error model and candidate sampling times for measured compartments; see Fig. 2, bottom panel.
- Type algebraic expressions on the right-hand side of differential equations into corresponding fields; see Fig. 3, top and bottom panels. In the algebraic expressions, by $\{A\}$ we denote amounts in different compartments, i.e., $A(1) = g_0, A(2) = g_1, \dots, A(5) = v$, and by $\{P\}$ we denote model parameters.

The user-defined option utilizes the numerical solution of the system of ODEs via Matlab built-in solver *ode45*.

In the PkStaMp run that was reported at PAGE 2011 meeting, we did not account for the dependence of the initial conditions of the system (30) on model parameters which led to minor differences in the reported coefficients of variation compared to other software tools; see Mentré et al. [15, Example 2]. This inconsistency was later corrected, and the resulting CVs from PkStaMp became identical to those from the other tools under the same assumptions.

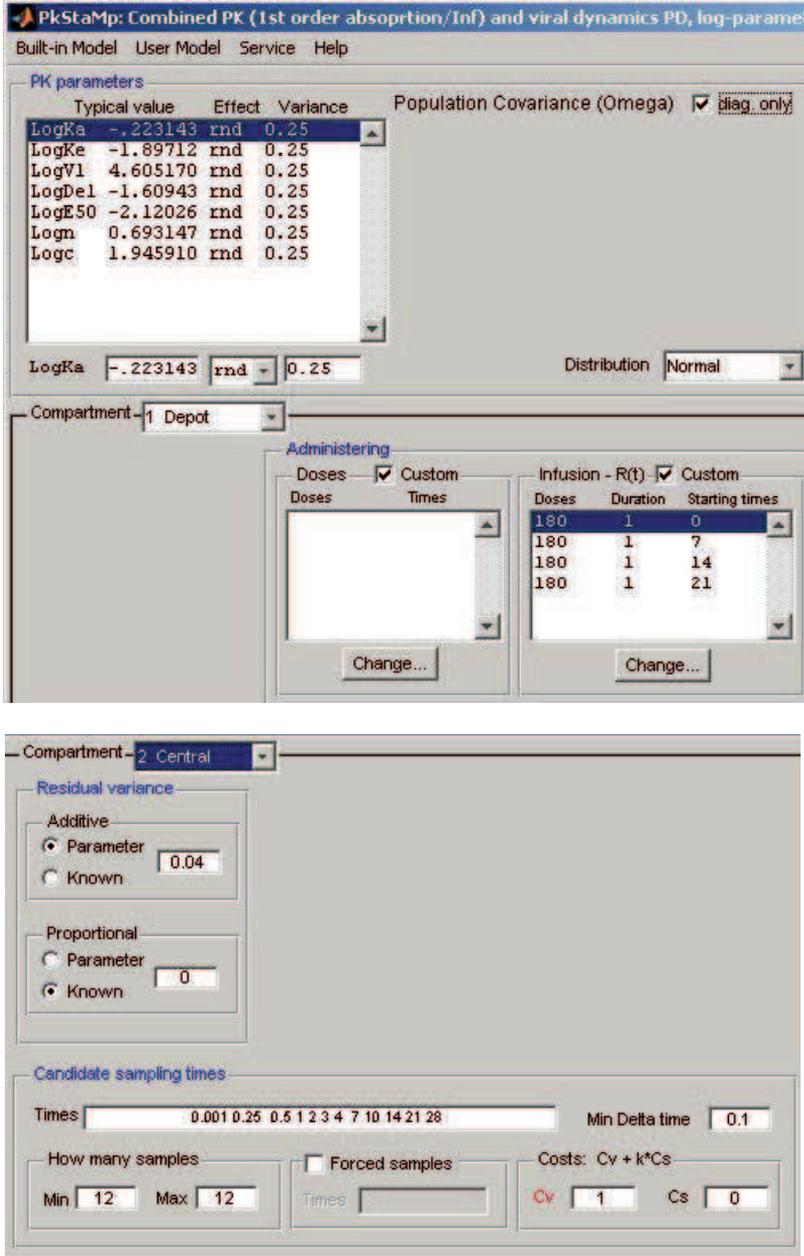


FIGURE 2. User-defined option, input screens: parameters and dosing regimen (top), residual error model and candidate sampling times (bottom).

OPTIMAL DESIGN FOR POPULATION PK/PD MODELS

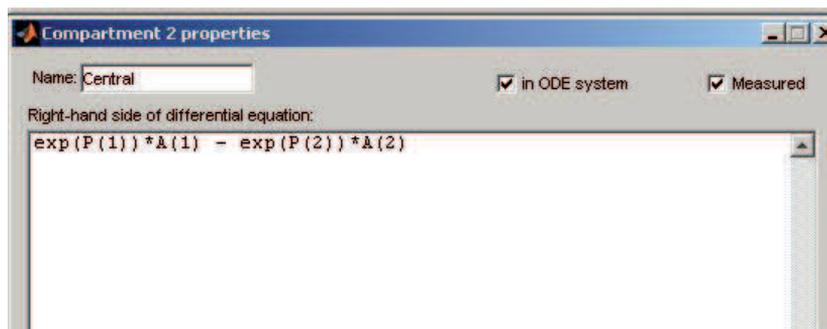
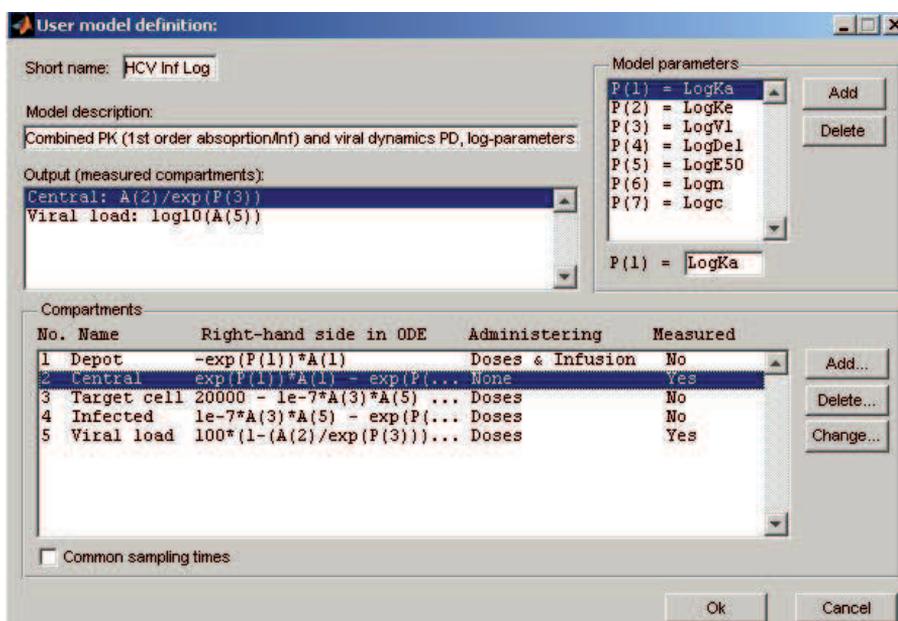


FIGURE 3. User-defined option, input screens: specifying differential equations and measured compartments.

Future work. Among potential extensions of the PkStaMp library is the use of Monte Carlo simulations (29) to approximate the individual information matrix $\mu(\mathbf{x}, \theta)$ in (13). It is also worthwhile to explore various measures of nonlinearity when using different approximation options; see Bates and Watts [5], Merlé and Todd [16], Pázman [21], Ratkowsky [22].

For information regarding availability of the PKStaMp library, please contact the corresponding author.

Acknowledgments. The authors are grateful to the anonymous referee for valuable comments on an earlier version of the paper that helped to improve the presentation of the results. The authors wish to thank participants of PODE workshop for many helpful discussions during the work on the two examples described in this paper, with special thanks to C. Bazzoli, S. Duffull, F. Mentré, J. Nyberg and K. Ogungbenro.

6. Appendix: derivation of formulae (14) and (15)

First consider the normal population distribution (5). (In fact, it is sufficient to assume the existence of the first two moments: $E\gamma_i = \gamma^0$, $\text{Var}(\gamma_i) = \Omega$.) Let

$$\begin{aligned}\mathbf{X}_i &= (x_{i1}, \dots, x_{ik})^T, \\ \mathbf{Y}_i &= (y_{i1}, \dots, y_{ik})^T, \\ \boldsymbol{\eta}(\mathbf{X}_i, \gamma_i) &= [\eta(x_{i1}, \gamma_i), \dots, \eta(x_{ik}, \gamma_i)]^T.\end{aligned}$$

Then the model (1), (2) can be written in the vector form

$$\mathbf{Y}_i = [\mathbf{I}_k + \sigma_P \text{Diag}(\varepsilon_{2i})] \boldsymbol{\eta}(\mathbf{X}_i, \gamma_i) + \sigma_A \varepsilon_{1i}. \quad (32)$$

It follows from the first-order Taylor expansion that

$$\boldsymbol{\eta}(\mathbf{X}_i, \gamma_i) \approx \boldsymbol{\eta}(\mathbf{X}_i, \gamma^0) + \mathbf{F}(\gamma_i - \gamma^0). \quad (33)$$

Since $E(\varepsilon_{1i}) = E(\varepsilon_{2i}) = \mathbf{0}$, and since $\varepsilon_{1i}, \varepsilon_{2i}$ and γ_i are independent, then it follows from (5), (32) and (33) that

$$E_{\gamma, \varepsilon}(\mathbf{Y}_i) \approx \boldsymbol{\eta}(\mathbf{X}_i, \gamma^0), \quad (34)$$

and

$$E_{\gamma, \varepsilon}(\mathbf{Y}_i \mathbf{Y}_i^T) \approx E_{\gamma, \varepsilon}[(\mathbf{U}_1 + \mathbf{U}_2 + \mathbf{U}_3)(\mathbf{U}_1 + \mathbf{U}_2 + \mathbf{U}_3)^T], \quad (35)$$

where

$$\begin{aligned}\mathbf{U}_1 &= [\mathbf{I}_k + \sigma_P \text{Diag}(\varepsilon_{2i})] \boldsymbol{\eta}(\mathbf{X}_i, \gamma^0), \\ \mathbf{U}_2 &= [\mathbf{I}_k + \sigma_P \text{Diag}(\varepsilon_{2i})] \mathbf{F}(\gamma_i - \gamma^0), \\ \mathbf{U}_3 &= \sigma_A \varepsilon_{1i}.\end{aligned}$$

The notation $E_{\gamma, \varepsilon}$ means the expectation with respect to the distribution of $\gamma_i, \varepsilon_{1i}$ and ε_{2i} . Note now that

$$E_{\gamma, \varepsilon}(\mathbf{U}_1 \mathbf{U}_1^T) = \boldsymbol{\eta}(\mathbf{X}_i, \gamma^0) \boldsymbol{\eta}^T(\mathbf{X}_i, \gamma^0) + \sigma_P^2 \text{Diag}[\boldsymbol{\eta}(\mathbf{X}_i, \gamma^0) \boldsymbol{\eta}^T(\mathbf{X}_i, \gamma^0)], \quad (36)$$

$$E_{\gamma, \varepsilon}(\mathbf{U}_2 \mathbf{U}_2^T) = \mathbf{F} \Omega \mathbf{F}^T + \sigma_P^2 \text{Diag}(\mathbf{F} \Omega \mathbf{F}^T), \quad (37)$$

$$E_{\gamma, \varepsilon}(\mathbf{U}_3 \mathbf{U}_3^T) = \sigma_A^2 \mathbf{I}_k. \quad (38)$$

OPTIMAL DESIGN FOR POPULATION PK/PD MODELS

Finally, $E_{\gamma, \varepsilon}(\mathbf{U}_s \mathbf{U}_{s'}^T) = \mathbf{0}$ if $s \neq s'$, because of independence of $\varepsilon_{1i}, \varepsilon_{2i}$ and γ_i . The formula (14) follows from (34)–(38).

When γ_i are log-normally distributed as in (6), the only adjustment in the derivation of the covariance matrix \mathbf{S} has to be made in the estimation of the term $E_{\gamma}[(\gamma_i - \gamma^0)(\gamma_i - \gamma^0)^T]$ for $E_{\gamma, \varepsilon}(\mathbf{U}_2 \mathbf{U}_2^T)$ in (37). The first-order approximation (21) entails

$$E(\gamma_i) = \gamma^0, \quad \text{Var}(\gamma_i) = \text{Diag}(\gamma^0) \boldsymbol{\Omega} \text{Diag}(\gamma^0), \quad (39)$$

and the formula (15) now follows from (39).

REFERENCES

- [1] ALIEV, A.—FEDOROV, V.—LEONOV, S.—MCHUGH, B.: *PkStaMp library for optimization of sampling schemes for PK/PD models*, in: Proc. of the 6th St. Petersburg Workshop on Simulation, Vol. 1 (S. M. Ermakov et al., eds.), VVM com Ltd., St. Petersburg, Russia, 2009, pp. 368–374.
- [2] ALIEV, A.—FEDOROV, V.—LEONOV, S.—MCHUGH, B.—MAGEE, M.: *PkStaMp library for constructing optimal population designs for PK/PD studies*, Commun. Stat., Simulation Comput. **41** (2012), 717–729.
- [3] ATKINSON, A. C.—DONEV, A.: *Optimum Experimental Design*. Clarendon Press, Oxford, 1992.
- [4] ATWOOD, C. L.: *Sequences converging to D-optimal designs of experiments*, Ann. Stat. **1** (1973), 342–352.
- [5] BATES, D. M.—WATTS, D. G.: *Nonlinear Regression Analysis and its Applications*. Wiley, New York, 1988.
- [6] BAZZOLI, C.—RETOU, S.—MENTRÉ, F.: *Fisher information matrix for nonlinear mixed effects multiple response models: evaluation of the appropriateness of the first order linearization using a pharmacokinetic/pharmacodynamic model*, Stat. Med. **28** (2009), 1940–1956.
- [7] DUFFULL, S. B.—WATERHOUSE, T. H.—ECCLESTON, J. A.: *Some considerations on the design of population pharmacokinetic studies*, J. Pharmacokin. Pharmacodyn. **32** (2005), 441–457.
- [8] FEDOROV, V. V.: *Theory of Optimal Experiment*. Academic Press, New York, 1972.
- [9] FEDOROV, V. V.—GAGNON, R.—LEONOV, S.—WU, Y.: *Optimal design of experiments in pharmaceutical applications*, in: Pharmaceutical Statistics Using SAS. A Practical Guide (A. Dmitrienko et al., eds.), SAS Press, Cary, NC, 2007, 151–195.
- [10] FEDOROV, V. V.—HACKL, P.: *Model-Oriented Design of Experiments*. Springer, New York, 1997.
- [11] FEDOROV, V. V.—LEONOV, S.: *Response driven designs in drug development*, in: Applied Optimal Designs (W. K. Wong and M. P. F. Berger, eds.), Wiley, Chichester, 2005, pp. 103–136.
- [12] GAGNON, R.—LEONOV, S.: *Optimal population designs for PK models with serial sampling*, J. Biopharm. Statist. **15** (2005), 143–163.
- [13] GUEORGUIEVA, I.—OGUNGBENRO, K.—GRAHAM, G.—GLATT, S.—AARONS, L.: *A program for individual and population optimal design for univariate and multivariate response pharmacokinetic-pharmacodynamic models*, Comput. Methods Programs Biomed. **86** (2007), 51–61.

- [14] MAGNUS, J. R.—NEUDECKER, H.: *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley, New York, 1988.
- [15] MENTRÉ, F.—NYBERG, J.—OGUNGBENRO, K.—LEONOV, S.—ALIEV, A.—DUFFULL, S.—BAZZOLI, C.—HOOKER, A.: *Comparison of results of the different software for design evaluation in population pharmacokinetics and pharmacodynamics*, in: Abstracts of the Annual Meeting of the Population Approach Group in Europe (PAGE), 2011, <http://www.page-meeting.org/default.asp?abstract=2066>.
- [16] MERLÉ, Y.—TODD, M.: *Impact of pharmacokineticpharmacodynamic model linearization on the accuracy of population information matrix and optimal design*, *J. Pharmacokin. Pharmacodyn.* **28** (2001), 363–388.
- [17] MIELKE, T.—SCHWABE, R.: *Some considerations on the Fisher information in nonlinear mixed effects models*, in: mODa 9—Advances in Model-Oriented Design and Analysis (A. Giovagnoli et al., eds.), Physica-Verlag/Springer, Berlin, 2010, pp. 129–136.
- [18] MONOLIX: <http://www.lixoft.net/monolix/overview/>, Lixoft, 2011.
- [19] NEUMANN, A. U.—LAM, N. P.—DAHARI, H.—GRETCH, D. R.—WILEY, T. E.—LAYDEN, T. J.—PERELSON, A. S.: *Hepatitis C viral dynamics in vivo and the antiviral efficacy of interferon- α therapy*, *Science* **282** (1998), 103–107.
- [20] NYBERG, J.—UECKERT, S.—STRÖMBERG, E.—KARLSSON, M. O.—HOOKER, A. C.: *PopED, version 2.12*, 2011, <http://poped.sourceforge.net>.
- [21] PÁZMAN, A.: *Nonlinear Statistical Models*. Kluwer Academic Publishers, Dordrecht, 1993.
- [22] RATKOWSKY, D. A.: *Nonlinear Regression Modeling*. Marcel Dekker, New York, 1983.
- [23] RETOUT, S.—MENTRÉ, F.: *Further developments of the Fisher information matrix in nonlinear mixed effects models with evaluation in population pharmacokinetics*, *J. Biopharm. Statist.* **13** (2003), 209–227.
- [24] WYNN, H. P.: *The sequential generation of D-optimal experimental designs*, *Ann. Math. Statist.* **41** (1970), 1655–1664.

Received January 21, 2012

Sergei Leonov
Vertex Pharmaceuticals
130 Waverly Street
Cambridge, MA 02139
U.S.A.
E-mail: sergei_leonov@vrtx.com

Alexander Aliev
Institute for Systems Analysis
Prospect 60-letia Octiabria, 9
Moscow 117312
RUSSIA
E-mail: ali@isa.ru