Biometrical Letters
Vol. 45 (2008), No. 1, 1–28

Optimum designs for discrimination between two nonlinear multivariate dynamic mixed–effects models

Bartosz Kuczewski[1,2], Barbara Bogacka[1], Dariusz Uciński[2]

[1]Queen Mary, University of London, School of Mathematical Sciences,
Mile End Road, London E1 4NS, UK, e-mail: b.bogacka@qmul.ac.uk
[2]University of Zielona Góra, Institute of Control and Computation Engineering,
ul. Podgórna 50, 65-246 Zielona Góra, Poland

Summary

The paper concerns a problem of finding powerful experimental designs in order to discriminate between two alternative nonlinear multivariate dynamic mixed-effects statistical models. The T-optimality criterion developed for fixed models with heteroscedastic errors is generalized and used after linearization of the candidate models by Taylor series expansion around the mean value of the parameters. The relevant equivalence theorem is proved. A numerical algorithm for finding optimal designs based on a Wynn-type iterative procedure is constructed. T-optimum designs for discrimination between two pharmacokinetic multiresponse models are calculated as an example.

Key words: model selection, T-optimum design, pharmacokinetic models, random parameters

1. Introduction

The model discrimination problem is an important part of statistical inference. There exist two main approaches to model discrimination, one based on hypothesis testing, e.g., likelihood ratio tests or goodness-of-fit tests (Lehmann and Romano, 2005), the second based on information theory and information criteria, such as Akaike and Bayes Information Criteria, respectively (see the monograph by Burnham and Anderson, 2002). Also, a chosen sampling plan (design) may influence the result of model discrimination. Thus, in this context, various criteria for designing experiments were considered in the literature (Box and Hill, 1967; Atkinson and Fedorov, 1975a; Fedorov and Khabarov, 1986; Ponce de Leon and Atkinson, 1991; Felsenstein, 1992; Müller and Ponce de Leon, 1996; Stewart, Shon and Box, 1998).

In this paper we consider the criterion called T-optimality, introduced in Atkinson and Fedorov (1975a) for two competing single response models, and then extended in Atkinson and Fedorov (1975b) for several competing models. For two competing linear models a T-optimum design provides the most powerful F-test for the lack of fit of one model when the other is assumed to be true. When the models are non-linear in the parameters, the exact F-test is replaced by an asymptotic one. In recent years, the criterion has been generalized to the case of nonlinear dynamic models (described by ordinary differential equations or partial differential equations) and also for heteroscedastic errors (see, e.g., Uciński and Bogacka, 2004; Uciński and Bogacka, 2005; Kuczewski, 2006; Kuczewski, Baranowski and Uciński, 2006). However, to the best of our knowledge, only fixed-effects models have been considered so far.

Mixed-effects models, in which some or all parameters are assumed to be random variables, are widely used in areas such as clinical trials, in particular in the so-called population pharmacokinetics and pharmacodynamics (Dawidian and Giltinan, 1995; Demidenko, 2004; Pinheiro and Bates, 2000; Lindsey, 2001). The interest is in a "mean model" and its so-called population parameters for a sample population of subjects (patients), rather than for individual cases.

Methods of parameter and variance component estimation for such models are well developed, see, e.g., Searle, Casella and McCulloch (1992), Pinheiro and Bates (2000), also the article by Yuh, Beal, Davidian, Harrison, Hester, Kowalski, Vonesh and Wolfinger (1994) providing a bibliography survey up to 1994. The design of experiments for precise estimation of the population parameters have been developed for such models. These are mainly D-optimum designs (see, e.g., Mentré, Mallet and Baccar, 1997; Jones and Wang, 1999; Retout, Duffull and Mentré, 2001; Retout and Mentré, 2003; Gagnon and Leonov, 2005; Patan and Bogacka, 2007). There is also some statistical literature on the Bayesian experimental design for such models (Han and Chaloner, 2004).

In contrast to the above topics, the model discrimination problem has attracted much less attention. Waterhouse, Woods, Eccleston and Lewis (2006) introduce product D-optimum population designs and claim that in many cases this strategy is efficient for both parameter estimation and model discrimination purposes. However, especially for non-linear models,

it may happen that the discrimination efficiency for such a design may be rather low when compared with a T-optimum one. In some situations it may be more efficient to first establish a correct model and then continue the experiment designed for parameter estimation.

In this paper we propose a method for computing T-optimum designs in order to discriminate between two competing population multiresponse models. In Section 2 we introduce the relevant notation for such models, present their linearized forms and define a generalized T-criterion complemented with its justification. In Section 3 we give a necessary and sufficient condition for a design to be T-optimum (Theorem 1). In Section 4 we present optimum designs for discriminating between two pharmacokinetic models. We conclude our work and give some discussion in Section 5. Details of the algorithm are given in the appendix.

2. Problem formulation

We consider a nonlinear, multi-response mixed-effects statistical model. The observations $\boldsymbol{y}_{ij} \in \mathbb{R}^d$ of process *responses* are described by

$$\boldsymbol{y}_{ij} = \boldsymbol{\eta}(t_i; \boldsymbol{\lambda}_{ij}) + \boldsymbol{\varepsilon}_{ij}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, r_i, \tag{1}$$

where $\boldsymbol{\eta} : \mathbb{R} \to \mathbb{R}^d$ is a nonlinear function with respect to the parameters $\boldsymbol{\lambda}$, $t_i \in \mathbb{R}$ is the explanatory variable (here a time instant of a measurement), $t_i \neq t_\kappa$ whenever $i \neq \kappa$, n is the number of different time points and r_i is the number of replications at t_i. The total number of observations is $\sum_{i=1}^{n} r_i = N$.

We assume that the model parameter $\boldsymbol{\lambda}_{ij}$ is a realization of a p-dimensional normal random vector such that

$$\boldsymbol{\lambda}_{ij} = \boldsymbol{\theta} + \boldsymbol{\vartheta}_{ij} \tag{2}$$

and

$$\mathrm{E}\big(\boldsymbol{\lambda}_{ij}\big) = \boldsymbol{\theta}, \quad \mathrm{Var}\big(\boldsymbol{\lambda}_{ij}, \boldsymbol{\lambda}_{\kappa\ell}\big) = \begin{cases} \boldsymbol{\Omega}(\boldsymbol{\theta}, \boldsymbol{\alpha}) & \text{if } i = \kappa \text{ and } j = \ell, \\ \mathbf{0}_{p\times p} & \text{otherwise,} \end{cases} \tag{3}$$

where $\boldsymbol{\theta}$ is called the population mean, $\boldsymbol{\Omega}(\boldsymbol{\theta}, \boldsymbol{\alpha}) \in \mathbb{R}^{p\times p}$ is a positive-definite symmetric matrix depending on $\boldsymbol{\theta}$ and, possibly, on an additional vector of constant parameters $\boldsymbol{\alpha} \in \mathbb{R}^q$.

Terms $\boldsymbol{\varepsilon}_{ij}$ represent random errors of measurements. We assume that the errors are normally distributed and for given $\boldsymbol{\lambda}_{ij}$ and $\boldsymbol{\lambda}_{\kappa\ell}$ we have

$$\mathrm{E}\big(\boldsymbol{\varepsilon}_{ij}\big) = \mathbf{0}_d, \quad \mathrm{Var}\big(\boldsymbol{\varepsilon}_{ij}, \boldsymbol{\varepsilon}_{\kappa\ell}\big) = \begin{cases} \boldsymbol{V}(t_i, \boldsymbol{\theta}, \boldsymbol{\beta}) & \text{if } i = \kappa \text{ and } j = \ell, \\ \mathbf{0}_{d\times d} & \text{otherwise,} \end{cases} \tag{4}$$

where $\mathbf{0}_d$ and $\mathbf{0}_{d\times d}$ are, respectively, the d-dimensional vector and the $(d \times d)$-dimensional matrix of zeros, $\boldsymbol{V}(t_i, \boldsymbol{\theta}, \boldsymbol{\beta}) \in \mathbb{R}^{d\times d}$ is a positive-definite matrix, possibly depending on an additional vector of constant parameters $\boldsymbol{\beta} \in \mathbb{R}^r$.

The properties (3) and (4) yield the observations independent in time, but correlated among the responses at a given t_i.

In what follows, we consider the situation in which we are given two competing heteroscedastic random-effects model structures

$$\mathcal{M}_1 : \quad \big(\boldsymbol{\eta}_1(t_i, \boldsymbol{\lambda}_{ij}^{(1)}),\, \boldsymbol{\Omega}_1(\boldsymbol{\theta}_1, \boldsymbol{\alpha}_1),\, \boldsymbol{V}_1(t_i, \boldsymbol{\theta}_1, \boldsymbol{\beta}_1)\big) \tag{5}$$

and

$$\mathcal{M}_2 : \quad \big(\boldsymbol{\eta}_2(t_i, \boldsymbol{\lambda}_{ij}^{(2)}),\, \boldsymbol{\Omega}_2(\boldsymbol{\theta}_2, \boldsymbol{\alpha}_2),\, \boldsymbol{V}_2(t_i, \boldsymbol{\theta}_2, \boldsymbol{\beta}_2)\big) \tag{6}$$

as candidate models for an unknown structure $\mathcal{M} : \big(\boldsymbol{\eta}(t_i, \boldsymbol{\lambda}_{ij}),\, \boldsymbol{\Omega}(\boldsymbol{\theta}, \boldsymbol{\alpha}),\, \boldsymbol{V}(t_i, \boldsymbol{\theta}, \boldsymbol{\beta})\big)$. The parameters of the competing models follow the assumptions of the form (2) - (4).

In order to determine a sampling schedule which would maximize the power of the test for lack of fit of model $\boldsymbol{\eta}_2$ against $\boldsymbol{\eta}_1$ we extend the T-optimality criterion developed for the heteroscedastic setting in Uciński and Bogacka (2004). For this purpose, we apply the first-order expansion of $\boldsymbol{\eta}$ around the random effects' mean (the approximation is reasonably accurate provided that the dispersion of the $\boldsymbol{\lambda}_{ij}$'s is small). Thus, Eqn. (1) can be approximated by (Dawidian and Giltinan, 1995, Ch. 6)

$$\boldsymbol{y}_{ij} \approx \boldsymbol{\eta}(t_i; \boldsymbol{\theta}) + \frac{\partial \boldsymbol{\eta}(t_i; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \boldsymbol{\vartheta}_{ij} + \boldsymbol{\varepsilon}_{ij}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, r_i. \tag{7}$$

Then we have

$$\mathrm{E}(\boldsymbol{y}_{ij}) \approx \boldsymbol{\eta}(t_i; \boldsymbol{\theta}),\ \mathrm{Var}\big(\boldsymbol{y}_{ij}, \boldsymbol{y}_{\kappa\ell}\big) = \begin{cases} \boldsymbol{W}(t_i, \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\beta}) & \text{if } i = \kappa \ \&\ j = \ell, \\ \mathbf{0}_{d\times d} & \text{otherwise,} \end{cases} \tag{8}$$

where

$$\boldsymbol{W}(t_i,\boldsymbol{\theta},\boldsymbol{\alpha},\boldsymbol{\beta}) \approx \frac{\partial\boldsymbol{\eta}(t_i;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\boldsymbol{\Omega}(\boldsymbol{\theta},\boldsymbol{\alpha})\left(\frac{\partial\boldsymbol{\eta}(t_i;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)^{\mathrm{T}} + \boldsymbol{V}(t_i,\boldsymbol{\theta},\boldsymbol{\beta}). \tag{9}$$

Observe that after such approximation we have a mixed-effects marginal model which is nonlinear with respect to $\boldsymbol{\theta}$ but is linear with respect to the random effects $\boldsymbol{\vartheta_{ij}}$ (Demidenko, 2004). Matrix $\boldsymbol{W}$ depends on both the expectation, $\boldsymbol{\theta}$, and the dispersion, $\boldsymbol{\Omega}$, of the random effects $\boldsymbol{\lambda}_{ij}$, as well as on the dispersion $\boldsymbol{V}$ of the additive errors. $\boldsymbol{W}$ is positive definite.

Defining $\boldsymbol{\gamma} = \left[\boldsymbol{\theta}^{\mathrm{T}},\boldsymbol{\alpha}^{\mathrm{T}},\boldsymbol{\beta}^{\mathrm{T}}\right]^{\mathrm{T}}$ as the vector of all population parameters, we can rewrite (7) as

$$\boldsymbol{y}_{ij} \approx \boldsymbol{\eta}(t_i;\boldsymbol{\theta}) + \boldsymbol{\epsilon}_{ij}, \quad i = 1,\ldots,n, \quad j = 1,\ldots,r_i, \tag{10}$$

where

$$\mathrm{E}\big(\boldsymbol{\epsilon}_{ij}\big) = \mathbf{0}_d, \quad \mathrm{Var}\big(\boldsymbol{\epsilon}_{ij},\boldsymbol{\epsilon}_{\kappa\ell}\big) = \begin{cases} \boldsymbol{W}(t_i,\boldsymbol{\gamma}) & \text{if } i=\kappa \text{ and } j=\ell, \\ \mathbf{0}_{d\times d} & \text{otherwise,} \end{cases} \tag{11}$$

where $\boldsymbol{W}(t_i,\boldsymbol{\gamma})$ has the form given by (9).

Consequently, the approximation (10) with properties (11) is a model equivalent to a fixed-effects multi-response nonlinear model with heteroscedastic errors. The competing models $\mathcal{M}_1$ and $\mathcal{M}_2$ can then be written as:

$$\begin{aligned} \mathcal{M}_1: &\quad \mathrm{E}(\boldsymbol{y}_{ij}) \approx \boldsymbol{\eta}_1(t,\boldsymbol{\theta}_1), \quad \mathrm{Var}(\boldsymbol{y}_{ij}) \approx \boldsymbol{W}_1(t_i,\boldsymbol{\gamma}_1), \\ \mathcal{M}_2: &\quad \mathrm{E}(\boldsymbol{y}_{ij}) \approx \boldsymbol{\eta}_2(t,\boldsymbol{\theta}_2), \quad \mathrm{Var}(\boldsymbol{y}_{ij}) \approx \boldsymbol{W}_2(t_i,\boldsymbol{\gamma}_2), \end{aligned} \tag{12}$$

where $\boldsymbol{\gamma}_\ell = \left[\boldsymbol{\theta}_\ell^{\mathrm{T}},\boldsymbol{\alpha}_\ell^{\mathrm{T}},\boldsymbol{\beta}_\ell^{\mathrm{T}}\right]^{\mathrm{T}} \in \Gamma_\ell \subset \mathbb{R}^{p_\ell+q_\ell+r_\ell}$, $\ell = 1,2$, denote vectors of constant but unknown parameters (Γ_1 and Γ_2 denote some known compact sets). Matrices $\boldsymbol{W}_\ell$, $\ell = 1,2$, are defined as in (9).

If we postulate that model $\mathcal{M}_1$ is true, i.e., $\boldsymbol{W}(t,\boldsymbol{\gamma}) = \boldsymbol{W}_1(t,\boldsymbol{\gamma}_1^0)$, $\boldsymbol{\eta}(t,\boldsymbol{\theta}) = \boldsymbol{\eta}_1(t,\boldsymbol{\theta}^0)$ for some known $\boldsymbol{\gamma}_1^0$ (which may have been inferred from some prior experiments) the discrimination between models given by (12) can be performed based on a design $\xi_N^\star$ maximizing the following generalization of the T-optimality criterion over a set of designs ξ_N:

$$T_{12}^0(\xi_N) = \min_{\boldsymbol{\gamma}_2\in\Gamma_2} J_{12}^0(\xi_N,\boldsymbol{\gamma}_2), \tag{13}$$

where

$$J_{12}^0(\xi_N, \boldsymbol{\gamma}_2) = \sum_{i=1}^{n} w_i g(t_i, \boldsymbol{\gamma}_2), \tag{14}$$

$$\begin{aligned} g(t, \boldsymbol{\gamma}_2) =& \Phi\big(\boldsymbol{W}_2^{-1}(t, \boldsymbol{\gamma}_2)\boldsymbol{W}(t)\big) \\ &+ [\boldsymbol{\eta}(t) - \boldsymbol{\eta}_2(t, \boldsymbol{\theta}_2)]^{\mathrm{T}} \boldsymbol{W}_2^{-1}(t, \boldsymbol{\gamma}_2)[\boldsymbol{\eta}(t) - \boldsymbol{\eta}_2(t, \boldsymbol{\theta}_2)] \end{aligned} \tag{15}$$

and

$$\Phi(\boldsymbol{A}) = \operatorname{trace}(\boldsymbol{A}) - \ln\det(\boldsymbol{A}). \tag{16}$$

For simplicity of notation we omitted the dependencies of $\boldsymbol{\eta}$ and $\boldsymbol{W}$ on $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$, respectively, since these two parameters are assumed to be known and fixed for the true model.

An important property of the function $\Phi(\,\cdot\,)$ is that it is strictly convex on the set of all symmetric positive-definite $(d \times d)$-dimensional matrices, cf. (Uciński and Bogacka, 2004, Lemma 1).

The normalized N-observation exact design ξ_N has the following standard form:

$$\xi_N \stackrel{\text{def}}{=} \begin{Bmatrix} t_1, & t_2, & \dots, & t_n \\ w_1, & w_2, & \dots, & w_n \end{Bmatrix}, \tag{17}$$

where $w_i = r_i/N$ and $\sum\limits_{i=1}^{n} w_i = 1$. We denote the set of support points $t_1, \dots, t_n$ by $\operatorname{supp}\xi_N$.

A rationale for the criterion (13) is the following. Assume that Model $\mathcal{M}_1$ is true for a given parameter vector $\boldsymbol{\gamma}_1^0$, and that $\boldsymbol{\gamma}_2^0$ is a possible value of parameter $\boldsymbol{\gamma}_2$ in $\mathcal{M}_2$. Moreover, for simplicity, assume that we have N distinct support points, all associated with equal weights $w_i = 1/N$, $i = 1, \dots, N$. Discrimination between the two competing models (12) can be viewed as testing the following simple hypothesis:

$$\mathcal{H}_0: \quad \boldsymbol{\eta}(t_i) = \boldsymbol{\eta}_1(t_i, \boldsymbol{\theta}_1^0), \quad \boldsymbol{W}(t_i) = \boldsymbol{W}_1(t_i, \boldsymbol{\gamma}_1^0), \quad i = 1, \dots, N, \tag{18}$$

against the alternative

$$\mathcal{H}_1: \quad \boldsymbol{\eta}(t_i) = \boldsymbol{\eta}_2(t_i, \boldsymbol{\theta}_2^0), \quad \boldsymbol{W}(t_i) = \boldsymbol{W}_2(t_i, \boldsymbol{\gamma}_2^0), \quad i = 1, \dots, N. \tag{19}$$

Since the disturbances $\boldsymbol{\epsilon}$ are approximately Gaussian and independent for different t_i's, the log-likelihood ratio function is given by

$$\mathcal{L} = \ln\left(\frac{\prod_{i=1}^{N}(2\pi)^{-\frac{d}{2}}|\boldsymbol{W}_1(t_i)|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}[\boldsymbol{y}_i-\boldsymbol{\eta}_1(t_i)]^{\mathrm{T}}\boldsymbol{W}_1^{-1}(t_i)[\boldsymbol{y}_i-\boldsymbol{\eta}_1(t_i)]\right)}{\prod_{i=1}^{N}(2\pi)^{-\frac{d}{2}}|\boldsymbol{W}_2(t_i)|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}[\boldsymbol{y}_i-\boldsymbol{\eta}_2(t_i)]^{\mathrm{T}}\boldsymbol{W}_2^{-1}(t_i)[\boldsymbol{y}_i-\boldsymbol{\eta}_2(t_i)]\right)}\right) \tag{20}$$

$$= -\frac{1}{2}\sum_{i=1}^{N}\ln|\boldsymbol{W}_2^{-1}(t_i)\boldsymbol{W}_1(t_i)| - \frac{1}{2}\sum_{i=1}^{N}[\boldsymbol{y}_i-\boldsymbol{\eta}_1(t_i)]^{\mathrm{T}}\boldsymbol{W}_1^{-1}(t_i)[\boldsymbol{y}_i-\boldsymbol{\eta}_1(t_i)]$$
$$+\frac{1}{2}\sum_{i=1}^{N}[\boldsymbol{y}_i-\boldsymbol{\eta}_2(t_i)]^{\mathrm{T}}\boldsymbol{W}_2^{-1}(t_i)[\boldsymbol{y}_i-\boldsymbol{\eta}_2(t_i)],$$

where the abbreviated notation $\boldsymbol{\eta}_\ell(t_i) \equiv \boldsymbol{\eta}_\ell(t_i, \boldsymbol{\theta}_\ell^0)$ and $\boldsymbol{W}_\ell(t_i) \equiv \boldsymbol{W}_\ell(t_i, \gamma_\ell^0)$, $\ell = 1, 2$, is applied.

Then

$$\begin{aligned} 2\,\mathrm{E}[\mathcal{L}] = &\underbrace{-\,\mathrm{E}\left[\sum_{i=1}^{N}\ln|\boldsymbol{W}_2^{-1}(t_i)\boldsymbol{W}_1(t_i)|\right]}_{\text{Term I}} \\ &\underbrace{-\,\mathrm{E}\left[\sum_{i=1}^{N}[\boldsymbol{y}_i-\boldsymbol{\eta}_1(t_i)]^{\mathrm{T}}\boldsymbol{W}_1^{-1}(t_i)[\boldsymbol{y}_i-\boldsymbol{\eta}_1(t_i)]\right]}_{\text{Term II}} \\ &\underbrace{+\,\mathrm{E}\left[\sum_{i=1}^{N}[\boldsymbol{y}_i-\boldsymbol{\eta}_2(t_i)]^{\mathrm{T}}\boldsymbol{W}_2^{-1}(t_i)[\boldsymbol{y}_i-\boldsymbol{\eta}_2(t_i)]\right]}_{\text{Term III}}. \end{aligned} \tag{21}$$

The explanation of the form of function g in the criterion (13) comes from the three terms indicated in the above formula, as follows:

1. Term I is the expectation of a constant, hence it is equal to

$$\sum_{i=1}^{N}\ln|\boldsymbol{W}_2^{-1}(t_i)\boldsymbol{W}_1(t_i)|.$$

2. Under the null hypothesis we have that $\boldsymbol{y}_i - \boldsymbol{\eta}_1(t_i) \underset{H_0}{\sim} \mathcal{N}(0, \boldsymbol{W}_1(t_i))$. Then, applying the formula for the expectation of a quadratic form, which says that for any vector random variable $\boldsymbol{z}$ such that $\mathrm{E}[\mathrm{z}] = \boldsymbol{\mu}$, $\mathrm{Var}(\boldsymbol{z}) = \boldsymbol{V}$ and for any symmetric constant matrix $\boldsymbol{A}$ of appropriate dimension, the expected value of a quadratic form $\boldsymbol{z}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{z}$ can be written as, (Searle, 1971, Ch. 2):

$$\mathrm{E}(\boldsymbol{z}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{z}) = \mathrm{trace}(\boldsymbol{A}\boldsymbol{V}) + \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{\mu}, \tag{22}$$

it follows that Term II reduces to

$$\sum_{i=1}^{N} \mathrm{trace}\big(\boldsymbol{W}_1^{-1}(t_i)\boldsymbol{W}_1(t_i)\big) = Nd.$$

3. Similarly as above, here we have $\boldsymbol{y}_i - \boldsymbol{\eta}_2(t_i) \underset{H_0}{\sim} \mathcal{N}(\boldsymbol{\eta}_1(t_i) - \boldsymbol{\eta}_2(t_i), \boldsymbol{W}_1(t_i))$ and from (22) it follows that Term III simplifies to

$$\sum_{i=1}^{N} \mathrm{trace}\big(\boldsymbol{W}_2^{-1}(t_i)\boldsymbol{W}_1(t_i)\big) + \sum_{i=1}^{N} [\boldsymbol{\eta}_1(t_i) - \boldsymbol{\eta}_2(t_i)]^{\mathrm{T}} \boldsymbol{W}_2^{-1}(t_i)[\boldsymbol{\eta}_1(t_i) - \boldsymbol{\eta}_2(t_i)].$$

Summarizing and dropping index of the true model we get

$$\begin{aligned} 2\,\mathrm{E}[\mathcal{L}] &= \sum_{i=1}^{N} \mathrm{trace}\big(\boldsymbol{W}_2^{-1}(t_i)\boldsymbol{W}(t_i)\big) - \sum_{i=1}^{N} \ln |\boldsymbol{W}_2^{-1}(t_i)\boldsymbol{W}(t_i)| \\ &\quad + \sum_{i=1}^{N} [\boldsymbol{\eta}(t_i) - \boldsymbol{\eta}_2(t_i)]^{\mathrm{T}} \boldsymbol{W}_2^{-1}(t_i)[\boldsymbol{\eta}(t_i) - \boldsymbol{\eta}_2(t_i)] - Nd \\ &= N\big(J_{12}^0(\xi_N, \boldsymbol{\gamma}_2^0) - d\big), \end{aligned}$$

where ξ_N is a uniform discrete N-point design. Thus, given N, the expected value of the likelihood ratio, which quantifies the discrepancy between both models, is proportional to $J_{12}^0(\xi_N, \boldsymbol{\gamma}_2^0)$. The explanation follows the same arguments if the design ξ_N is not uniform.

Further we relax the original optimization problem (13) and consider an experimental design ξ as a continuous probability measure. It gives the following continuous generalization of the optimality criterion:

$$T_{12}(\xi) = \min_{\boldsymbol{\gamma}_2 \in \Gamma_2} J_{12}(\xi, \boldsymbol{\gamma}_2) \tag{23}$$

for

$$J_{12}(\xi, \boldsymbol{\gamma}_2) = \int_T g(t, \boldsymbol{\gamma}_2)\, \xi(\mathrm{d}t), \tag{24}$$

where $T = [t_{\min}, t_{\max}]$ is a time interval within which observation instants may be scheduled.

A probability measure

$$\xi^\star = \arg \max_{\xi \in \Xi(T)} T_{12}(\xi) \tag{25}$$

is called the locally T_{12}-optimum design in the set of all probability measures ξ over T denoted by $\Xi(T)$.

3. Characterization of optimum solutions

In this section we present theoretical results for which we assume the following:

(A1) T and Γ_2 are compact sets.

(A2) $\boldsymbol{\eta}(\,\cdot\,)$ and $\boldsymbol{W}(\,\cdot\,)$ are continuous functions on T.

(A3) $\boldsymbol{\eta}_2(\,\cdot\,,\,\cdot\,)$ and $\boldsymbol{W}_2(\,\cdot\,,\,\cdot\,)$ are continuous functions on $T \times \Gamma_2$.

Then T_{12}-optimum designs fulfil the equivalent condition given in the following:

Theorem 1 (Equivalence Theorem for T_{12}-optimum designs). *Assume that the minimization problem* (23) *possesses a unique solution* $\boldsymbol{\gamma}_2^\star \in \Gamma_2$ *for a measure* $\xi^\star$. *Under Assumptions (A1)–(A3) a necessary and sufficient condition for* $\xi^\star \in \Xi(T)$ *to be* T_{12}*-optimum is*

$$g(t, \boldsymbol{\gamma}_2^\star) \le T_{12}(\xi^\star), \quad \forall t \in T. \tag{26}$$

The equality in (26) *is attained at all support points of* $\xi^\star$. *Furthermore, the set of all optimal measures* $\xi^\star$ *is convex.*

Proof. First we examine the one-sided directional derivative of T_{12} defined by (23). The continuity of $\boldsymbol{\eta}$ and $\boldsymbol{\eta}_2$, together with the Bounded Convergence Theorem (Rao, 1987, Cor. 6, p. 161), yield the continuity of the mappings

$$(\alpha, \boldsymbol{\gamma}_2) \mapsto J_{12}(\xi + \alpha\delta\xi, \boldsymbol{\gamma}_2) \tag{27}$$

and

$$(\alpha, \boldsymbol{\gamma}_2) \mapsto \frac{\partial J_{12}}{\partial \alpha}(\xi + \alpha\delta\xi, \boldsymbol{\gamma}_2). \tag{28}$$

The directional differentiability property of a max (or equivalently min) function of the form $f(x) = \max_{y \in Y} \varphi(x, y)$ given, e.g., by Theorem 3.3 of (Pshenichnyi, 1971) or Theorem 5.4.7 of (Polak, 1997), states that

$$\delta f(\boldsymbol{x}_0; \delta\boldsymbol{x}) = \max_{\boldsymbol{y} \in \bar{Y}(\boldsymbol{x}_0)} \delta\varphi(\boldsymbol{x}_0, \boldsymbol{y}; \delta\boldsymbol{x}), \tag{29}$$

where $\bar{Y}(\boldsymbol{x}_0) = \{\boldsymbol{y} \in Y : f(\boldsymbol{x}) = \varphi(\boldsymbol{x}, \boldsymbol{y})\}$ (it is sometimes called the *answering set*). It means that the one-sided directional derivative $\delta f(\boldsymbol{x}_0; \delta\boldsymbol{x})$ is equal to the largest (or equivalently smallest for min functions) of the directional derivatives of the functions $x \mapsto \varphi(\boldsymbol{x}, \boldsymbol{y})$ that are 'active' at $\boldsymbol{x}_0$, i.e. for which there exists $\boldsymbol{y}_0 \in Y$ such that $\varphi(\boldsymbol{x}_0, \boldsymbol{y}_0) = f(\boldsymbol{x}_0)$.

This result applied to the T_{12} function implies that

$$\delta T_{12}(\xi; \delta\xi) = \min_{\boldsymbol{\gamma}_2 \in \Gamma_2(\xi)} \delta J_{12}(\xi, \boldsymbol{\gamma}_2; \delta\xi), \tag{30}$$

where

$$\Gamma_2(\xi) = \left\{ \bar{\boldsymbol{\gamma}}_2 \in \Gamma_2 \,:\, \bar{\boldsymbol{\gamma}}_2 = \arg\min_{\boldsymbol{\gamma}_2 \in \Gamma_2} J_{12}(\xi, \boldsymbol{\gamma}_2) \right\}, \tag{31}$$

and $\delta J_{12}(\xi, \boldsymbol{\gamma}_2; \delta\xi)$ stands for the one-sided differential of J_{12} at ξ with increment $\delta\xi$, $\boldsymbol{\gamma}_2$ is interpreted as a fixed parameter.

By the assumption of the theorem, for an optimal design $\xi^\star$, i.e. the one which maximizes $T_{12}(\xi)$, the set $\Gamma_2(\xi^\star)$ consists of only one point $\boldsymbol{\gamma}_2^\star$, and therefore

$$\delta T_{12}(\xi^\star; \delta\xi) = \delta J_{12}(\xi, \boldsymbol{\gamma}_2^\star; \delta\xi) = \int_T g(t, \boldsymbol{\gamma}_2^\star)\, \delta\xi(\mathrm{d}t). \tag{32}$$

Putting $\delta\xi = \xi - \xi^\star$, we get

$$\delta T_{12}(\xi^\star; \xi - \xi^\star) = \int_T g(t, \boldsymbol{\gamma}_2^\star)\, \xi(\mathrm{d}t) - \underbrace{\int_T g(t, \boldsymbol{\gamma}_2^\star)\, \xi^\star(\mathrm{d}t)}_{T_{12}(\xi^\star)}. \tag{33}$$

Finally, since $\int_T T_{12}(\xi^\star)\, \xi(\mathrm{d}t) = T_{12}(\xi^\star)$, we can write

$$\delta T_{12}(\xi^\star; \xi - \xi^\star) = \int_T \varphi(t, \xi^\star)\, \xi(\mathrm{d}t), \tag{34}$$

where

$$\varphi(t, \xi^\star) = g(t, \boldsymbol{\gamma}_2^\star) - \underbrace{\int_T g(t, \boldsymbol{\gamma}_2^\star)\, \xi^\star(\mathrm{d}t)}_{T_{12}(\xi^\star)}. \tag{35}$$

It follows that

$$\max_{\xi \in \Xi(T)} \delta T_{12}(\xi^\star; \xi - \xi^\star) = \max_{\xi \in \Xi(T)} \int_T \varphi(t, \xi^\star)\, \xi(\mathrm{d}t). \tag{36}$$

Note, that the criterion T_{12} is concave due to the linearity of J_{12} in ξ. Hence the optimality of the design $\xi^\star$ implies that (36) must be nonpositive for any $\xi \in \Xi(T)$ (Uciński, 2005, Thm. B. 25, p. 266). Note, however, that we have

$$\int_T \varphi(t, \xi^\star)\, \xi^\star(\mathrm{d}t) = 0 \tag{37}$$

what forces the nonnegativity of the maximum on the right-hand side of (36). From this we see that the optimality of $\xi^\star$ is equivalent to

$$\max_{\xi \in \Xi(T)} \int_T \varphi(t, \xi^\star)\, \xi(\mathrm{d}t) = 0. \tag{38}$$

It is easy to check that the last condition is satisfied if, and only if,

$$\max_{\xi \in \Xi(T)} \varphi(t, \xi^\star) = 0, \tag{39}$$

i.e. in (38) it suffices to restrict attention to one-point measures $\xi = \{\begin{smallmatrix} t \\ 1 \end{smallmatrix}\}$. This gives (26).

Condition (38) is sufficient, too. In fact, for a fixed $\boldsymbol{\gamma}_2 \in \Gamma_2$, J_{12} in (24) is a linear function of ξ, and hence T_{12} becomes concave (Uciński, 2005, Thm. B. 21, p. 265), which means that the necessary condition (38) becomes sufficient as well.

We also claim that the mapping $t \mapsto \varphi(t, \xi^\star)$ attains its maximum value of zero at all the support points of $\xi^\star$. Indeed, suppose that this were false. Then, we could find a set $T' \subset \operatorname{supp} \xi^\star$ and a scalar a such that

$$\int_{T'} \varphi(t, \xi^\star)\, \xi^\star(\mathrm{d}t) \leq a < 0 \tag{40}$$

and

$$\varphi(t, \xi^\star) = 0 \quad \text{for } t \in \operatorname{supp} \xi^\star \setminus T'. \tag{41}$$

But this would yield

$$\int_{T} \varphi(t, \xi^\star)\, \xi^\star(\mathrm{d}t) \leq a < 0 \tag{42}$$

which contradicts (37).

It remains to show that the set of all optimal measures $\xi^\star$ is convex. But this is immediate, since mapping $\xi \mapsto T_{12}(\xi)$ is concave. This completes the proof. □

This theorem is very useful for several reasons. First of all, it provides a tool for checking whether a numerically computed optimum design is indeed optimum; secondly, it helps a numerical procedure to find an optimal solution; thirdly, it can also be used for finding intervals around the optimum support points which produce a high efficiency of the design. Such intervals, the so-called efficient windows, give a practitioner some flexibility in selecting the times of taking measurements, while keeping high design efficiency, see Bogacka, Johnson, Jones and Volkov (2008).

4. Computational example

As an example, we consider two competing compartmental models describing pharmacokinetics of intraconazole drug in patients with Cystic

Fibrosis (for details of the development of statistical models in medical applications see, e.g., Lindsey, 2001). This example was also analyzed in Waterhouse, Redmann, Duffull and Eccleston (2005), where the authors use a D-optimum product criterion in order to calculate a population design.

The expected response of the first model (assumed as the 'true' one in our experiment) is a solution of the following non-linear ODE set:

$$\begin{aligned}
\frac{\mathrm{d}A_1}{\mathrm{d}t} &= -F_{12}k_aA_1 - F_{14}k_aA_1,\\
\frac{\mathrm{d}A_2}{\mathrm{d}t} &= F_{12}k_aA_1 + \frac{Q}{V_3}A_3 - \frac{Q}{V_2}A_2 - F_{24}\frac{CL_{24}}{V_2}A_2 - F_{20}\frac{V_{max}}{K_m + A_2/V_2}A_2,\\
\frac{\mathrm{d}A_3}{\mathrm{d}t} &= \frac{Q}{V_2}A_2 - \frac{Q}{V_3}A_3,\\
\frac{\mathrm{d}A_4}{\mathrm{d}t} &= F_{14}k_aA_1 + F_{24}\frac{CL_{24}}{V_2}A_2 - \frac{CL_4}{V_4}A_4,
\end{aligned} \tag{43}$$

whereas the alternative model is obtained from the following set of linear ODE:

$$\begin{aligned}
\frac{\mathrm{d}A_1}{\mathrm{d}t} &= -F_{12}k_aA_1 - F_{14}k_aA_1,\\
\frac{\mathrm{d}A_2}{\mathrm{d}t} &= F_{12}k_aA_1 + \frac{Q}{V_3}A_3 - \frac{Q}{V_2}A_2 - F_{24}\frac{CL_{24}}{V_2}A_2 - F_{20}\frac{CL_2}{V_2}A_2,\\
\frac{\mathrm{d}A_3}{\mathrm{d}t} &= \frac{Q}{V_2}A_2 - \frac{Q}{V_3}A_3,\\
\frac{\mathrm{d}A_4}{\mathrm{d}t} &= F_{14}k_aA_1 + F_{24}\frac{CL_{24}}{V_2}A_2 - \frac{CL_4}{V_4}A_4,
\end{aligned} \tag{44}$$

where A_1, A_2, A_3 denote the amount of intraconazole in the gut, central and peripheral compartments, respectively, and A_4 stands for the amount of intraconazole metabolite (hydroxyintraconazole). Table 1 contains a description of the model parameters, whereas Fig. 1 presents the two possible kinetics of the drug, see Waterhouse et al. (2005). Note that the solutions A_1–A_4 of these sets of differential equations are nonlinear with respect to the parameters.

Table 1. Interpretation of model parameters and constant values.

	Unit	Interpretation
k_a	[h^{-1}]	absorbtion rate
F_{12}	-	fraction of parent in the gut absolute bioavailability
F_{14}	-	fraction of metabolite after first pass metabolism
F_{24}	-	fraction of parent converted to metabolite
F_{20}	-	fraction of parent eliminated
CL_2	[L/h]	clearance of itraconazole
CL_4	[L/h]	clearance of metabolite
CL_{24}	[L/h]	clearance of itraconazole by metabolism to hydroxyitraconazole
Q	[L/h]	inter-compartmental clearance
V_{max}	[mg/(mL·h)]	theoretical maximum rate of the process
K_m	[mg/mL]	Michaelis-Menten constant
V_2	[L]	volume of central compartment
V_4	[L]	volume of metabolite compartment
V_3	[L]	volume in peripheral compartment

Experimental conditions used for calculation of $\xi^{\star}_{mixed}$ were set as follows:

- For both models the initial conditions are $A_1(0) = 200mg$ (dose), $A_2(0) = A_3(0) = A_4(0) = 0$.
- In both models some of the parameters are treated as known constants. These are $F_{12} = 0.55$, $F_4 = 0.043$, $F_{24} = F_{20} = 0.5$, $V_3 = 307$, $CL_{24} = 1.07$, $Q = 46$, $V_{max} = 0.00954$ and $K_m = 0.329$.
- For both models we assume that $\boldsymbol{\Omega}_\ell = \omega_\ell \mathrm{diag}\{\theta_{1_\ell}, \ldots, \theta_{p_\ell}\}$ and $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2_{const}\boldsymbol{I}_d + \sigma^2_\ell \boldsymbol{V}_\ell(t, \boldsymbol{\theta}_\ell))$, $\sigma^2_{const} = 0.005$, $\boldsymbol{V}_\ell(t, \boldsymbol{\theta}_\ell) = \mathrm{diag}\{A_1(t, \boldsymbol{\theta}_\ell), A_2(t, \boldsymbol{\theta}_\ell), A_3(t, \boldsymbol{\theta}_\ell), A_4(t, \boldsymbol{\theta}_\ell)\}$, $\ell = 1, 2$.
- The 'true' model population parameter vector is $\boldsymbol{\gamma}_1^0 = [k_{a_1}, V_{2_1}, V_{4_1}, CL_{4_1}, \omega_1, \sigma_1^2]^{\mathrm{T}} = [0.945, 365.0, 23.0, 1.75, 0.05, 0.05]^{\mathrm{T}}$.
- For the alternative model the population parameter vector is $\boldsymbol{\gamma}_2 = [k_{a_2}, V_{2_2}, V_{4_2}, CL_{4_2}, CL_{2_2}, \omega_2, \sigma_2^2]^{\mathrm{T}} \in \Gamma_2$, where

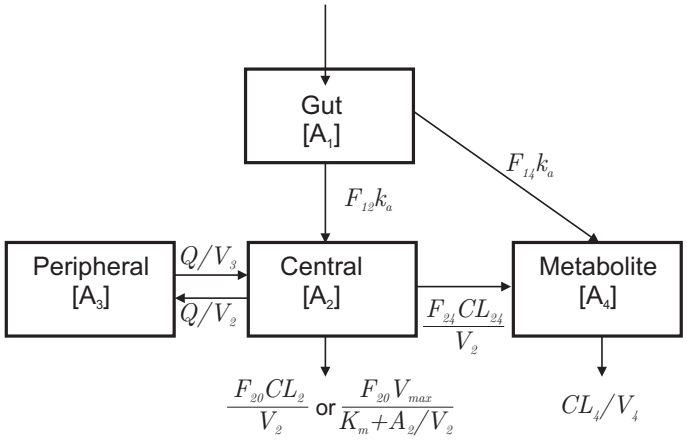

Figure 1. Scheme of the compartmental model of the pharmacokinetics of itraconazole and hydroxyitraconazole.

$\Gamma_2 = [0.9, 1.1] \times [360.0, 370.0] \times [20.0, 25.0] \times [1.6, 2.0] \times [1.0, 7.0] \times [0.025, 0.075] \times [0.025, 0.075]$.

– The design range is $T = [0, 48]$.

For comparison, the T-optimum design $\xi^{\star}_{fixed}$ (for the same models, parameter values and ranges) under the assumption of no random effects and $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}_d, \sigma^2 \boldsymbol{I}_d)$ was also calculated (with $\sigma^2 = 1$). Under this assumption, obviously, $\boldsymbol{\gamma}_2 \equiv \boldsymbol{\theta}_2$ and the term (15) simplifies to $g(t_i, \boldsymbol{\theta}_2) = \|\boldsymbol{\eta}(t_i) - \boldsymbol{\eta}_2(t_1, \boldsymbol{\theta}_2)\|^2$ (see, e.g., Uciński and Bogacka, 2005; Kuczewski et al., 2006).

Additionally, a series of mixed-effects designs $\xi^{j\star}_{mixed},\ j = 1, \ldots, 5$, were computed in order to investigate the influence of variance terms on the resultant design (for the same values and ranges of the fixed-effects parameters $\boldsymbol{\theta_1}$ and $\boldsymbol{\theta_2}$, but different combinations of values and ranges of the parameters $\omega_i, \sigma_i^2,\ i = 1, 2$). For this designs, the experimental conditions were set as follows:

– ξ^1_{mixed}: $\omega_1 = \sigma_1^2 = 0.5$, $[\omega_2, \sigma_2^2] \in [0.25, 0.75] \times [0.25, 0.75]$,

– ξ^2_{mixed}: $\omega_1 = \sigma_1^2 = 5.0$, $[\omega_2, \sigma_2^2] \in [2.5, 7.5] \times [2.5, 7.5]$,

– ξ^3_{mixed}: $\omega_1 = \sigma_1^2 = 50.0$, $[\omega_2, \sigma_2^2] \in [25.0, 75.0] \times [25.0, 75.0]$,
– ξ^4_{mixed}: $\omega_1 = 5.0, \sigma_1^2 = 0.5$, $[\omega_2, \sigma_2^2] \in [2.0, 5.0] \times [0.25, 0.75]$,
– ξ^5_{mixed}: $\omega_1 = 50.0, \sigma_1^2 = 0.5$, $[\omega_2, \sigma_2^2] \in [20.0, 50.0] \times [0.25, 0.75]$.

Also the T-efficiencies of the fixed-effects design in reference to each particular mixed-effects design were calculated as follows:

$$T_{\text{eff}}\left(\xi^\star_{fixed}, \xi^\star_{mixed}\right) = \frac{J_{12}\left(\xi^\star_{fixed}, \boldsymbol{\gamma}^\star_{2_{mixed}}\right)}{J_{12}\left(\xi^\star_{mixed}, \boldsymbol{\gamma}^\star_{2_{mixed}}\right)}. \tag{45}$$

Actually, it is a generalized counterpart of T^b_{eff} from (Waterhouse, Eccleston and Duffull, 2004).

Computer programs were implemented in Lahey-Fujitsu Fortran 95 compiler v.5.6, additionally using some of the routines from the IMSL library. The direct differentiation method (Uciński, 2005) was used in order to calculate sensitivity coefficients $\partial\boldsymbol{\eta}(t_i;\boldsymbol{\theta})/\partial\boldsymbol{\theta}$. The computation of the designs (with tolerance δ varying from 0.001 up to 0.005) took from 3 to 4 hours, depending on the considered case (fixed or mixed effect), on a 1.7 GHz PC Pentium 4 with 512GB of RAM. This is due to an unavoidable cost paid for the necessity of using global optimization. For a discussion of the algorithm see the appendix and comments in Section 5.

The designs obtained for mixed- and fixed-effects models, respectively, are the following:

$$\xi^\star_{mixed} = \left\{ \begin{array}{cc} 11.0030, & 48.0 \\ 0.0589, & 0.9411 \end{array} \right\}, \qquad \xi^\star_{fixed} = \left\{ \begin{array}{cc} 13.1171, & 48.0 \\ 0.7106, & 0.2894 \end{array} \right\}$$

The most competitive values of the alternative model parameters obtained in the optimization procedure are

$$\boldsymbol{\gamma}^\star_{2_{mixed}} = [\underbrace{0.9337, 364.1804, 22.2013, 1.6660, 6.0070}_{\boldsymbol{\theta}^\star_{2_{mixed}}}, 0.0360, 0.0447],$$

$$\boldsymbol{\theta}^\star_{2_{fixed}} = [0.9441, 365.3012, 22.2884, 1.6950, 5.9152].$$

Moreover, T-efficiency $T_{\text{eff}}\left(\xi^\star_{fixed}, \xi^\star_{mixed}\right) = 0.9975$.

Regarding the additional mixed-effects designs, the results are the following:

$$\xi^{1\star}_{mixed} = \left\{ \begin{array}{ll} 0.3363, & 48.0 \\ 0.1136, & 0.8864 \end{array} \right\}, \qquad \xi^{2\star}_{mixed} = \left\{ \begin{array}{ll} 0.4805, & 48.0 \\ 0.2135, & 0.7865 \end{array} \right\},$$
$$\xi^{3\star}_{mixed} = \left\{ \begin{array}{lll} 0.3363, & 5.9670, & 48.0 \\ 0.1573, & 0.0449, & 0.7978 \end{array} \right\}, \quad \xi^{4\star}_{mixed} = \left\{ \begin{array}{ll} 7.5435, & 48.0 \\ 0.1251, & 0.8749 \end{array} \right\},$$
$$\xi^{5\star}_{mixed} = \left\{ \begin{array}{l} 48.0 \\ 1.0 \end{array} \right\},$$

$$\boldsymbol{\gamma}^{1\star}_{2_{mixed}} = [0.9249, 364.2665, 22.2073, 1.6001, 5.9750, \mathbf{0.3276}, \mathbf{0.4540}],$$
$$\boldsymbol{\gamma}^{2\star}_{2_{mixed}} = [0.9334, 364.9910, 23.0203, 1.6000, 5.7314, \mathbf{3.3092}, \mathbf{4.5486}],$$
$$\boldsymbol{\gamma}^{3\star}_{2_{mixed}} = [0.9399, 360.0025, 23.7860, 1.7022, 2.5162, \mathbf{37.4291}, \mathbf{42.1687}],$$
$$\boldsymbol{\gamma}^{4\star}_{2_{mixed}} = [0.9159, 366.0529, 24.9959, 1.6004, 5.5250, \mathbf{2.1115}, \mathbf{0.5210}],$$
$$\boldsymbol{\gamma}^{5\star}_{2_{mixed}} = [0.9276, 364.3781, 24.2408, 1.6002, 1.2532, \mathbf{25.9694}, \mathbf{0.5998}],$$

$$T_{\text{eff}}\left(\xi^{\star}_{fixed}, \xi^{1\star}_{mixed}\right) = 0.9830, \qquad T_{\text{eff}}\left(\xi^{\star}_{fixed}, \xi^{2\star}_{mixed}\right) = 0.9808,$$
$$T_{\text{eff}}\left(\xi^{\star}_{fixed}, \xi^{3\star}_{mixed}\right) = 0.9850, \qquad T_{\text{eff}}\left(\xi^{\star}_{fixed}, \xi^{4\star}_{mixed}\right) = 0.9230,$$
$$T_{\text{eff}}\left(\xi^{\star}_{fixed}, \xi^{5\star}_{mixed}\right) = 0.8710.$$

We can see that the most competitive values of parameters $\boldsymbol{\theta}$ for the alternative model are similar in both mixed- and fixed-effects models (an exception occurs for $\xi^{3\star}_{mixed}$ and $\xi^{5\star}_{mixed}$ in the case of parameter CL_{2_2}). Also, comparing $\xi^{\star}_{fixed}$ and $\xi^{\star}_{mixed}$ (computed for relatively small values of variance terms), we observe that the support points for the two cases are similar. The weights, however, are very different, putting almost all experimental effort at the end of the region in the mixed-effects case.

It is a hallmark of the T-optimum criterion that it puts the support points where there is a large distance between the competing models. Here the main difference between the two kinetics is in the clearance of intraconazole, which has more effect on the expected model responses at the end of the design region than at the beginning. When the clearance parameter Cl_2 is treated as a random variable, it introduces additional variability, which is higher at the end of the design region. This is a possible reason why there is such a high weight put on $t = 48.0$ in the mixed model, while it is much smaller (but still important) in the fixed model (it can be observed in all computed mixed-effects designs, since the smallest weight obtained

for $t = 48.0$ equals 0.7865). Note that a similar behaviour, i.e. changes in weight values when increasing the variance of a random coefficient, was indicated by Schmelter, Benda and Schwabe (2006) in the case of D-optimum designs for a simple linear model with random slope.

As can be observed, when comparing $\xi^{\star}_{mixed}$ with $\xi^{i\star}_{mixed}$, $i = 1, 2, 3$, an increase in variance term values results in a radical change in the location of the first support point or even adds a third point into the design (in the case of $\xi^{3\star}_{mixed}$). However, when we take a look at the T-efficiencies, we can conclude that practically, the fixed-effects design $\xi^{\star}_{fixed}$ is as efficient as mixed-effects designs $\xi^{\star}_{mixed}, \xi^{1\star}_{mixed}$ to $\xi^{3\star}_{mixed}$ in discrimination between our mixed-effects models. Such a strange behaviour is probably caused by the fact that in the computation scenario the variance of the additional Gaussian noise σ_1^2 was equal to the parameter ω_1 defining the variance of the random parts of the parameters $\boldsymbol{\theta}$.

The situation changes, however, when the variance of the additional Gaussian noise is considerably smaller than that of ω_1 (designs $\xi^{4\star}_{mixed}$ and $\xi^{5\star}_{mixed}$). Now we observe that, when the value of the parameter ω increases, the T-efficiency of the fixed-effects design with respect to the particular mixed-effects design decreases. However, the efficiency is still relatively high (e.g., 0.8710 for $\xi^{5\star}_{mixed}$). Again, it is worth pointing out that such a behaviour coincides with the results obtained by Schmelter et al. (2006). Finally, we observe further changes in the location of the first support point, up to its dropping for $\xi^{5\star}_{mixed}$ (however, the resulting one-point design is singular from the viewpoint of parameter estimation).

Plots (shown in Fig. 2) aiming at checking the optimality condition given in Theorem 1 do confirm the optimality of the computed designs.

5. Conclusions and Discussion

In order to minimize the indispensable computational burden, we restricted our attention to the basic case of T-optimality considering two competing multi-response models. The results can be easily extended to discrimination between three or more competing models, cf. Atkinson and Fedorov (1975b). We also made the strong assumption about the correctness of the first model, which may not be realistic and, moreover, requires

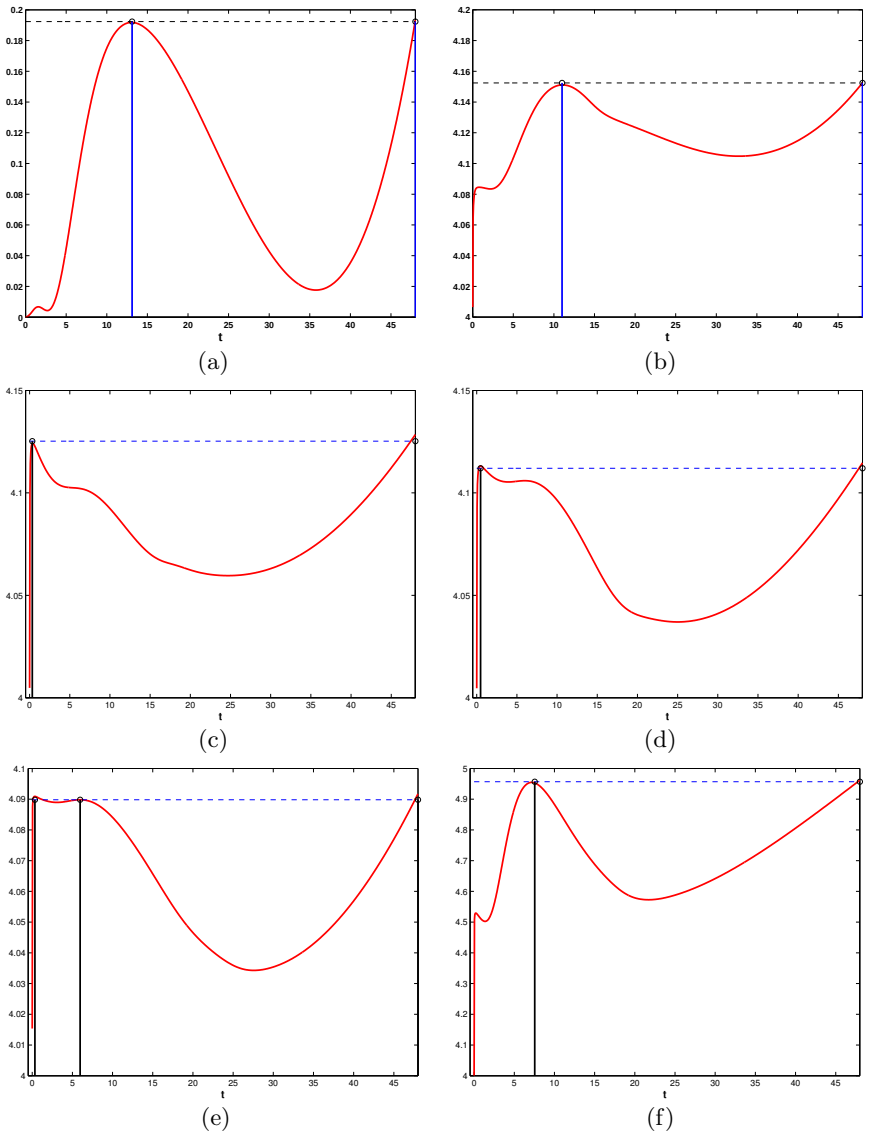

Figure 2. Sensitivity function $g(t, \boldsymbol{\gamma}_2^\star)$ vs support location (vertical lines) obtained for fixed-effects design $\xi_{fixed}^\star$ (a) and for the mixed-effects design $\xi_{mixed}^\star$ (b), $\xi_{mixed}^{1\star}$ (c), $\xi_{mixed}^{2\star}$ (d), $\xi_{mixed}^{3\star}$ (e), $\xi_{mixed}^{4\star}$ (f). This function identifies the optimum solution.

the knowledge of its, at least initial, parameter estimates. However, with some technical alterations and, unfortunately, a substantial increase in the computational cost, the approach can be extended to incorporate prior distributions for the truth of the postulated models (cf. Ponce de Leon and Atkinson (1991) providing the approach which, to some extent, also provides a solution to the parameter uncertainty problem, producing the so-called robust designs).

Another important issue concerns the efficiency of parameter estimation, which is usually rather low for T-optimum designs. There exist a number of approaches partially overcoming this drawback and providing a reasonable balance between parameter estimation and model discrimination. They base mainly on using or combining D-optimality with other criteria. The most common are (Atkinson, 2008; Waterhouse et al., 2004): DT-optimality, product D-optimality and hybrid designs (e.g., combining D- and T-optimum designs computed separately).

First of all, we have to notice that there exist situations, in which the experimenter's attention is focused on the discrimination efficiency without any compromise in the direction of parameter estimation efficiency. Such a situation often takes place when it is of primary importance to establish the type of the process when the model represents a physical phenomenon and its form gives the experimenter information about the process; for example if it is a reversible or a non-reversible chemical reaction.

Using DT-optimality, we can control the level of balance between the parameter estimation and discrimination efficiency or we can look for the design maximizing product of both the efficiencies; see, e.g. Atkinson, Donev and Tobias (2007) or Kuczewski (2006). But we have to remember that D-optimality concerns only the first of the models (the 'true' one).

A different situation takes place in the case of product D-optimum designs. It seems that they may provide quite reasonable discrimination efficiency, but at the cost of the necessity of providing initial estimates for both the models. Secondly, there are suggestions, based on empirical results, see, e.g. Waterhouse et al. (2004), that the usefulness of this approach for both the goals strongly depends on particular circumstances, mainly on the phenomenon under investigation and its models being considered.

At last, hybrid designs lose the compactness and the valuable property of the minimal size of the design produced by the aforementioned

methods. Thus, there exists no universal solution and a natural extension of further research heads towards a more thorough comparison of the proposed method with other existing approaches and combining it with the approaches increasing the parameter estimation efficiency mentioned above.

Regarding the problem of the design sensitivity with respect to the parameters $\boldsymbol{\gamma}_{\mathbf{1}}^{\mathbf{0}}$, it has to be emphasized that its direct analytical analysis is highly complicated by the non-differentiability of the criterion. However, it is obvious that the resultant design strongly depends on the random part of $\boldsymbol{\gamma}$ (see examples and remarks in the previous section). The same dependence takes place in the case of parameters $\boldsymbol{\theta}$, but our practical experience shows that its intensity correlates with the degree of the model nonlinearity with respect to the parameters $\boldsymbol{\theta}$.

The T-optimum design criterion is introduced under the assumption of normally distributed observations. Recently, López-Fidalgo, Tommasi and Trandafir (2007) proposed a new criterion based on the notion of Kullback-Leibner distance called KL-optimality. It is useful for discriminating between rival models with non-normally, e.g., log-normally or gamma, distributed observations (see the comprehensive monograph by Burnham and Anderson (2002) for application of information-based criteria, e.g., the Akaike Information Criterion, in model selection). They showed that different cases of the T-optimality may be considered as particular cases of the KL-optimality. Thus, KL-optimality seems to be a natural direction of further investigations in the considered case of mixed-effects models. However, even in the case of non-trivial fixed-effect models (e.g., the models which do not posses analytical solutions), computation of the KL distance (which involves calculation of the integral over the sample space of the possible observations) requires an immense increase in the computational burden. Moreover, one of the rival models must still be assumed as a true one.

The proposed approach can be directly applied to population pharmacokinetic studies when all subjects follow the same observational scheme. A further development of the method for groups of subjects following different designs would also be interesting.

Regarding the numerical algorithm (see Appendix), we have to address the convergence issues. As was already indicated in Fedorov and Hackl (1997, p. 95), the classical sequential Wynn-Fedorov algorithm tailored to

optimize the T-optimality criterion may suffer from the lack of convergence if a global minimizer for

$$\min_{\boldsymbol{\theta}_2 \in \Theta_2} \int_T \|\boldsymbol{\eta}(t) - \boldsymbol{\eta}_2(t, \boldsymbol{\theta}_2)\|^2 \, \xi(\mathrm{d}t), \tag{46}$$

(which has to be solved when computing $T_{12}(\xi)$ for the fixed-effects models) is not unique. The same situation takes place concerning our generalized problem of

$$\min_{\boldsymbol{\gamma}_2 \in \Gamma_2} \int_T g(t, \boldsymbol{\gamma}_2) \, \xi(\mathrm{d}t). \tag{47}$$

Although this phenomenon has been well known for many years, no viable alternative for this scheme has been proposed in the optimum experimental design literature up to recent days. The only improvements recommended in Fedorov and Hackl (1997) concern some kind of regularization. Namely, the authors suggest to replace successive designs ξ^k by $\bar{\xi}^k = (1-\gamma)\xi^k + \gamma\bar{\xi}$, where $0 < \gamma \ll 1$ and $\bar{\xi}$ is a regular design, i.e., the design for which the minimization problem (46) possesses a unique solution. But this is only a vague hint whose utility has never been formally proved and it can only be qualified as a heuristic. Kuczewski (2006) and Uciński and Kuczewski (2006) proposed a relaxation procedure which is robust to the presence of non-unique minimizers for (46). They also proved its convergence in a finite number of steps. The same work gives also a generalization of the Equivalence Theorem for T-optimum designs in the case of the existence of non-unique minimizers for the problem (46). Proceeding in much the same way, its counterpart for the situation considered in this paper can be developed.

Appendix

Details of numerical algorithm

The equivalence theorem proved in Section 3 lets us to adopt the classical iterative scheme of Wynn-Fedorov type (Wynn, 1970), which was recommended by Atkinson and Fedorov (1975a) for construction of T-optimum

designs for static single-output systems and since then generalized for the case of dynamic multivariate models (see, e.g., Kuczewski et al., 2006). The numerical algorithm can be represented by the following steps (n^k stands for size of the design in the k-th step):

Algorithm 1 (Generalized Wynn-Fedorov algorithm for T_{12}-optimum designs).

Step 1: *Guess an initial design* ξ^0 *of the form* $\xi^0 = \begin{Bmatrix} t_1^0, & \dots, & t_{n^0}^0 \\ w_1^0, & \dots, & w_{n^0}^0 \end{Bmatrix}$ *for some arbitrary* n^0. *Choose some positive tolerance* $\delta \ll 1$. *Set* $k = 0$.

Step 2: *Determine*

$$\begin{aligned} \widehat{\boldsymbol{\gamma}}_2^k &= \arg \min_{\boldsymbol{\gamma}_2 \in \Gamma_2} \sum_{i=1}^{n^k} w_i^k g(t_i^k, \boldsymbol{\gamma}_2), \\ \widehat{t}^k &= \arg \max_{t \in T} g(t, \widehat{\boldsymbol{\gamma}}_2^k). \end{aligned} \tag{48}$$

Step 3: *If* $g(\widehat{t}^k, \widehat{\boldsymbol{\gamma}}_2^k) \leq T_{12}(\xi^k) + \delta$, *then* $\xi^\star = \xi^k$ *and STOP. Otherwise go to Step 4.*

Step 4: *Choose the appropriate value of* α^k, $0 \leq \alpha^k \leq 1$, *and compute the convex combination of designs:*

$$\xi^{k+1} = (1 - \alpha^k)\xi^k + \alpha^k \delta(\widehat{t}^k), \tag{49}$$

where $\delta(\widehat{t}^k)$ *is the design concentrated only at one support point* $\widehat{t}^k$. *Set* $k \leftarrow k + 1$ *and go to Step 2.*

In contrast to the D-optimality criterion, even in the single response case, selection of an optimum value of the steplength α^k, i.e., finding

$$\alpha^k = \arg \max_{\alpha \in [0,1]} T_{12}((1 - \alpha)\xi^k + \alpha \delta(\widehat{t}^k)),$$

requires a numerical search. A simple search procedure, e.g., the golden section method (Press, Teukolsky, Vetterling and Flannery, 1996) can be

used. Alternatively, like in the D-optimality case, another common choice is

$$\alpha^k = \frac{1}{k+1} \tag{50}$$

or, in general, α^k can be chosen using any sequence satisfying the conditions (Fedorov and Hackl, 1997)

$$\lim_{k\to\infty} \alpha^k = 0, \quad \sum_{k=0}^{\infty} \alpha^k = \infty, \quad \sum_{k=0}^{\infty} (\alpha^k)^2 < \infty. \tag{51}$$

Generally, the convergence speed of the presented scheme is rather low, since it actually belongs to the group of first-order algorithms. In practice, the optimum support points are usually found relatively quickly (when using efficient global optimizers during Step 2 of the algorithm), but a precise determination of the corresponding weights takes much more time. In the statistical literature there have been some attempts to modify the basic scheme to enhance the convergence speed (Fedorov and Hackl, 1997). The resulting heuristics are intended mainly for the D-optimum design criterion, but with minor changes they can be adapted to the T-optimum criterion (see Kuczewski, 2006).

For example, one of the characteristic features of the algorithm is that the weights of the non-optimal support points gradually decrease. This eventually results in the existence of support points with negligible weights. Moreover, in each iteration a new support point is included into the design. Usually, after several iterations, the location of the new points becomes similar or very close to the existing ones. This is caused by the numerical inaccuracies of the optimization process. In order to obtain minimal-support solutions it is worthwhile to equip the implementation with procedures aimed at removing support points with negligible weights from the current design and also replacing clustered points by a single support point. The weights of the points to be replaced are added in each cluster and the clustered points are substituted by only one point with the weight equal to the resulting sum. Then removing points with negligible weights can be performed. The thresholds defining a maximal radius of the clusters (and, consequently, the number of the replaced points) and a minimum acceptable weight should be set *a priori*. The appropriate choice can speed up the

convergence, but it may happen that by setting excessively high thresholds we will obtain the effect of repeatedly removing and adding the same points into the design (and, consequently, the lack of the convergence). In the case of D-optimality a useful test exists (Pronzato, 2003) which allows for a safe removal of points which have no chance to be located in the optimum design. Unfortunately, there is no such counterpart here and the appropriate thresholds need to be chosen empirically.

A vital role in the iterative algorithm presented in this section is played by an efficient global optimization method (global optimization problems have to be solved in Step 2 of the algorithm). Since common nonlinear programming algorithms are known to converge to local optima, we have turned our attention to a stochastic optimization method called the adaptive random search (ARS), which is widely used in the engineering optimization literature (Walter and Pronzato, 1997). Based on numerous computer experiments it was found that this extremely simple strategy is especially suited for the purpose of global optimization problems in calculating T-optimum designs.

Originally, the algorithm solves a maximization problem $\max_{\boldsymbol{x}\in X} J(\boldsymbol{x})$ for an admissible set X being a hypercube.

The ARS does not use the information about the gradient of the performance index. Thus a significant numerical efficiency could hardly be expected. However, because of its valuable properties regarding global convergence and simplicity, the ARS seems to be more flexible and suitable in the case of dynamic systems than many other classical non-linear programming methods. Furthermore, gradient evaluation can be very costly or the approximation of the gradient may fail to be satisfactory (e.g., there may occur some scaling problems or insufficient smoothness of the underlying functions). Nevertheless, the performance of the ARS can be improved by combination with various other methods, so occasionally we can switch to local maximization in order to make the results more accurate.

Concerning computation complexity of the ARS algorithm, it is, like the majority of global optimization algorithms, polynomial of higher order (Spall, 2003). Thus, an increase in the number of parameters will result in a considerable increase in the run time.

Acknowledgments

This research was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) grant EP/C54171/1. Also, we are grateful to the referee of the paper for constructing comments and remarks.

References

Atkinson A.C. (2008): DT-optimum designs for model discrimination and parameter estimation, Journal of Statistical Planning and Inference 138: 56–64.

Atkinson A.C., Donev A.N., Tobias R. (2007): Optimum Experimental Designs, with SAS, Oxford University Press, Oxford.

Atkinson A.C., Fedorov V.V. (1975a): The design of experiments for discriminating between two rival models, Biometrika 62(1): 57–70.

Atkinson A.C., Fedorov V.V. (1975b): Optimal design: Experiments for discriminating between several models, Biometrika 65(2): 289–303.

Bogacka B., Johnson P., Jones B., Volkov O. (2008): D-efficient window experimental designs, Journal of Statistical Planning and Inference 138(1): 160–168.

Box G.E.P., Hill W.J. (1967): Discrimination among mechanistics models, Technometrics 9(1): 57–71.

Burnham K.P., Anderson D.R. (2002): Model Selection and Multimodal Inference. A Practical Information-Theoretic Approach, 2nd edn, Springer-Verlag, New York.

Dawidian M., Giltinan D.M. (1995): Nonlinear Models for Repeated Measurement Data, Vol. 62 of Monographs on Statistics and Applied Probability, Chapman & Hall/CRC, Boca Raton.

Demidenko E. (2004): Mixed Models. Theory and Applications, Wiley Series in Probability and Statistics, Wiley Interscience, New York.

Fedorov V.V., Hackl P. (1997): Model-Oriented Design of Experiments, Lecture Notes in Statistics, Springer-Verlag, New York.

Fedorov V.V., Khabarov V. (1986): Duality of optimal design for model discrimination and parameter estimation, Biometrika 73(1): 183–190.

Felsenstein K. (1992): Optimal Bayesian design for discriminating among rival models, Computational Statistics & Data Analysis 14: 427–436.

Gagnon R., Leonov S. (2005): Optimal population designs for PK models with serial sampling, Journal of Biopharmaceutical Statistics 15(1): 143–163.

Han C., Chaloner K. (2004): Bayesian experimental design for nonlinear mixed-effects models with application to HIV dynamics, Biometrics 60(1): 25–33.

Jones B., Wang, J. (1999): Constructing optimal designs for fitting pharmacokinetic models, Statistics and Computing 9(3): 209–218.

Kuczewski B. (2006): Computational Aspects of Discrimination between Models of Dynamic Systems, Vol. 10 of Lecture Notes in Control and Computer Science, University of Zielona Góra Press, Zielona Góra. Available at: http://zbc.uz.zgora.pl/publication/2553.

Kuczewski B., Baranowski P., Uciński D. (2006): Parallel processing in discrimination between models of dynamic systems, in R. Wyrzykowski, J. Dongarra, N. Meyer and J. Waśniewski (eds), Parallel Processing and Applied Mathematics, PPAM 2005, Vol. 3911 of LNCS, Springer-Verlag, Berlin Heidelberg: 340–348.

Lehmann E.L., Romano J.P. (2005): Testing Statistical Hypothesis, Springer Texts in Statistics, third edn, Springer, New York.

Lindsey J.K. (2001): Nonlinear Models in Medical Statistics, Oxford Statistical Science Series, Oxford University Press, Oxford.

López-Fidalgo J., Tommasi C., Trandafir P. (2007): An optimal experimental design criterion for discriminating between non-normal models, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 69(2): 231–242.

Mentré F., Mallet A., Baccar D. (1997): Optimal design in random-effects regression models, Biometrika 84(2): 429–442.

Müller W.G., Ponce de Leon A.C. (1996): Discriminating between two binary data models: Sequentially designed experiments, Journal of Statistical Computation and Simulation 55: 87–100.

Patan M., Bogacka B. (2007): Efficient sampling windows for parameter estimation in mixed effects models, in J. López-Fidalgo, J. M. Rodrígez-Díaz and B. Torsney (eds), mODa 8, Proc. 8th Int. Workshop on Model-Oriented Data Analysis, Almagro, Spain, 2007, Physica-Verlag, Heidelberg: 165–174. In print.

Pinheiro J.C., Bates D.M. (2000): Mixed-Effects Models in S and S-PLUS, Statistics and Computing, Springer-Verlag, New York.

Polak E. (1997): Optimization. Algorithms and Consistent Approximations, Applied Mathematical Sciences, Springer-Verlag, New York.

Ponce de Leon A.C., Atkinson A.C. (1991): Optimum experimental design for discriminating between two rival models in the presence of prior information, Biometrika 78(3): 601–608.

Press W.H., Teukolsky S.A., Vetterling W.T., Flannery B.P. (1996): Numerical Recipes in Fortran 90. The Art of Parallel Scientific Computing, 2-nd edn, Cambridge University Press, Cambridge.

Pronzato L. (2003): Removing non-optimal support points in D-optimum design algorithms, Statistics & Probability Letters 63: 223–228.

Pshenichnyi B.N. (1971): Necessary Conditions for an Extremum, Marcel Dekker, Inc., New York.

Rao M.M. (1987): Measure Theory and Integration, John Wiley & Sons, New York.

Retout S., Duffull S., Mentré F. (2001): Developement and implementation of the population fisher information matrix for the evaluation of population pharmacokinetic designs, Computer Methods and Programs in Biomedicine 65(2): 141–151.

Retout S., Mentré F. (2003): Further developement of the fisher information matrix in nonlinear mixed effects models with evaluation in population pharmacokinetics, Journal of Biopharmaceutical Statistics 13(2): 209–227.

Schmelter T., Benda N., Schwabe R. (2006): Some curiosities in optimal designs for random slopes, Preprint, Institute for Mathematical Stochastics, Otto-von-Guernicke University, Magdeburg. Available at: http://www.math.uni-magdeburg.de/∼schwabe/Preprints/2006_39.pdf.

Searle S.R. (1971): Linear Models, John Wiley & Sons, New York.
Searle S.R., Casella G., McCulloch C.E. (1992): Variance Components, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics Section, John Wiley & Sons, New York.
Spall J.C. (2003): Introduction to Stochastic Search and Optimization. Estimation, Simulation, and Control, John Wiley & Sons, Hoboken, NJ.
Stewart W.E., Shon Y., Box G.E.P. (1998): Discrimination and goodness of fit of multiresponse mechanistic models, AIChE Journal 44(6): 1404–1412.
Uciński D. (2005): Optimal Measurement Methods for Distributed Parameter System Identification, CRC Press, Boca Raton.
Uciński D., Bogacka B. (2004): T-optimum designs for multiresponse dynamic heteroscedastic models, in A. Di Bucchianico, H. Läuter and H. P. Wynn (eds), mODa 7, Proc. 7th Int. Workshop on Model-Oriented Data Analysis, Heeze, The Netherlands, 2004, Physica-Verlag, Heidelberg: 191–199.
Uciński D., Bogacka B. (2005): T-optimum designs for discrimination between two multivariate dynamic models, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67: 3–18.
Uciński D., Kuczewski B. (2006): Sensor placement for discrimination between rival models of distributed parameter systems, Proc. 45th IEEE Conference on Decision and Control, San Diego, USA. On CD-ROM.
Walter É., Pronzato L. (1997): Identification of Parametric Models from Experimental Data, Communications and Control Engineering, Springer-Verlag, Berlin.
Waterhouse T.H., Eccleston J.A., Duffull S.B. (2004): On optimal design for discrimination and estimation, in J. Antoch (ed.), COMPSTAT 2004: Proceedings in Computational Statistics, Physica-Verlag, Heidelberg: 1963–1970.
Waterhouse T.H., Redmann S., Duffull S.B., Eccleston J.A. (2005): Optimal design for model discrimination and parameter estimation for intraconazole population pharmacokinetics in cystic fibrosis patients, Journal of Pharmacokinetics & Pharmacodynamics 32(3-4): 521–545.
Waterhouse T.H., Woods D.C., Eccleston J.A., Lewis S.M. (2006): Design selection criteria for discrimination between nested models for binomial data, Methodology Working Paper M06/02, Southampton Statistical Sciences Research Institute, University of Southampton, Southampton. Available at: http://www.s3ri.soton.ac.uk/publications/methodology.php.
Wynn H.P. (1970): The sequential generation of D-optimum experimental designs, The Annals of Mathematical Statistics 41: 1655–1664.
Yuh L., Beal S., Davidian M., Harrison F., Hester A., Kowalski K., Vonesh E., Wolfinger R. (1994): Population pharmacokinetic/pharmacodynamic metodology and applications: A bibliography, Biometrics 50(2): 566–575.

Legend

1 Interpretation of models parameters and constant values (after Waterhouse et al., 2005).
2 Scheme of the compartmental model of the pharmacokinetics of itraconazole and hydroxyitraconazole (after Waterhouse et al., 2005).
3 Sensitivity function $g(t, \boldsymbol{\gamma}_2^\star)$ vs support location (vertical lines) obtained for fixed-effects models design $\xi_{fixed}^\star$ (a) and for the mixed-effects models design $\xi_{mixed}^\star$ (b), $\xi_{mixed}^{1\star}$ (c), $\xi_{mixed}^{2\star}$ (d), $\xi_{mixed}^{3\star}$ (e), $\xi_{mixed}^{4\star}$ (f) indicates the optimum solution.