Optimum designs for non-linear mixed effects models in the presence of covariates

Barbara Bogacka [a] ⋆, A. H. M. Mahbub Latif [b], Steven Gilmour [c], Kuresh Youdim [d]

[a] School of Mathematical Sciences, Queen Mary, University of London, London E1 4NS, UK

[b] Institute of Statistical Research and Training, University of Dhaka, Dhaka-1000, Bangladesh

[c] Southampton Statistical Sciences Research Institute, University of Southampton, Highfield, Southampton SO17 1BJ, UK

[d] Roche Pharmaceutical Research and Early Development, Pharmaceutical Sciences, Roche Innovation Center Basel, F. Hoffmann-La Roche Ltd, Grenzacherstrasse 124, 4070 Basel, Switzerland

⋆*e-mail: b.bogacka@qmul.ac.uk*

Abstract

In this paper we present a new method for optimizing designs of experiments for non-linear mixed effects models, where a categorical factor with covariate information is a design variable combined with another design factor. The work is motivated by the need to efficiently design pre-clinical experiments in enzyme kinetics for a set of Human Liver Microsomes. However, the results are general and can be applied to other experimental situations where the variation in the response due to a categorical factor can be partially accounted for by a covariate. The covariate included in the model explains some systematic

variability in a random model parameter. This approach allows better understanding of the population variation as well as estimation of the model parameters with higher precision.

Keywords: Covariates, Enzyme Kinetics, Planning Experiments, Random Model Parameters

1 Introduction

Nonlinear mixed effects models have been extensively used in various applications, in particular in evaluation of the population pharmacokinetic and pharmacodynamic parameters in drug development studies, see for example Chapter 5 of Fitzmaurice et al. (2009). It has been shown that design optimization in the mixed models approach can reduce the number of observations per individual and at the same time provide good predictions of the individual responses as well as good estimates of the population parameters including their variances. This is particularly important in studies in which many observations per individual are not feasible, for example when the patients are small children. Further improvements are possible by including covariates which are partially responsible for the inter-individual variability of the observed responses.

There has been some work done on covariate selection in non-linear mixed effects model building. Wu and Wu (2002) consider such a method in application to HIV data (viral decay rate), which are usually very sparse for individuals, although the number of individuals and the inter-individual variation may be large. There is also a large number of covariates available but only some of the covariates are related to the base model parameters and account for inter-individual variation in the measured viral decay rate. The authors compared various methods of choosing a model given the data when there are missing covariate values. There is however an important

open question of how to choose the experimental variables so that the observed responses (data) are most informative. The main design variable in this instance is the time of measuring decay rate, but then the question is which individuals, that is which covariate values, will give best information for the efficient model building and evaluation.

Ding and Wu (2001) added an indicator variable as a covariate in their mixed effects model of viral decay for antiviral drugs in HIV to represent a treatment. The model parameters depend on the covariate. They compared type I errors for various tests for the hypothesis of equality of the treatment effects. In Wu and Ding (2002) they examined the effect of various sampling times on the power for identifying a treatment difference. Retout et al. (2007) showed the efficiency of D-optimum designs on the Wald test for such differences. However, they did not consider the choice of covariate values in the design optimization.

Denti et al. (2010) showed that relating the model parameters to a selection of covariates can decrease significantly the inter-subject variability due to random individual effects. A part of the population variability can be explained by the covariates. They also pointed out that this can lead to savings in the numbers of individual observations and so to increasing the efficiency of the experiments. They also noted the potential of including covariate information at the design stage. We are not aware, however, that this has been done.

Our work was initially motivated by applications in pre-clinical studies of evaluating potential drug-drug interactions. The *in-vitro* experiments are done in Human Liver Microsomes (HLM) as it is in the liver that most of the enzymes responsible for metabolism occur. These are Cytochrome P450 enzymes (CYPs) and they can be partly responsible for the inter-HLM (inter-subject) variability in respect of the drug metabolism, cf. Hasler et al. (1999).

Belle et al. (2000) showed that population analysis of sparse data can reduce coefficients of vari-

ation of some parameters in enzyme kinetic models. In experiments with HLMs they expressed the inter-HLM variability in terms of the parameter V_{max} of a two-enzyme kinetic model. They also found that the variability in this parameter was related to activity of CYP1A2. Including the activity as a covariate into the model reduced the coefficient of variation (CV) of the parameter estimator from 70% to 39% and it also reduced slightly the error variance estimator's CV (intra-HLM variability). They do not however consider optimum planning of the experiment.

In this paper we present a method of planning experiments where covariates are treated as design variables whose values are chosen, from the set of values in the available subjects, by the experimenter in combination with other treatment factors. In Section 2 we present notation for a general model and a population design, as well as a design optimality criterion. We also derive the model approximation and the information matrix form. Section 3 is devoted to the application in enzyme kinetics. We also show how the criterion of optimality can be extended to allow for a model transformation in case the residuals do not follow the model assumptions. We briefly present the numerical optimization algorithm and the results. Finally we comment on our findings and in Section 4 we give brief conclusions.

2 Theory and methods

2.1 Modelling

Assume that we intend to study the effect of a treatment factor on members of a population. We denote by $\boldsymbol{x}$ a vector of levels of a treatment factor, continuous or discrete, although in regression models they are usually continuous, such as concentration of a drug injected into blood plasma, time of taking measurements or temperature at which a chemical reaction is run. We

also assume that the population under investigation is diverse in its nature and the diversity can be partially explained by some concomitant variables (covariates) of the population members. These, for example, could be size of a tumor, its grade and a number of affected lymph nodes of cancer patients taking part in a clinical trial or enzyme activity in drug metabolism in preclinical studies. We denote by $\boldsymbol{z}$ a vector of such covariates. If the purpose of the experiment is to estimate and make inferences on some treatment parameters, then the following question arises: are some values of the covariates more informative than others and, if so, what combinations of individuals with the values of the treatment factors should be used? In Section 3 we show that we can improve the efficiency of estimation of Michaelis-Menten model parameters by an optimum selection of liver tissue preparations characterized by enzymes' activities combined with levels of concentration of the drug under investigation.

We denote by I_s a set of available elements of the population S, that is, $I_s = \{1, 2, \ldots, s\}$ and by $I_{(n)}$ any subset of I_s of size n, that is $I_{(n)} = \{(1), (2), \ldots, (n)\}$, where the round brackets are used to index the elements in $I_{(n)}$. For example, I_s could be all potential HLMs which could be included in an experiment and $I_{(n)}$ a set of HLMs actually selected for the experiment. We could also think of S as a treatment factor with s available levels. For simplicity, we will call the elements of population S *subjects*. Furthermore, let $\boldsymbol{x} \in \mathcal{X} \subset \mathbb{R}^t$ and let $\boldsymbol{z} \in \mathcal{Z} \subset \mathbb{R}^q$, where t and q are some natural numbers.

In general terms, we can write a response model for the j-th value of $\boldsymbol{x}$, (k)-th subject and i-th replication as follows,

$$y_{ij(k)} = \eta\big(\boldsymbol{x}_{j(k)}, \boldsymbol{\beta}_{(k)}\big) + \varepsilon_{ij(k)}, \quad i = 1, \ldots, r_{j(k)},\ j = 1, \ldots, n_{(k)},\ (k) \in I_{(n)}, \tag{2.1}$$

where η is a model function relating the response to the runs of the experiment. In most applications relevant to the work presented in this paper it is a non-linear function with respect to

both the explanatory variables and the parameters. Furthermore, $\boldsymbol{\beta}_{(k)}$ denotes a p-dimensional vector of functions of the covariates $\boldsymbol{z}_{(k)}$ associated with the (k)-th subject, unknown constant parameters $\boldsymbol{\beta} \in \mathbb{R}^{p_1}$ and a p_2-dimensional vector of random effects $\boldsymbol{b}_{(k)}$, that is,

$$\boldsymbol{\beta}_{(k)} = \begin{pmatrix} g_{(k)1}(\boldsymbol{\beta}, \boldsymbol{b}_{(k)}, \boldsymbol{z}_{(k)}) \\ \vdots \\ g_{(k)p}(\boldsymbol{\beta}, \boldsymbol{b}_{(k)}, \boldsymbol{z}_{(k)}) \end{pmatrix},$$

where functions g can be linear or non-linear with respect to both the parameters and the covariates.

The total number of observations is $N = \sum_{(k)\in I_{(n)}} \sum_{j=1}^{n_{(k)}} r_{j(k)} = \sum_{(k)\in I_{(n)}} m_{(k)}$, where $m_{(k)} = \sum_{j=1}^{n_k} r_{j(k)}$. We assume that

$$\boldsymbol{b}_{(k)} \sim \mathcal{N}_{p_2}(\boldsymbol{0}, \boldsymbol{\Sigma}), \;\; \boldsymbol{\varepsilon}_{(k)} | b_{(k)} \sim \mathcal{N}_{m_{(k)}}(\boldsymbol{0}, \sigma_\varepsilon^2 \boldsymbol{I}) \quad \text{for all } (k) \in I_{(n)}, \tag{2.2}$$

where $\boldsymbol{\varepsilon}_{(k)}$ denotes the vector of random errors for subject (k). We denote by $\boldsymbol{\gamma}$ a vector of all the model parameters of interest, that is

$$\boldsymbol{\gamma} = (\boldsymbol{\beta}^{\mathrm{T}}, \boldsymbol{\sigma}^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{p_1 + \frac{p_2(p_2-1)}{2}} \times \mathbb{R}_+^{p_2+1}, \tag{2.3}$$

where $\boldsymbol{\beta}^{\mathrm{T}} = (\beta_1, \ldots, \beta_{p_1})$ and $\boldsymbol{\sigma}^{\mathrm{T}} = (\sigma_1^2, \ldots, \sigma_{p_2}^2, \sigma_{12}, \ldots, \sigma_{p_2-1,p_2}, \sigma_\varepsilon^2)$ consists of the variances and covariances of the random effects vector $\boldsymbol{b}$ (elements of matrix $\boldsymbol{\Sigma}$) and the error variance.

2.2 Design

Each subject $(k) \in I_{(n)}$ is characterized by some covariates $\boldsymbol{z}_{(k)}$. For an efficient experiment we need to choose subjects, which implies the relevant levels of the covariates, paired with values

of the vector $\boldsymbol{x}$. A set of such pairs for subject (k) is denoted by

$$\xi_{(k)} = \left\{ \begin{array}{ccc} \begin{pmatrix} \boldsymbol{z}_{1(k)} \\ \boldsymbol{x}_{1(k)} \end{pmatrix} & \dots & \begin{pmatrix} \boldsymbol{z}_{n_{(k)}(k)} \\ \boldsymbol{x}_{n_{(k)}(k)} \end{pmatrix} \\ r_{1(k)} & \dots & r_{n_{(k)}(k)} \end{array} \right\}, \quad (k) \in I_{(n)},\ \boldsymbol{x}_{j(k)} \in \mathcal{X},\ \boldsymbol{z}_{j(k)} \in \mathbb{R}^{q_{(k)}},\ r_{j(k)} \in \mathbb{N}. \tag{2.4}$$

The replications $r_{j(k)}$ of the support points $\begin{pmatrix} \boldsymbol{z}_{j(k)} \\ \boldsymbol{x}_{j(k)} \end{pmatrix}$ are natural numbers, that is we consider exact designs. Further in this paper we assume that for a given level (k) the same covariates are used and we assume that their values do not change with the changes in $\boldsymbol{x}$. Hence, index j in $\boldsymbol{z}_{j(k)}$ can be omitted and the design (2.4) can be written in the following way

$$\xi_{(k)} = \left\{ \begin{array}{ccc} \boldsymbol{x}_{1(k)} & \dots & \boldsymbol{x}_{n_{(k)}(k)} \\ r_{1(k)} & \dots & r_{n_{(k)}(k)} \end{array} ;\ \boldsymbol{z}_{(k)} \right\}. \tag{2.5}$$

The experiment is performed over a subset $I_{(n)}$ of the available subjects of population S and the design for the subset is denoted by

$$\zeta = \left\{ \xi_{(1)}, \dots, \xi_{(n)} \right\}. \tag{2.6}$$

We call ζ a *population design*, where subjects (k) and values of the explanatory variables $\boldsymbol{x}$ are the *design variables*. Each subject chosen for the experiment has its individual plan of the experiment (*individual design*) $\xi_{(k)}$. In some cases, all subjects will have the same individual design. Sometimes we will have groups of subjects having the same individual design. The theory presented in this paper covers all such cases as long as the subjects are independent.

2.2.1 Criterion of design optimality

We are interested in efficient estimation of the model parameters $\boldsymbol{\gamma}$ as defined in (2.3) and we choose the D-criterion for finding optimum designs. We have q covariates available, but in fact

we may be interested in a subset of the covariates only. Two special cases are all q covariates or a single covariate included in the model.

We denote the information matrix corresponding to model (2.1) by $\boldsymbol{M}(\zeta, \boldsymbol{\gamma})$. Then the criterion of optimality can be written as

$$\psi_{\boldsymbol{\gamma}}(\zeta, \boldsymbol{\gamma}) = \log \det \boldsymbol{M}(\zeta, \boldsymbol{\gamma}). \tag{2.7}$$

By definition, the information matrix for $\boldsymbol{\gamma}$ is equal to

$$\boldsymbol{M} = -\,\mathrm{E}\begin{pmatrix} \frac{\partial^2 \ell}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^{\mathrm{T}}} & \frac{\partial^2 \ell}{\partial \boldsymbol{\beta} \partial \boldsymbol{\sigma}^{\mathrm{T}}} \\ \left(\frac{\partial^2 \ell}{\partial \boldsymbol{\beta} \partial \boldsymbol{\sigma}^{\mathrm{T}}}\right)^{\mathrm{T}} & \frac{\partial^2 \ell}{\partial \boldsymbol{\sigma} \partial \boldsymbol{\sigma}^{\mathrm{T}}} \end{pmatrix},$$

where ℓ denotes the log-likelihood function for the parameters given the observations. Here, however, the marginal density function of vector $\boldsymbol{y}$ whose entries are as in (2.1) does not have a closed form. To approximate the distribution a Taylor series expansion of the model is often applied. The resulting linear combination of random variables gives a normal distribution if the variables are normal.

2.3 Model approximation

Lindstrom and Bates (1990) and Gilberg et al. (1999) use the first order approximation of the Taylor expansion of the model function η around the fixed parameters and the random effects at their estimates. They are interested in methods for parameter estimation and they assume that there are data available to calculate the required initial estimates. At the design stage there are no data to hand. Hence, we evaluate the approximation at a prior guess of the fixed effects and at the assumed expectation of the random effects. In our paper we are using a point prior,

denoted by $\boldsymbol{\beta}^0$, and we assume that $\mathrm{E}(\boldsymbol{b}_{(k)}) = \mathbf{0}$. That is, we expand the model for

$$\boldsymbol{\vartheta} = \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{b}_{(k)} \end{pmatrix} \quad \text{about} \quad \boldsymbol{\vartheta}^0 = \begin{pmatrix} \boldsymbol{\beta}^0 \\ \mathrm{E}(\boldsymbol{b}_{(k)}) \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta}^0 \\ \mathbf{0} \end{pmatrix}$$

and approximate the model by truncating the expansion as follows:

$$\begin{aligned} \eta_{j(k)} &\cong \eta_{j(k)}\big|_{\boldsymbol{\vartheta}^0} + \left(\frac{\partial \eta_{j(k)}}{\partial \boldsymbol{\vartheta}}\big|_{\boldsymbol{\vartheta}^0}\right)^{\mathrm{T}} \left(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}^0\right) \\ &= \eta_{j(k)}\big|_{\boldsymbol{\vartheta}^0} + \left(\frac{\partial \eta_{j(k)}}{\partial \boldsymbol{\beta}}\big|_{\boldsymbol{\vartheta}^0}\right)^{\mathrm{T}} \left(\boldsymbol{\beta} - \boldsymbol{\beta}^0\right) + \left(\frac{\partial \eta_{j(k)}}{\partial \boldsymbol{b}_{(k)}}\big|_{\boldsymbol{\vartheta}^0}\right)^{\mathrm{T}} \boldsymbol{b}_{(k)} \\ &= \alpha_{j(k)} + \left(\frac{\partial \eta_{j(k)}}{\partial \boldsymbol{\beta}}\big|_{\boldsymbol{\vartheta}^0}\right)^{\mathrm{T}} \boldsymbol{\beta} + \left(\frac{\partial \eta_{j(k)}}{\partial \boldsymbol{b}_{(k)}}\big|_{\boldsymbol{\vartheta}^0}\right)^{\mathrm{T}} \boldsymbol{b}_{(k)}, \end{aligned} \tag{2.8}$$

where

$$\alpha_{j(k)} = \eta_{j(k)}\big|_{\boldsymbol{\vartheta}^0} - \left(\frac{\partial \eta_{j(k)}}{\partial \boldsymbol{\beta}}\big|_{\boldsymbol{\vartheta}^0}\right)^{\mathrm{T}} \boldsymbol{\beta}^0.$$

Here the derivatives are evaluated at known values $\boldsymbol{\vartheta}^0$ and so are known functions. Furthermore, due to the chain rule, the derivatives can be written as

$$\begin{aligned} \frac{\partial \eta_{j(k)}}{\partial \beta_l} &= \frac{\partial \eta_{j(k)}}{\partial\, g_{(k)l'}} \frac{\partial g_{(k)l'}}{\partial \beta_l} \qquad l' = 1, \ldots, p,\ l = 1, \ldots, p_1 \\ \frac{\partial \eta_{j(k)}}{\partial b_{(k)l}} &= \frac{\partial \eta_{j(k)}}{\partial\, g_{(k)l'}} \frac{\partial g_{(k)l'}}{\partial b_{(k)l}} \qquad l' = 1, \ldots, p,\ l = 1, \ldots, p_2 \end{aligned} \tag{2.9}$$

Writing the above in the matrix notation we have

$$\eta_{j(k)} = \alpha_{j(k)} + \boldsymbol{f}_{j(k)}^{\mathrm{T}} \boldsymbol{Z}_{(k)} \boldsymbol{\beta} + \boldsymbol{f}_{j(k)}^{\mathrm{T}} \boldsymbol{H}_{(k)} \boldsymbol{b}_{(k)}, \tag{2.10}$$

where

$$\begin{aligned} \boldsymbol{f}_{j(k)}^{\mathrm{T}} &= \left(\frac{\partial \eta_{j(k)}}{\partial\, g_{(k)1}}, \ldots, \frac{\partial \eta_{j(k)}}{\partial\, g_{(k)p}}\right)\Big|_{\boldsymbol{\vartheta}^0} \\ \boldsymbol{Z}_{(k)} &= \left(\left\{\frac{\partial g_{(k)l'}}{\partial \beta_l}\right\}_{l'=1,\ldots,p;\, l=1,\ldots,p_1}\right)\Big|_{\boldsymbol{\vartheta}^0} \\ \boldsymbol{H}_{(k)} &= \left(\left\{\frac{\partial g_{(k)l'}}{\partial b_{(k)l}}\right\}_{l'=1,\ldots,p;\, l=1,\ldots,p_2}\right)\Big|_{\boldsymbol{\vartheta}^0} \end{aligned}$$

Then, model (2.1) including all observations for individual (k) can be written as

$$\boldsymbol{y}_{(k)} \cong \boldsymbol{\alpha}_{(k)} + \boldsymbol{F}_{(k)} \boldsymbol{Z}_{(k)} \boldsymbol{\beta} + \boldsymbol{F}_{(k)} \boldsymbol{H}_{(k)} \boldsymbol{b}_{(k)} + \boldsymbol{\varepsilon}_{(k)}, \tag{2.11}$$

where $\boldsymbol{\alpha}_{(k)}$ is the $m_{(k)}$-dimensional vector of constants $\alpha_{j(k)}$, each repeated $r_{j(k)}$ times, $\boldsymbol{F}_{(k)}$ is the $(m_{(k)} \times p)$-dimensional matrix whose rows are $\boldsymbol{f}_{j(k)}^{\mathrm{T}}$, each row repeated $r_{j(k)}$ times.

We assume that random vectors $\boldsymbol{b}_{(k)}$ and $\boldsymbol{\varepsilon}_{(k)}$ are independent and have multivariate normal distributions as in (2.2). Hence, the approximate expectation and the dispersion matrix of vector $\boldsymbol{y}_{(k)}$ are, respectively,

$$\boldsymbol{\mu}_{(k)} = \mathrm{E}(\boldsymbol{y}_{(k)}) \cong \boldsymbol{\alpha}_{(k)} + \boldsymbol{F}_{(k)}\boldsymbol{Z}_{(k)}\boldsymbol{\beta}$$

and

$$\boldsymbol{V}_{(k)} = \mathrm{Var}(\boldsymbol{y}_{(k)}) \cong \boldsymbol{F}_{(k)}\boldsymbol{H}_{(k)}\boldsymbol{\Sigma}\boldsymbol{H}_{(k)}^{\mathrm{T}}\boldsymbol{F}_{(k)}^{\mathrm{T}} + \sigma_{\varepsilon}^{2}\boldsymbol{I}_{m_{(k)}}.$$

The distribution of $\boldsymbol{y}_{(k)}$ is approximately normal, that is,

$$\boldsymbol{y}_{(k)} \underset{approx}{\sim} \mathcal{N}_{m_{(k)}}(\boldsymbol{\mu}_{(k)}, \boldsymbol{V}_{(k)}),$$

where $\boldsymbol{\mu}_{(k)}$ depends on the vector of parameters $\boldsymbol{\beta}$, while $\boldsymbol{V}_{(k)}$ depends on the vector of the variances and covariances $\boldsymbol{\sigma}$.

2.4 Fisher Information Matrix

The log-likelihood function for the full vector of parameters, given the responses of (k)-th subject, is

$$\ell_{(k)} = \ell(\boldsymbol{\beta}, \boldsymbol{\sigma}|\boldsymbol{y}_{(k)}) = const. - \frac{1}{2}\big(\boldsymbol{y}_{(k)} - \boldsymbol{\mu}_{(k)}\big)^{\mathrm{T}}\boldsymbol{V}_{(k)}^{-1}\big(\boldsymbol{y}_{(k)} - \boldsymbol{\mu}_{(k)}\big) - \frac{1}{2}\log\det(\boldsymbol{V}_{(k)}).$$

The Fisher Information Matrix for subject $(k) \in I_{(n)}$ is

$$\boldsymbol{M}_{(k)} = -\,\mathrm{E}\begin{pmatrix} \frac{\partial^2 \ell_{(k)}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\mathrm{T}}} & \frac{\partial^2 \ell_{(k)}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\sigma}^{\mathrm{T}}} \\ \left(\frac{\partial^2 \ell_{(k)}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\sigma}^{\mathrm{T}}}\right)^{\mathrm{T}} & \frac{\partial^2 \ell_{(k)}}{\partial\boldsymbol{\sigma}\partial\boldsymbol{\sigma}^{\mathrm{T}}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{B}_{(k)} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{C}_{(k)} \end{pmatrix},$$

where

$$\boldsymbol{B}_{(k)} = \frac{\partial \boldsymbol{\mu}_{(k)}^{\mathrm{T}}}{\partial \boldsymbol{\beta}} \boldsymbol{V}_{(k)}^{-1} \left(\frac{\partial \boldsymbol{\mu}_{(k)}^{\mathrm{T}}}{\partial \boldsymbol{\beta}} \right)^{\mathrm{T}} = \boldsymbol{Z}_{(k)}^{\mathrm{T}} \boldsymbol{F}_{(k)}^{\mathrm{T}} \boldsymbol{V}_{(k)}^{-1} \boldsymbol{F}_{(k)} \boldsymbol{Z}_{(k)}$$

$$\boldsymbol{C}_{(k)} = \frac{1}{2} \left\{ \mathrm{trace} \left(\boldsymbol{V}_{(k)}^{-1} \frac{\partial \boldsymbol{V}_{(k)}}{\partial \sigma_i} \boldsymbol{V}_{(k)}^{-1} \frac{\partial \boldsymbol{V}_{(k)}}{\partial \sigma_{i'}} \right) \right\}_{i,i'=1,\ldots,p_3},$$

where σ_i are elements of the p_3-dimensional vector $\boldsymbol{\sigma}$. We assume that the subjects are independent. Hence, the information matrix for the whole design, which we will call the *Population Fisher Information Matrix* and denote by $\boldsymbol{M}$, is the sum of the individual information matrices. That is,

$$\boldsymbol{M} = \sum_{(k)\in I_{(n)}} \boldsymbol{M}_{(k)} = \sum_{(k)\in I_{(n)}} \begin{pmatrix} \boldsymbol{B}_{(k)} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{C}_{(k)} \end{pmatrix} = \begin{pmatrix} \sum_{(k)\in I_{(n)}} \boldsymbol{B}_{(k)} & \boldsymbol{0} \\ \boldsymbol{0} & \sum_{(k)\in I_{(n)}} \boldsymbol{C}_{(k)} \end{pmatrix} \tag{2.12}$$

Then, the D-optimality criterion function is

$$\log\det \boldsymbol{M} = \log\det \begin{pmatrix} \boldsymbol{B} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{C} \end{pmatrix} = \log\det \boldsymbol{B} \det \boldsymbol{C} = \log\det \boldsymbol{B} + \log\det \boldsymbol{C},$$

where

$$\boldsymbol{B} = \sum_{(k)\in I_{(k)}} \boldsymbol{B}_{(k)}, \text{ and } \boldsymbol{C} = \sum_{(k)\in I_{(k)}} \boldsymbol{C}_{(k)}.$$

3 Application

3.1 Enzyme kinetics model

In a typical enzyme kinetics reaction enzymes bind to substrates and turn them into products. The binding step is reversible while the catalytic step irreversible. In chemical notation

$$S + E \longleftrightarrow ES \rightarrow E + P,$$

where S, E and P denote substrate, enzyme and product. The reaction rate is represented by the Michaelis-Menten model $v = \dfrac{V_{\max} x}{K_{\mathrm{m}} + x}$, where x is the concentration of the substrate ($[S]$) and $V_{\max}$ and K_{m} are the model parameters: $V_{\max}$ denotes the maximum velocity of the enzyme and K_{m} is the Michaelis-Menten constant; it is the value of x at which half of the maximum velocity $V_{\max}$ is reached.

In our example I_s is the set of all available HLM preparations. We have $s = 47$ and the substrate concentration x is assumed to belong to the interval $\mathcal{X} = [0.3,\ 50]$.

Typically, there would be several concentration levels and each would be used for measuring the response (the reaction rate) from each HLM. In Figure 4.2 we present observations for the 47 liver microsomal preparations of such a standard experiment. We will call the design *rich*. The differences among the subjects are clearly seen in the values of parameter $V_{\max}$, the horizontal asymptote of the Michaelis-Menten model.

There are six cytochrome P450 enzymes specific for each HLM, characterized by the enzyme activities (covariates $\boldsymbol{z}$, $q = 6$).

We model the response function η as the Michaelis-Menten function, where the parameters may depend on the covariate values of the subjects (HLMs), which is indicated by the index (k), that is,

$$\eta\big(x_{j(k)}, \boldsymbol{\beta}_{(k)}\big) = \frac{V_{\max(k)} x_{j(k)}}{K_{\mathrm{m}(k)} + x_{j(k)}}.$$

If we include just one covariate related to $V_{\max(k)}$ only (activity of one enzyme) then vector $\boldsymbol{\beta}_{(k)}$ can be written in the matrix form as follows

$$\boldsymbol{\beta}_{(k)} = \begin{pmatrix} V_{\max(k)} \\ \\ K_{\mathrm{m}} \end{pmatrix} = \begin{pmatrix} g_{(k)1} \\ \\ g_{(k)2} \end{pmatrix} = \begin{pmatrix} e^{\beta_0 + \beta_1 z_{(k)} + b_{(k)}} \\ \\ e^{\beta_2} \end{pmatrix}$$

The dimensions here are $p = 2, p_1 = 3, p_2 = 1, p_3 = 2$. Also, $b_{(k)} \sim \mathcal{N}(0, \sigma_b^2)$ and $\boldsymbol{\gamma}^{\mathrm{T}} =$

$(\boldsymbol{\beta}^{\mathrm{T}}, \boldsymbol{\sigma}^{\mathrm{T}}) = (\beta_0, \beta_1, \beta_2, \sigma_b^2, \sigma_\varepsilon^2)$. Matrices $\boldsymbol{F}_{(k)}$, $\boldsymbol{Z}_{(k)}$ and $\boldsymbol{H}_{(k)}$ are as follows.

$$\boldsymbol{F}_{(k)} = \left\{ \left(\mathbf{1}_{r_{j(k)}} \times \boldsymbol{f}_{j(k)}^{\mathrm{T}} \right) \big|_{\boldsymbol{\vartheta}^0} \right\}_{j=1,\ldots,n_{(k)},\,(k)\in I_{(n)}},$$

where

$$\boldsymbol{f}_{j(k)}^{\mathrm{T}} = \left(\frac{x_{j(k)}}{K_{\mathrm{m}} + x_{j(k)}},\ -\frac{V_{\max(k)} x_{j(k)}}{(K_{\mathrm{m}} + x_{j(k)})^2} \right) \big|_{\boldsymbol{\vartheta}^0} = \left(\frac{x_{j(k)}}{e^{\beta_2^0} + x_{j(k)}},\ -\frac{e^{\beta_0^0 + \beta_1^0 z_{(k)}} x_{j(k)}}{(e^{\beta_2^0} + x_{j(k)})^2} \right),$$

$$\boldsymbol{Z}_{(k)} = \begin{pmatrix} e^{\beta_0 + \beta_1 z_{(k)} + b_{(k)}} & e^{\beta_0 + \beta_1 z_{(k)} + b_{(k)}} z_{(k)} & 0 \\ 0 & 0 & e^{\beta_2} \end{pmatrix} \big|_{\boldsymbol{\vartheta}^0} = \begin{pmatrix} e^{\beta_0^0 + \beta_1^0 z_{(k)}} & e^{\beta_0^0 + \beta_1^0 z_{(k)}} z_{(k)} & 0 \\ 0 & 0 & e^{\beta_2^0} \end{pmatrix}$$

and

$$\boldsymbol{H}_{(k)} = \begin{pmatrix} e^{\beta_0 + \beta_1 z_{(k)} + b_{(k)}} \\ 0 \end{pmatrix} \big|_{\boldsymbol{\vartheta}^0} = \begin{pmatrix} e^{\beta_0^0 + \beta_1^0 z_{(k)}} \\ 0 \end{pmatrix},$$

dropping the notation $|_{\boldsymbol{\vartheta}^0}$ for notational clarity. These applied to (2.12) will give the population Fisher information matrix for our example.

3.2 Transformation

In practical situations relevant to this work it is often the case that the random errors of observations are not normally distributed. If this is a possible scenario, we propose to transform the model and adjust the optimality criterion to take into account the transformation parameter.

We apply the transform both sides method as follows

$$y_{ij(k)}^{(\lambda)} = \left\{ \eta\big(\boldsymbol{x}_{j(k)}, \boldsymbol{\beta}_{(k)}\big) \right\}^{(\lambda)} + \varepsilon_{ij(k)}, \quad i = 1, \ldots, r_{j(k)},\ j = 1, \ldots, n_{(k)},\ (k) \in I_{(n)}, \tag{3.1}$$

where (λ) indicates the Box-Cox transformation function, that is,

$$y^{(\lambda)} = \begin{cases} \frac{y^\lambda - 1}{\lambda}, & \text{when } \lambda \neq 0; \\ \log y, & \text{when } \lambda = 0. \end{cases} \tag{3.2}$$

Gilberg et al. (1999) compared estimation results for a Michaelis-Menten mixed effects model which is weighted and/or transformed on both sides by various methods. For their example the transform both sides models worked better than the non-transformed ones.

The transformation requires adjustments in the linearized model (2.8) and in the information matrix. For a given value of $\lambda \neq 0$, the derivative of $\eta^{(\lambda)}$ with respect to β_l is $\eta^{\lambda-1}\frac{\partial \eta}{\partial \beta_l}$ and similarly for the derivative with respect to $b_{(k)l}$. Hence, for $\lambda \neq 0$ we have

$$\begin{aligned} \eta^{(\lambda)}_{j(k)} &\cong \eta^{(\lambda)}_{j(k)}\big|_{\boldsymbol{\vartheta}^0} + \eta^{\lambda-1}_{j(k)}\big|_{\boldsymbol{\vartheta}^0} \left(\frac{\partial \eta_{j(k)}}{\partial \boldsymbol{\vartheta}}\big|_{\boldsymbol{\vartheta}^0}\right)^{\mathrm{T}} \left(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}^0\right) \\ &= \widetilde{\alpha}_{j(k)} + \eta^{\lambda-1}_{j(k)}\big|_{\boldsymbol{\vartheta}^0} \left(\frac{\partial \eta_{j(k)}}{\partial \boldsymbol{\beta}}\big|_{\boldsymbol{\vartheta}^0}\right)^{\mathrm{T}} \boldsymbol{\beta} + \eta^{\lambda-1}_{j(k)}\big|_{\boldsymbol{\vartheta}^0} \left(\frac{\partial \eta_{j(k)}}{\partial \boldsymbol{b}_{(k)}}\big|_{\boldsymbol{\vartheta}^0}\right)^{\mathrm{T}} \boldsymbol{b}_{(k)}, \end{aligned} \tag{3.3}$$

where

$$\widetilde{\alpha}_{j(k)} = \eta^{(\lambda)}_{j(k)}\big|_{\boldsymbol{\vartheta}^0} - \eta^{\lambda-1}_{j(k)}\big|_{\boldsymbol{\vartheta}^0} \left(\frac{\partial \eta_{j(k)}}{\partial \boldsymbol{\beta}}\big|_{\boldsymbol{\vartheta}^0}\right)^{\mathrm{T}} \boldsymbol{\beta}^0.$$

Let

$$\widetilde{\boldsymbol{f}}^{\mathrm{T}}_{j(k)} = \eta^{\lambda-1}_{j(k)}\big|_{\boldsymbol{\vartheta}^0} \boldsymbol{f}^{\mathrm{T}}_{j(k)}.$$

Then, the linearized transformed model including all observations for individual (k) has the same form as (2.11) with $\widetilde{\boldsymbol{\alpha}}_{(k)}$ replacing $\boldsymbol{\alpha}_{(k)}$ and $\widetilde{\boldsymbol{F}}_{(k)}$ replacing $\boldsymbol{F}_{(k)}$, where $\widetilde{\boldsymbol{\alpha}}_{(k)}$ is the $m_{(k)}$-dimensional vector of constants $\widetilde{\alpha}_{j(k)}$, each repeated $r_{j(k)}$ times, $\widetilde{\boldsymbol{F}}_{(k)}$ is the $(m_{(k)} \times p)$-dimensional matrix whose rows are $\widetilde{\boldsymbol{f}}^{\mathrm{T}}_{j(k)}$, each row repeated $r_{j(k)}$ times. That is,

$$\boldsymbol{y}^{(\lambda)}_{(k)} \cong \widetilde{\boldsymbol{\alpha}}_{(k)} + \widetilde{\boldsymbol{F}}_{(k)}\boldsymbol{Z}_{(k)}\boldsymbol{\beta} + \widetilde{\boldsymbol{F}}_{(k)}\boldsymbol{H}_{(k)}\boldsymbol{b}_{(k)} + \boldsymbol{\varepsilon}_{(k)}. \tag{3.4}$$

This model should be now used to obtain the information matrix $\widetilde{\boldsymbol{M}}$ which will have the same form as (2.12) with the adjusted derivatives to account for the transformation.

In our example vector $\widetilde{\boldsymbol{f}}^{\mathrm{T}}_{j(k)}$ takes the form

$$\widetilde{\boldsymbol{f}}^{\mathrm{T}}_{j(k)} = \left(\frac{\left(\eta^0_{j(k)}\right)^{\lambda}}{e^{\beta^0_0 + \beta^0_1 z_{(k)}}}, \frac{-\left(\eta^0_{j(k)}\right)^{\lambda}}{e^{\beta^0_2} + x_{j(k)}}\right),$$

where

$$\eta^0_{j(k)} = \frac{e^{\beta^0_0+\beta^0_1 z_{(k)}} x_{j(k)}}{e^{\beta^0_2} + x_{j(k)}}.$$

The transformation parameter λ is usually unknown and has to be estimated. This means that now we have a multiple objective: efficient estimation of both λ and $\boldsymbol{\gamma}$.

Since the joint estimation leads to difficulties with computation and interpretation, to find λ stabilizing the random errors we use the simple fixed effects model (Latif and Gilmour, 2012)

$$y^{(\lambda)}_{ij(k)} = \tau_{j(k)} + \delta_{ij(k)}, \quad \delta_{ij(k)} \sim \mathcal{N}(0, \sigma^2_\delta). \tag{3.5}$$

This can be considered as a simple ANOVA model, with Box-Cox transformation, with the vector $\boldsymbol{\tau}$ of treatment effects $\tau_{j(k)}$.

We denote by $\widetilde{\boldsymbol{M}}(\zeta, \boldsymbol{\tau}, \sigma^2_\delta, \lambda)$ the information matrix for all the parameters of model (3.5), that is for the treatment effects, the unknown variance and the unknown transformation parameter λ. As we are interested in efficient estimation of the transformation parameter we choose a criterion to minimize the variance of its estimator, that is

$$\psi_\lambda(\zeta, \lambda) = -\log\left[\operatorname{var}(\widehat{\lambda})\right] = -\log\left[\boldsymbol{c}^{\mathrm{T}} \widetilde{\boldsymbol{M}}(\zeta, \boldsymbol{\tau}, \sigma^2_\delta, \lambda)^{-1} \boldsymbol{c}\right], \tag{3.6}$$

where $\boldsymbol{c}^{\mathrm{T}} = (0, 0, \ldots, 0, 1)$ has dimension $\sum_{(k)\in I_{(n)}} n_{(k)} + 2$.

This is a part of the compound criterion, which also includes a criterion for efficient estimation of the parameters $\boldsymbol{\gamma}$.

That is, the compound optimality criterion function is

$$\begin{aligned}
\psi(\zeta, \boldsymbol{\gamma}, \lambda) &= \psi_{\lambda}(\zeta, \lambda) + \psi_{\boldsymbol{\gamma}}(\zeta, \boldsymbol{\gamma}) \\
&= -\log\left[\boldsymbol{c}^{\mathrm{T}} \widetilde{\boldsymbol{M}}(\zeta, \boldsymbol{\tau}, \sigma_{\delta}^2, \lambda)^{-1} \boldsymbol{c}\right] + \log\det \widetilde{\boldsymbol{M}}(\zeta, \boldsymbol{\gamma} | \lambda = \lambda^0) \\
&= \log \frac{\det \widetilde{\boldsymbol{M}}(\zeta, \boldsymbol{\gamma} | \lambda = \lambda^0)}{\boldsymbol{c}^{\mathrm{T}} \widetilde{\boldsymbol{M}}(\zeta, \boldsymbol{\tau}, \sigma_{\delta}^2, \lambda)^{-1} \boldsymbol{c}} = \log \frac{\det \widetilde{\boldsymbol{B}}(\zeta, \boldsymbol{\beta} | \lambda = \lambda^0) \det \widetilde{\boldsymbol{C}}(\zeta, \boldsymbol{\sigma} | \lambda = \lambda^0)}{\boldsymbol{c}^{\mathrm{T}} \widetilde{\boldsymbol{M}}(\zeta, \boldsymbol{\tau}, \sigma_{\delta}^2, \lambda)^{-1} \boldsymbol{c}} \\
&= \log \frac{\det \widetilde{\boldsymbol{B}}(\zeta, \boldsymbol{\beta} | \lambda = \lambda^0)}{\sqrt{\boldsymbol{c}^{\mathrm{T}} \widetilde{\boldsymbol{M}}(\zeta, \boldsymbol{\tau}, \sigma_{\delta}^2, \lambda)^{-1} \boldsymbol{c}}} + \log \frac{\det \widetilde{\boldsymbol{C}}(\zeta, \boldsymbol{\sigma} | \lambda = \lambda^0)}{\sqrt{\boldsymbol{c}^{\mathrm{T}} \widetilde{\boldsymbol{M}}(\zeta, \boldsymbol{\tau}, \sigma_{\delta}^2, \lambda)^{-1} \boldsymbol{c}}}
\end{aligned} \tag{3.7}$$

3.3 Optimal design search algorithm

The typical method of finding optimal designs for moderate to large run sizes is to seek continuous optimal designs and round them to the run size required. However, in this case this is not possible since the optimality criterion is not proportional to the run size N. This is a feature of mixed models where we are interested in the variance components. Instead, we search for exact locally optimal designs under the criterion $\psi(\zeta, \boldsymbol{\gamma}, \lambda)$, given in (3.7), for different run sizes, using the Fedorov exchange algorithm.

For each run of the experiment, we must choose an HLM, which implies a choice of a covariate value and of a substrate concentration. Since the substrate concentrations are on a continuous scale, we reduce the problem by choosing values from a candidate set. In practice, this candidate set should be spaced out according to what could be expected to be practically different concentrations. In the following illustration, we used a discrete subset of concentrations from the interval [0.3, 50] and the set of 47 HLMs, which were used to obtain the rich design.

We start with a candidate set of treatments which is obtained by all possible combinations of selected concentration levels and 47 HLMs.

A randomly (with replacement) selected N treatments from the candidate set is considered as the initial design, which is then updated using the exchange algorithm. To compete with the initial design, a competing design is obtained by interchanging the first treatment of the initial design by a treatment of the candidate design. The initial and competing designs are compared with respect to the design criterion $\psi(\zeta, \boldsymbol{\gamma}, \lambda)$ and the competing design is considered as the current best design (*Case I*) if it corresponds to the higher design criterion value than that of the initial design, otherwise, the initial design is considered as the current best design (*Case II*). To compete with the current best design, the new competing design is obtained either by replacing the second treatment (for *Case I*) or the first treatment of the current best design (for *Case II*) with a new treatment of the candidate set. This search procedure is continued while there is a competing design with higher design criterion value than the current best design, otherwise, the current best design is considered as the optimal design. For exact design, there is no guarantee that the exchange algorithm leads to the global optimal design, so in practice it is preferable to repeat the search procedure for a number of different initial designs to obtain the optimal design.

3.4 Numerical results

The surface $\eta^{(\lambda)}$ as a function of the concentration and of the covariate (activity of enzyme 2B6, standardized) is shown in Figure 4.3. We can see that as the covariate value increases so does the asymptote of the response, that is the value of the parameter $V_{\max}$. The covariate explains some of the variability in the response.

For the single covariate, we obtained designs for $N = 50, 100, 150, 200, 250$ and compared them here with the design used for the rich data, which had $2N = 846$, being all combinations of 47 HLMs with the 9 concentrations

$$\{0.3125, 0.625, 1.25, 2.5, 5.0, 10.0, 20.0, 40.0, 50.0\}$$

repeated twice. We used estimates of the parameters obtained from the rich design as prior values for finding the optimum designs which are $\widehat{V}_{\max} = \exp\{\widehat{\beta}_0 + \widehat{\beta}_1 z\}, \widehat{\beta}_0 = -3.155, \widehat{\beta}_1 = 0.744, \widehat{K}_m = \exp\{\widehat{\beta}_2\}, \widehat{\beta}_2 = 0.5463, \widehat{\sigma}_b^2 = 0.373, \widehat{\sigma}_\varepsilon^2 = 0.059$. For our optimization we used a regular grid of concentrations in the region of $[0.3,\ 50]$ in steps of 0.1 after refining the more course grid and the same set of HLMs as in the rich design. The points selected by the algorithm for $N = 50, 100, 150, 200, 250$ are shown in Figure 4.4 together with the rich design points.

Some patterns, with minor variations, are clear. The structure of these optimum designs is very different from the typical set-up used in practice. Only four concentrations are used, apart from the design for $N = 250$ where there is one point chosen at a fifth concentration. The lowest possible concentration is combined with the HLMs with the lowest covariate values. The second and third chosen concentration values, close to the prior value assumed for K_m, are combined with several HLMs with the highest values of the covariate. Finally, the highest possible concentration is used with many HLMs, starting with those at each extreme and working towards the middle as the run size increases.

The marginal optimal designs shown in Figure 4.5 indicate the number of replications of the support points for the considered cases of N. Both concentration and the activity values on the borders of the design region are chosen much more often than the internal points.

In particular, it is interesting to observe two characteristics of the marginal design for the choice of covariate values: first, that all the available values are chosen and second that the weight

is mostly put on the border values. The former characteristic is related to the fact that the response depends on the covariate z via the parameter $V_{\max}$ and latter is related to the fact that the response depends linearly on $V_{\max}$ and so the end points of the parameter region will be most informative.

Note that this shows only the covariate value and does not distinguish between different HLMs with the same value, some of which existed in the data set used. The designs are given in full in Tables 1–5 in the Web Appendix (on-line supplementary material).

Furthermore, looking at the marginal design for the concentration values, we see that a large weight is put on the biggest concentration value, which gives information on the $V_{\max}$ parameter, which is considered as random. On the other hand, parameter K_m is assumed to be constant and so only some of the covariate values are combined with the concentrations which give information on this parameter.

Relative efficiency measure

We use a relative efficiency measure for a population design ζ compared with another design, for example a standard design used in practice. In our case we compare the optimum design $\zeta^\star$ and the rich design ζ^{rich} in a space of population designs Δ, that is

$$Eff = \left\{ \frac{\det \boldsymbol{M}(\zeta^\star, \boldsymbol{\gamma}^0)}{\det \boldsymbol{M}(\zeta^{rich}, \boldsymbol{\gamma}^0)} \right\}^{\frac{1}{p}}, \tag{3.8}$$

where both designs are evaluated at some prior parameter values $\boldsymbol{\gamma}^0$.

The relative efficiency of the optimal designs obtained for various values of N to the rich design are shown in Figure 4.6. We also calculated the relative efficiency for the optimal

design replicated 2, 3, 4 and 5 times. We can see that two replications of a 250-point optimal design is better then the rich design (which is a 423 point design replicated twice). This means that we could use 500 runs of the experiment rather than 846 to obtain the same efficiency of estimation if we do the design optimization including a covariate which explains a significant part of the response variability.

The 50-point design, even when replicated 5 times, is worse than the rich design. This is because it does not carry enough information on the variable parameter $V_{\max}$; not all available covariates are combined with the largest value of the concentration. On the other hand, the 100-point design replicated five times is almost equally efficient as the 250-point design replicated twice. In fact, in both designs all covariate values are combined with the largest concentration value. Furthermore, if we could use similar number of runs as in the rich design, then for example, the 200-point design replicated four times would give a higher efficiency, still with slightly fewer runs.

We performed some simulations to examine the precision of parameter estimation. The results are given in Table 1.

The table gives values of the bias of the estimates and their standard deviations, estimated from the simulations, and the mean estimated standard errors, as well as the ratio of the last two, calculated for simulated observations based on four designs: twice replicated 150, 200, 250 point design and the rich design. For all the cases there were 10000 simulations and the estimates of the parameters from the real data set were used as prior values. As we can see there is not much difference in the precision of estimation, but the rich design gives us slightly smaller bias in estimating λ.

4 Conclusions

In this paper we presented a new method of designing experiments for non-linear mixed effects models where covariates are design variables combined with an explanatory variable or another treatment factor. We gave forms of the information matrix and of the D-optimality criterion for such a case and also expanded the criterion to allow for transformation of the response in case of non-normal random errors.

The theory is exemplified by the real-life data on the Human Liver Microsomes with various enzyme activities. Several optimized designs are presented and their properties studied. We observe that substantial savings can be achieved by using such designs. The new designs can be equally efficient using less experimental material than is needed in standard practice or, if a similar experimental effort is allowed, then we can achieve higher efficiency.

Furthermore, using mixed-effects models with covariates we also get information on the population variability and are able to asses the variation of the response due to the covariates. This can be particularly useful for stratification of the population and also for personalizing the treatments.

In this paper we assume that we know which covariate is important for the response and we chose values of this covariate to optimize the design. More work is needed to further develop the methodology to optimally choose among several covariates during the stage of designing experiments.

Web Appendix referenced in Section 3.4 is available with this paper at the Biometrics website on Wiley Online Library.

Acknowledgments

This research was partly supported by the UK Engineering and Physical Sciences Research Council (EPSRC) grant EP/C54171/1 and partly by the INI where it was initiated and largely executed.

References

Belle, D., B. Ring, S. Allerheiligen, M. Heathman, L. O'Brian, V. Sinha, L. Roskos, and S. Wrighton (2000). A population approach to enzyme characterization and identification: Application to Phenacetin O-Deethylation. *Pharmaceutical Research 17*(12), 1531–1536.

Denti, P., A. Bertoldo, P. Vicini, and C. Cobelli (2010). IVGTT glucose minimal model covariate selection by nonlinear mixed-effects approach. *American Journal of Physiological Endocrinol Metabolism 298*, E950–E960.

Ding, A. and H. Wu (2001). Assessing aniviral potency of anti-HIV therapies in vivo by comparing viral decay rates in viral dynamic models. *Biostatistics 2*(1).

Fitzmaurice, G., M. Davidian, G. Verbeke, and G. Molenberghs (2009). *Longitudinal Data Analysis*. Boca Raton: Chapman and Hall.

Gilberg, F., W. Urfer, and L. Edler (1999). Heteroscedastic nonlinear regression models with random effects and their application to enzyme kinetic data. *Biometrical Journal 41*(5).

Hasler, J., R. Estabrook, M. Murray, I. Pikuleva, M. Waterman, J. Capdevila, V. Holla, C. Helvig, J. Falck, and G. Farrell (1999). Human cytochromes P450. *Mol Aspects Med 20*, 1–137.

Latif, M. and S. Gilmour (2012). Transform-both-sides nonlinear models for randomized experiments. *Submitted*.

Lindstrom, M. and D. Bates (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics 46*, 673–687.

Retout, S., E. Comets, A. Samson, and F. Mentré (2007). Design in nonlinear mixed effects models: Optimization using the Fedorov-Wynn algorithm and power of the Wald test for binary covariates. *Statistics in Medicine 26*, 5162–5179.

Wu, H. and A. Ding (2002). Design for viral dynamic studies for efficiently assessing potency of anti-HIV therapies in AIDS clinical trials. *Biometrical Journal 44*(2).

Wu, H. and L. Wu (2002). Identification of significant host factors for hiv dynamics modelled by non-linear mixed effects models. *Statistics in Medicine 21*, 753–771.

Figures and Tables

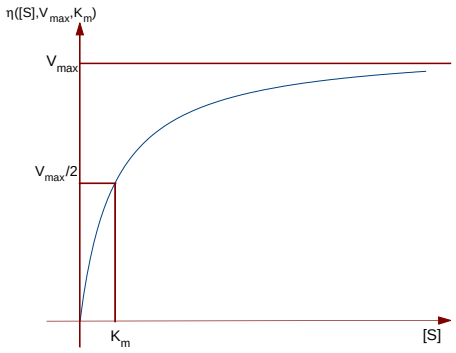


Figure 4.1: Michaelis-Menten Model. The response function: $\eta\big([S]; V_{\max}, K_{\mathrm{m}}\big)$.

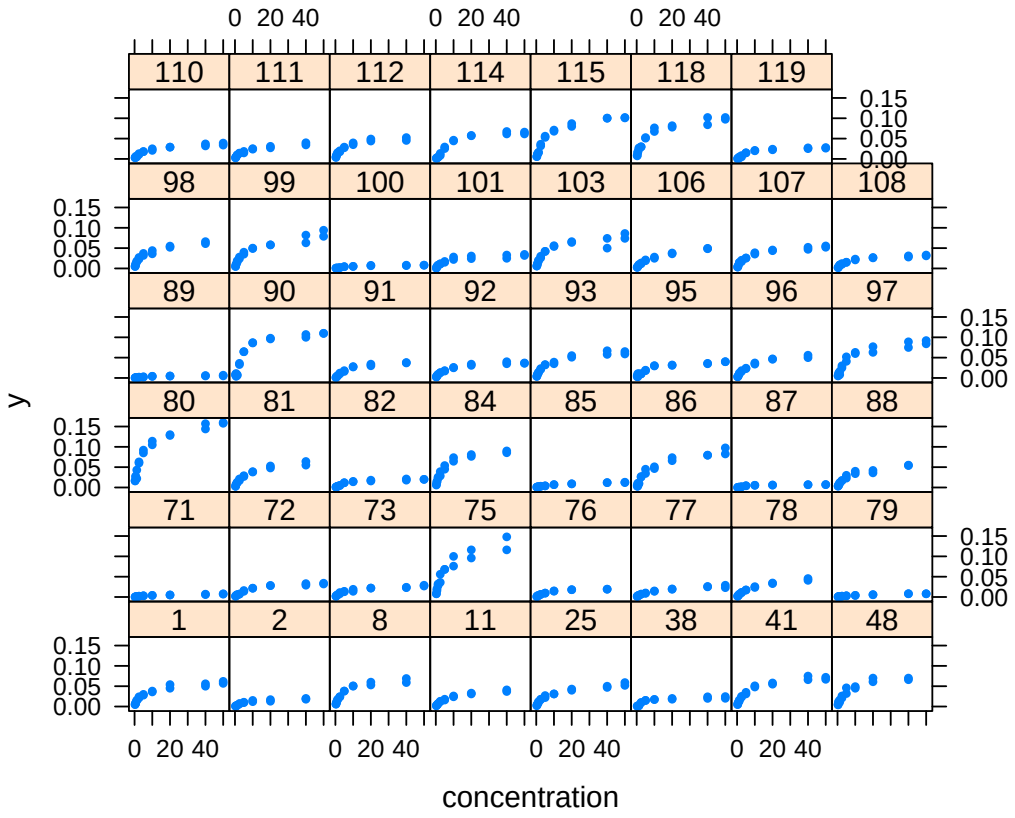


Figure 4.2: Observations of reaction rates for a substrate in 47 HLM preparations.

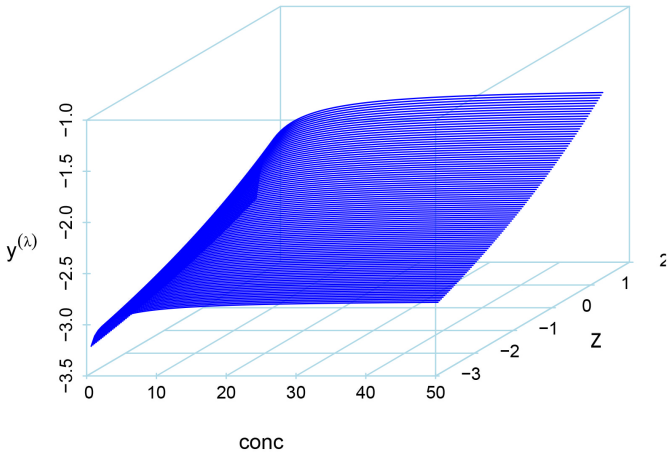
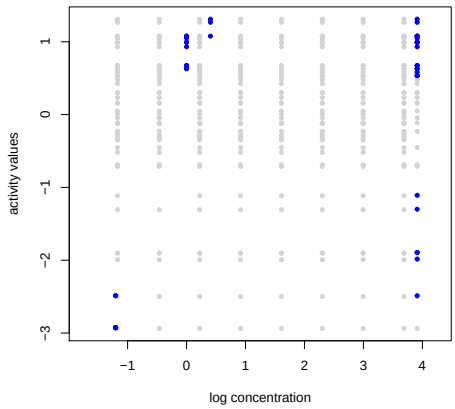
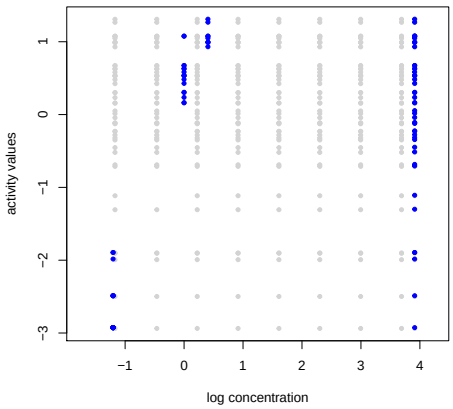


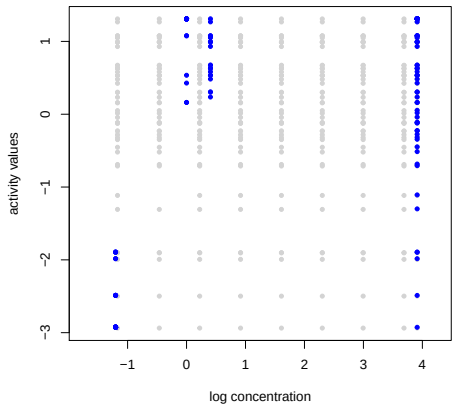
Figure 4.3: The transformed model surface fitted to the rich data set.



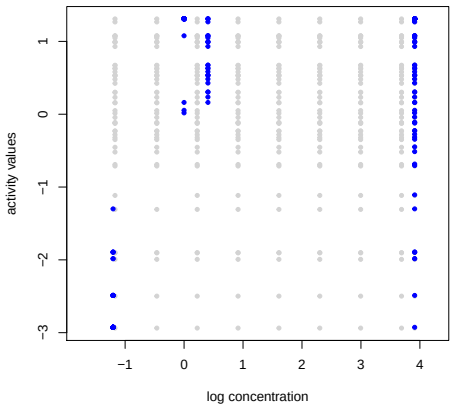
(a) $N = 50$



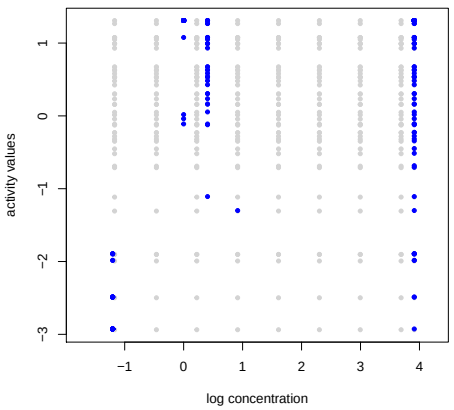
(b) $N = 100$



(c) $N = 150$

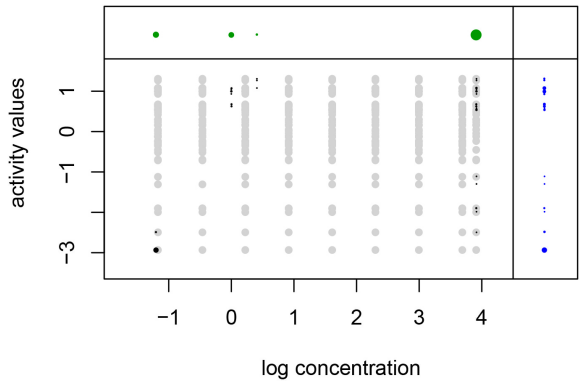


(d) $N = 250$

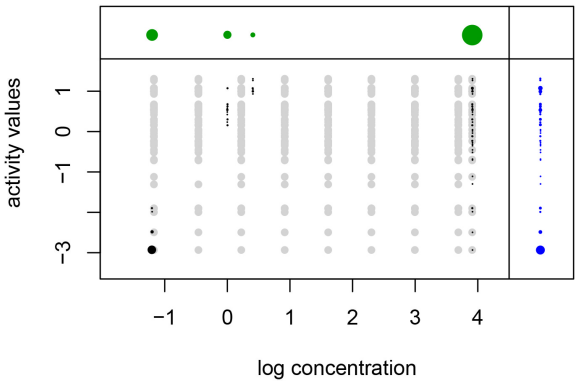


(e) $N = 250$

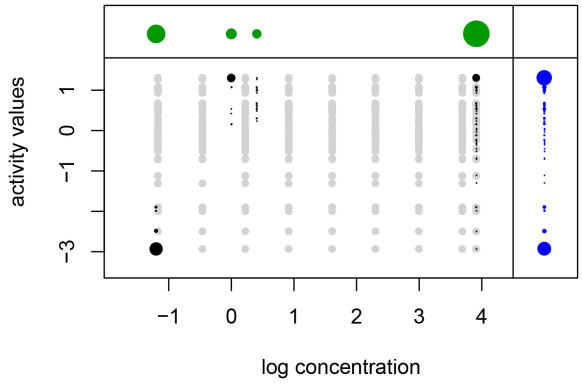
Figure 4.4: Optimum concentration levels and activity values obtained for five different values of the total number of observations N allowed (grey points indicate rich design).



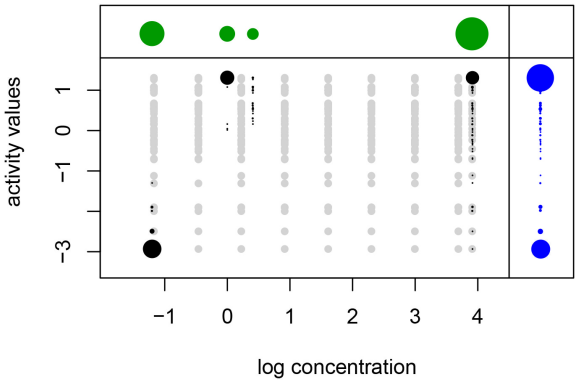
(a) $N = 50$



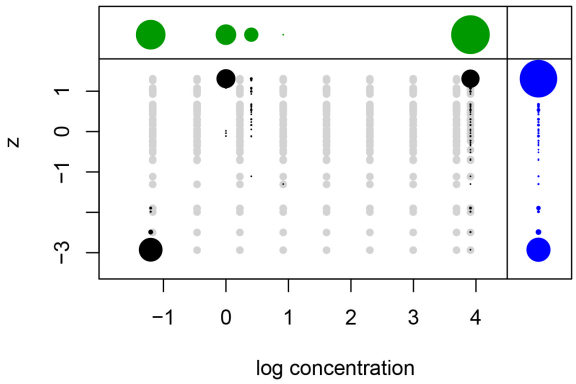
(b) $N = 100$



(c) $N = 150$



(d) $N = 250$



(e) $N = 250$

Figure 4.5: Support points of the optimal designs and the marginal designs.

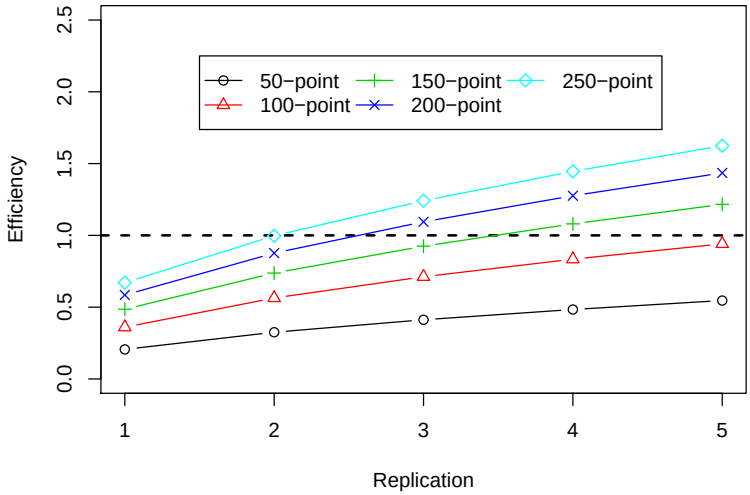


Figure 4.6: Relative efficiencies of the optimal designs, and replicates of these designs, with respect to the rich design.

Table 1: Simulation results

$2N$		β_0	β_1	β_2	σ_b	σ	λ
300	bias	0.0044	-0.0005	-0.0009	-0.0083	-0.0076	0.0292
	sd	0.0556	0.0320	0.0567	0.0406	0.0057	0.0198
	se	0.0556	0.0312	0.0560			0.0174
	sd/se	1.0006	1.0246	1.0128			1.1383
400	bias	0.0043	-0.0001	-0.0011	-0.0100	-0.0054	0.0203
	sd	0.0556	0.0278	0.0570	0.0410	0.0050	0.0171
	se	0.0552	0.0264	0.0554			0.0149
	sd/se	1.0076	1.0534	1.0293			1.1462
500	bias	0.0027	-0.0005	-0.0010	-0.0096	-0.0044	0.0163
	sd	0.0559	0.0237	0.0560	0.0405	0.0044	0.0150
	se	0.0550	0.0233	0.0551			0.0133
	sd/se	1.0157	1.0206	1.0152			1.1239
Rich	bias	0.0006	0.0002	-0.0005	-0.0102	0.0009	-0.0016
(846)	sd	0.0551	0.0223	0.0550	0.0388	0.0072	0.0270
	se	0.0543	0.0222	0.0542			0.0193
	sd/se	1.0162	1.0073	1.0145			1.3977
Data	est	-3.1547	0.7439	0.5463	0.3730	0.0591	0.2800
	se	0.0556	0.0555	0.0223			0.0192