Nonlinear Mixed Effects Models: Approximations of the Fisher Information and Design

T. Mielke

Abstract The problem of the missing closed form representation of the probability density of the observations in nonlinear mixed effects models carries forward to the calculation of the Fisher information. Linearizations of the response function are often applied for approximating the underlying statistical model. The impact of different linearizations on the design of experiments will be briefly discussed in this article and an alternative motivation for an approximation of the Fisher information will be presented. The different results will be illustrated in the example of a simple population pharmacokinetic model.

1 Introduction

Mixed effects models are often applied for the analysis of grouped data. The difference of observations of different groups are in these models assumed to depend on observation errors and group wise varying parameter vectors. Specially in pharmacological studies each individual can be interpreted as a single group and insight in the population behavior can be obtained by modeling the individual parameter vectors as identically distributed random variables. Nonlinear mixed effects models are in the literature often described under the assumption of normally distributed random effects. Estimators based on weighted sums of squares (e.g. Pinheiro and Bates (2000)) or on stochastic approximations (e.g. Kuhn and Lavielle (2001)) are proposed for the analysis of the population behavior, as no closed form representation of the likelihood function in these models exists. The used estimators are typically assumed to be consistent with an information matrix behaving as in linear mixed effects models or nonlinear models with heteroscedastic normal errors. However, the limitations of the stochastic behavior of some estimators were discussed on the monoexponential model by Demidenko (2005). Linearized models build the foundation for experimental designs in nonlinear mixed effects models. The different linearization approaches yield distinct information matrices, such that the approximations might have a big influence on the design of studies, as illustrated by

---

T. Mielke , Otto-von-Guericke University Magdeburg, e-mail: tobias.mielke@ovgu.de

This work was supported by the BMBF grant SKAVOE 03SCPAB3.

2000 Mathematics Subject Classification: 62K05, 62J02.

Keywords: Fisher information, mixed effects models, nonlinear models, optimal design

Mielke and Schwabe (2010) in a simple example.
After introducing the mixed effects model in the second section, an alternative motivation of an approximation of the Fisher information, based on approximations of conditional moments, is presented in the third section. Designs in population studies are briefly defined in the fourth section, before an example of a pharmacokinetic model shows the influence of different approximations on the design in the fifth section.

2 Mixed effects models

In the mixed effects models, the j-th observation of the i-th individual under experimental setting $x_{ij} \in \mathscr{X}$ is with a response function η described by

$$Y_{ij} = \eta(\beta_i, x_{ij}) + \varepsilon_{ij},$$

with an individual parameter vector $\beta_i \in \mathbb{R}^p$ and real valued observation errors ε_{ij}. The response function η is assumed to be differentiable in β_i and continuous on the compact design space $\mathscr{X}$. The m_i dimensional observation vector Y_i is for given individual parameter vectors β_i with the exact individual design

$$\xi_i = (x_{i1}, ..., x_{im_i}) \text{ where } x_{ij} \in \mathscr{X},\ j = 1, ..., m_i$$

and the vector valued response function

$$\eta(\beta_i, \xi_i) := \big(\eta(\beta_i, x_{i1}), ..., \eta(\beta_i, x_{im_i})\big)^T,$$

assumed to be normally distributed:

$$Y_i \sim N\big(\eta(\beta_i, \xi_i), \sigma^2 I_{m_i}\big).$$

The inter-individual variation is induced by the individual wise varying parameter vectors β_i which are assumed to be realizations of normally distributed random variables:

$$\beta_i \sim N(\beta, \sigma^2 D).$$

The observation errors $\varepsilon_i = (\varepsilon_{i1}, ..., \varepsilon_{im_i})$ and individual parameter vectors β_i are assumed to be independent and observations of different individuals are considered to be stochastically independent as well. Throughout this article we assume the variance parameter $\theta = (\sigma^2, D)$ to be known and the matrix D to be positive definite.
The likelihood of observations y_i results in integral form in

$$L(\beta; y_i, \theta) := f_{Y_i}(y_i) = \int_{\mathbb{R}^p} \phi_{Y_i|\beta_i}(y_i)\phi_{\beta_i}(\beta_i)d\beta_i,$$

where the influence of the parameters β and θ on the likelihood is contained in the normal densities

$$\phi_{Y_i|\beta_i}(y_i) = \sqrt{2\pi\sigma^2}^{\,-m_i} \exp\Big[-\frac{1}{2\sigma^2}\big(y_i - \eta(\beta_i, \xi_i)\big)^T \big(y_i - \eta(\beta_i, \xi_i)\big)\Big]$$

$$\phi_{\beta_i}(\beta_i) = \sqrt{2\pi\sigma^2}^{\,-p}\sqrt{\| D \|}^{\,-1}\exp[-\frac{1}{2\sigma^2}(\beta_i-\beta)^T D^{-1}(\beta_i-\beta)],$$

with $\| \cdot \|$ denoting the determinant. For nonlinear response functions η, the likelihood function cannot be written in a closed from. To circumvent this problem, the model is transformed by a linearization of the response function as described by Retout and Mentré (2003). With a design matrix defined as

$$F_\beta := \frac{\partial \eta(\beta_i,\xi_i)}{\partial \beta_i^T}|_{\beta_i=\beta}$$

follows for a linearization in the true population mean β and under the assumption of negligible linearization errors:

$$Y_i = \eta(\beta,\xi_i) + F_\beta(\beta_i-\beta) + \varepsilon_i.$$

The distribution assumptions on β_i and the observation error vector ε_i imply normally distributed vectors of observations with heteroscedastic errors:

$$Y_i \sim N\big(\eta(\beta,\xi_i), \sigma^2[I_{m_i} + F_\beta D F_\beta^T]\big).$$

Under the assumption of negligible linearization errors, a linear mixed effects model is obtained, when alternatively linearizing the response function in some point β_0:

$$Y_i \sim N\big(\eta(\beta_0,\xi_i) + F_{\beta_0}(\beta-\beta_0), \sigma^2[I_{m_i} + F_{\beta_0} D F_{\beta_0}^T]\big).$$

Both linearizations yield for linear response functions η the true linear mixed effects model, as $F_\beta = F_{\beta_0}$ is then independent of β. Note that the information matrices in heteroscedastic normal and linear mixed effects models are distinct. The linearization in the true population mean under the assumption of negligible linearization errors yields with $V_\beta := I_{m_i} + F_\beta D F_\beta^T$:

$$\mathbf{M}_{1,\beta}(\xi_i) := \frac{1}{\sigma^2} F_\beta^T V_\beta^{-1} F_\beta + \frac{1}{2} S,$$

where $S \geq 0$, with

$$S_{j,k} = \mathrm{tr}\,[V_\beta^{-1}\frac{\partial V_\beta}{\partial \beta_j} V_\beta^{-1}\frac{\partial V_\beta}{\partial \beta_k}],\ \ j,k = 1,...,p,$$

as approximation of the Fisher information, whereas the information resulting from a linearization in a point β_0 is of the form

$$\mathbf{M}_{2,\beta}(\xi_i) := \frac{1}{\sigma^2} F_{\beta_0}^T V_{\beta_0}^{-1} F_{\beta_0},$$

such that specially for $\beta_0 = \beta$ additional information is drawn by $\mathbf{M}_{1,\beta}$ from the variance structure of the observations. This difference in the information matrices was discussed by Mielke and Schwabe (2010) in an example with the result that the additional matrix term S in some situations might misleadingly generate information.

3 Approximation of the Fisher information

Alternatively the Fisher information can be directly computed as the covariance of the score function. The score function is obtained for a positive definite inter-individual variance matrix $\sigma^2 D$ as in Mielke (2011) by

$$\frac{\partial l(\beta; y_i, \theta)}{\partial \beta} = \frac{1}{\sigma^2} D^{-1} \big(E(\beta_i | Y_i = y_i) - \beta \big),$$

with the log-likelihood function $l(\beta; y_i, \theta) := \log \big(L(\beta; y_i, \theta) \big)$. The Fisher information can hence be written in the form

$$\begin{aligned} \mathfrak{M}_\beta(\xi_i) &= E\left(\frac{\partial l(\beta; Y_i, \theta)}{\partial \beta} \frac{\partial l(\beta; Y_i, \theta)}{\partial \beta^T} \right) \\ &= \frac{1}{\sigma^2} D^{-1} - \frac{1}{\sigma^4} D^{-1} E\big(Var(\beta_i | Y_i) \big) D^{-1}, \end{aligned}$$

since with the distributional assumptions on the individual parameter vector β_i follows

$$Var(\beta_i) = E\big(Var(\beta_i|Y_i)\big) + Var\big(E(\beta_i|Y_i)\big) = \sigma^2 D,$$

such that apporixmations of the Fisher information result as approximations of the expectation of the conditional variance. For limited numbers of possible experimental settings, small individual sample sizes m_i and low dimensional parameter vectors β, the Fisher information can then be approximated using quadrature rules or Monte-Carlo methods. Generally the computational burden for approximating the dependence of the Fisher information on the experimental settings will already for relatively small sample sizes and small dimensions of β be very high, such that analytical approximations are of interest. Tierney and Kadane (1986) propose fully exponential Laplace approximations for approximating posterior moments. Therefor a second order Taylor approach in the minimizing argument β_i^* of the penalized least squares term

$$\tilde{l}(\beta_i; y_i, \beta, \theta) := \big(y_i - \eta(\beta_i, \xi_i)\big)^T \big(y_i - \eta(\beta_i, \xi_i)\big) + (\beta_i - \beta)^T D^{-1} (\beta_i - \beta)$$

is applied for approximating the occurring integrals. The nonlinear response function η implies a nonlinear dependence of the support point β_i^* of the Taylor approach on the observation vector y_i, such that approximations of the expectation of the conditional variance cannot be obtained without yet another level of approximations. Similarly Mielke (2011) suggests the approximation of the conditional density of β_i for given observation y_i

$$f_{\beta_i|Y_i=y_i}(\beta_i) := \frac{\phi_{Y_i|\beta_i}(y_i)\phi_{\beta_i}(\beta_i)}{f_{Y_i}(y_i)}$$

by approximations to the denominator integral and an according approximation to the numerator with a first order Taylor approach in the response function η. In dependence of the support point $\hat{\beta}$ of the Taylor approach, the resulting density is then approximated by a normal density:

$$\beta_i|_{Y_i=y_i} \overset{app.}{\sim} N\big(\mu(y_i, \hat{\beta}_i, \beta), \sigma^2 M_{\hat{\beta}_i}^{-1}\big), \text{ with}$$

$$\begin{aligned}\mu(y_i,\hat{\beta}_i,\beta) &:= M_{\hat{\beta}_i}^{-1}\left(F_{\hat{\beta}_i}^T[y_i-\eta(\hat{\beta}_i,\xi_i)+F_{\hat{\beta}_i}\hat{\beta}_i]+D^{-1}\beta\right) \text{ and}\\ M_{\hat{\beta}_i} &:= F_{\hat{\beta}_i}^T F_{\hat{\beta}_i}+D^{-1}.\end{aligned}$$

The big benefit of this approach is that approximations of the Fisher information in nonlinear mixed effects models can be directly deduced, as the approximation of the conditional variance not necessarily depends on the observations y_i. The Fisher information is in dependence of the support point of $\hat{\beta}$ with this approach approximated by

$$\mathbf{M}_{2,\beta}(\xi_i) := \frac{1}{\sigma^2}F_{\hat{\beta}}^T V_{\hat{\beta}}^{-1}F_{\hat{\beta}},$$

what corresponds for $\hat{\beta}=\beta_0$ to the linear mixed effects model approximation.

A more refined approximation of the information might however be obtained by taking the distribution of the observations into account:

$$\begin{aligned}E(Var(\beta_i|Y_i)) &= \int_{\mathbb{R}^{m_i}} Var(\beta_i|Y_i=y_i)\int_{\mathbb{R}^p}\phi_{Y_i|\beta_i}(y_i)\phi_{\beta_i}(\beta_i)d\beta_i dy_i\\ &= \int_{\mathbb{R}^p}\int_{\mathbb{R}^{m_i}} Var(\beta_i|Y_i=y_i)\phi_{Y_i|\beta_i}(y_i)\phi_{\beta_i}(\beta_i)dy_i d\beta_i\\ &\approx \int_{\mathbb{R}^p}\sigma^2 M_{\beta_i}^{-1}\phi_{\beta_i}(\beta_i)d\beta_i\\ &= \sigma^2 E(M_{\beta_i}^{-1}).\end{aligned}$$

where the approximation holds by the argument, that the solution of the penalized least squares problem should be not too far located from the true individual parameter vector, which is here β_i. Note however, that for this approximation the existence of the expectation has to be guaranteed. With similar transformations as for $\mathbf{M}_{2,\beta}$ the Fisher information can then be approximated by

$$\mathbf{M}_{3,\beta}(\xi_i) := \frac{1}{\sigma^2}E(F_{\beta_i}^T V_{\beta_i}^{-1}F_{\beta_i}).$$

Unfortunately this approximation generally cannot be written in a closed form, such that the expectation has to be calculated numerically. Note that the true Fisher information matrix is obtained for all presented approximations in the case of linear response functions η.

4 Design

The two stages of the mixed effects models carry forward to the design. Individual designs

$$\xi_i := (x_{i1},\ldots,x_{im_i}) \text{ with } x_{ij}\in\mathscr{X}$$

describe the experimental settings for individuals, whereas the population designs summarize the proportions ω_i of individual designs ξ_i in the population:

$$\zeta := (\xi_1, ..., \xi_k, \omega_1, ..., \omega_k) \text{ with } \omega_i \geq 0,\ \sum_{i=1}^{k-1} \omega_i \leq 1 \ \text{ and } \ \omega_k = 1 - \sum_{i=1}^{k-1} \omega_i.$$

We here assume the numbers of observations per individual to be identical, such that $m_i = m$ for all $i = 1, ..., k$. Population designs ζ can hence be interpreted as approximate designs on the m-dimensional design space $\mathscr{X}^m$. The normalized population Fisher information matrix is with the independence of the observations of different individuals obtained as the weighted sum of the individual information matrices:

$$\mathfrak{M}_{pop,\beta}(\zeta) = \sum_{i=1}^{k} \omega_i \mathfrak{M}_\beta(\xi_i).$$

Target of the design optimization is the minimization of some real valued design criteria of the information matrix $\mathfrak{M}_{pop,\beta}$ with respect to populations designs ζ. We restrict ourselves in this article to the D-optimality criterion:

$$\Phi_D(\zeta) := -\log(\| \mathfrak{M}_{pop,\beta}(\zeta) \|),$$

as the content of the confidence ellipsoid for β is inverse proportional to the determinant of the information matrix. Results for other optimality criteria can be similarly deduced. The design optimization and the verification of optimal designs can be conducted with applications of Fedorov's equivalence theorem for designs of experiments in the case of simultaneous observations of several quantities (Fedorov (1972),p.211), yielding the sensitivity function

$$g_\zeta(\xi) := \text{tr}\, [\mathfrak{M}^{-1}_{pop,\beta}(\zeta)\mathfrak{M}_\beta(\xi)].$$

For the estimation of the p dimensional parameter vector β, a design ζ^* is hence D-optimal, if and only if

$$g_{\zeta^*}(\xi) \leq p \ \ \forall \xi \in \mathscr{X}^m$$

in the case of m observations per individual.

5 Example

We consider a one compartment model with first order absorption and the model structure as in Schmelter (2007), but with one observation at the time $x_i \in \mathscr{X} = [0.1, 24]$ per individual only:

$$Y_i = \frac{\beta_{1,i}}{\beta_{3,i}\beta_{1,i} - \beta_{2,i}} [\exp(-\frac{\beta_{2,i}}{\beta_{3,i}} x_i) - \exp(-\beta_{1,i} x_i)] \exp(\varepsilon_i).$$

Numbers from some previous experiments were used for planning purposes for the population location parameter:

$$\beta = (\beta_1, \beta_2, \beta_3) = (0.61, 25, 88).$$

The random individual effects are here considered to enter the individual parameter vectors proportionally:

$$\beta_{j,i} = \beta_j \exp(b_{j,i}),\ j = 1,2,3 \ \text{ with } (b_{1,i}, b_{2,i}, b_{3,i}) \overset{iid}{\sim} N_3(0, \sigma^2 D),$$

with a known diagonal variance matrix $D = diag(89.3, 12.5, 9.0)$ and the observation errors ε_i are assumed to be normally distributed with a variance $\sigma^2 = 0.01$. The model can then be easily transformed in a nonlinear mixed effects model as introduced in the second section.

The optimal designs based on the approximations $\mathbf{M}_{1,\beta}$ and $\mathbf{M}_{2,\beta}$ were calculated with the use of the equivalence theorem:

$$\zeta_1^* = \begin{pmatrix} (3.10) & (5.18) & (24.00) \\ 0.61 & 0.08 & 0.31 \end{pmatrix} \text{ for } \mathbf{M}_{1,\beta} \text{ and}$$

$$\zeta_2^* = \begin{pmatrix} (0.10) & (4.18) & (24.00) \\ 0.33 & 0.33 & 0.33 \end{pmatrix} \text{ for } \mathbf{M}_{2,\beta}.$$

Simulations were applied in order to compute optimal designs for the approximation $\mathbf{M}_{3,\beta}$ and for the Fisher information $\mathfrak{M}_\beta$ and the dependences of the components of the information matrices on the experimental settings were approximated by using polynomial splines. Based on these approximations, the D-optimal designs:

$$\zeta_3^* = \begin{pmatrix} (2.85) & (24.00) \\ 0.59 & 0.41 \end{pmatrix} \text{ for } \mathbf{M}_{3,\beta} \text{ and}$$

$$\zeta_F^* = \begin{pmatrix} (2.32) & (6.40) & (24.00) \\ 0.61 & 0.01 & 0.38 \end{pmatrix} \text{ for } \mathfrak{M}_\beta$$

were obtained. The uniqueness of the D-optimal designs follows for the considered approximations with the structure of the sensitivity functions, which are illustrated in Figure1. Table 1 shows the efficiency of the proposed designs for the different approximations. The efficiency of designs is here defined by the terms

$$\delta_{\zeta_i^*}(\zeta) := \left(\frac{\| \mathbf{M}_{i,\beta}(\zeta) \|}{\| \mathbf{M}_{i,\beta}(\zeta_i^*) \|} \right)^{\frac{1}{3}},\ i = 1,2,3; \quad \text{and} \quad \delta_{\zeta_F^*}(\zeta) := \left(\frac{\| \mathfrak{M}_\beta(\zeta) \|}{\| \mathfrak{M}_\beta(\zeta_F^*) \|} \right)^{\frac{1}{3}}.$$

The efficiency of all proposed designs is in relation to the simulation based approximation of the Fisher information relatively high. The similarity of the designs ζ_F^* for the Fisher information and ζ_3^* for the approximation $\mathbf{M}_{3,\beta}$ causes the high efficiency of 0.95. The efficiency of designs with respect to the often used linear mixed effects approximation $\mathbf{M}_{2,\beta}$ of the Fisher information is of special interest. The designs ζ_1^*, ζ_3^* and ζ_F^* are here relatively inefficient, what might be caused by the weights of the designs, and the missing observations on the left border of the design space. Note that the number of support points of the design ζ_3^* is smaller than the number of parameters of interest, such that the corresponding information matrix in terms of the linear mixed effects approximation is singular. The additional information of the approximation $\mathbf{M}_{1,\beta}$ was mentioned by Mielke and Schwabe (2010) and can be seen in the here presented example as well. The D-optimality criterion takes for the optimal design in the approximation $\mathbf{M}_{1,\beta}$ with respect to the different approximations the values

$$\Phi_{D;1}(\zeta_1) = 11.81 \text{ for } \mathbf{M}_{1,\beta}; \ \ \Phi_{D;2}(\zeta_1) = 15.42 \text{ for } \mathbf{M}_{2,\beta};$$

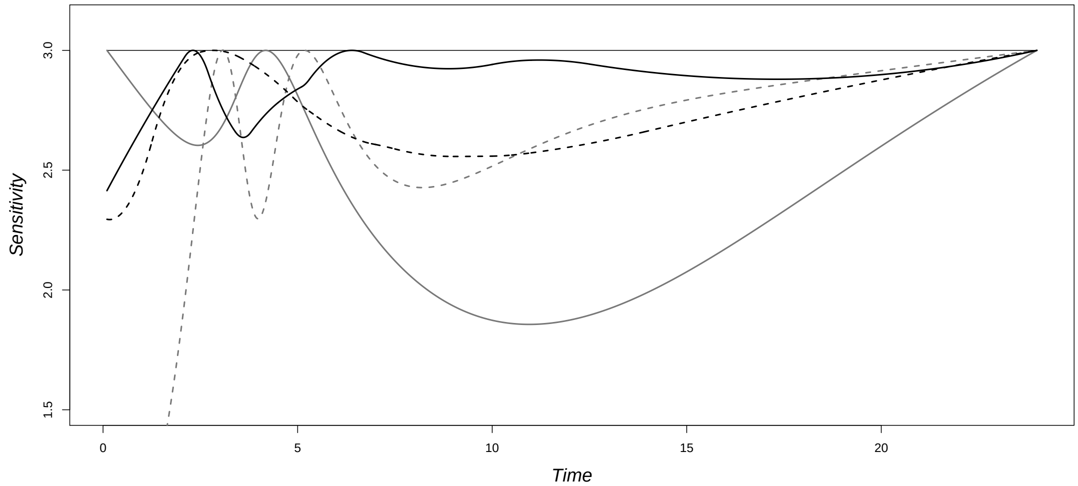

Fig. 1 Sensitivity functions on the design region $\mathscr{X} = [0.1, 24]$.
Grey dashed: $\mathbf{M}_{1,\beta}$; Grey solid: $\mathbf{M}_{2,\beta}$; Black dashed: $\mathbf{M}_{3,\beta}$; Black solid: $\mathfrak{M}_{\beta}$.

$$\Phi_{D;3}(\zeta_1) = 13.81 \text{ for } \mathbf{M}_{3,\beta}; \;\; \Phi_{D;F}(\zeta_1) = 14.77 \text{ for } \mathfrak{M}_{\beta}.$$

The information gain here takes place for the approximation $\mathbf{M}_{3,\beta}$ as well, as the value of the optimality criterion is smaller than the value given by the Fisher information, which is approximated by the criterion $\Phi_{D;F}(\zeta_1)$. In the present example, it can be even analytically shown that the approximation $\mathbf{M}_{1,\beta}$ suggests more information then the true Fisher information.

Table 1 Efficiency of the proposed designs

	ζ_1^*	ζ_2^*	ζ_3^*	ζ_F^*
$\delta_{\zeta_1^*}$	1.00	0.55	0.95	0.77
$\delta_{\zeta_2^*}$	0.66	1.00	0.00	0.37
$\delta_{\zeta_3^*}$	0.98	0.84	1.00	0.98
$\delta_{\zeta_F^*}$	0.88	0.83	0.95	1.00

6 Discussion

Although some new motivations propose the use of the information matrix resulting from linear mixed effects models for the design of experiments in nonlinear mixed effects models, the results for specific situations can be unsatisfactory. Designs which are obtained by analytical approximations of the Fisher information should be handled very carefully in the considered models, if the target of the design problem is the minimization of some criteria related to the Fisher information matrix. These designs can however be used as benchmarks or as starting points for the location of designs based on numeric approximations of the Fisher information. Of special interest is the optimization of designs with respect to the expected information $\mathbf{M}_{3,\beta}$. Pinheiro and Bates (2000) suggest a similar structure of the inverse of the covariance matrix for estimators of β, which are based on penalized sums of squares, such that the optimization of designs with respect to the information matrix $\mathbf{M}_{3,\beta}$ might improve the quality of estimates not only theoretically. The presented example unfortunately showed that designs with less support points than parameters of interest might result with this information matrix as optimal designs, what might however cause new problems for the estimation. A second problem for the approximation $\mathbf{M}_{3,\beta}$ was the value of the optimality criterion, which was in the example below the bound provided by the approximated Fisher information. These points require some further investigations.

Acknowledgment

The author would like to thank Rainer Schwabe (Univ. of Magdeburg) for his very useful remarks and is grateful for the support by the BMBF grant SKAVOE 03SCPAB3.

References

Demidenko, E. (2005). *Mixed Effects Models: Theory and Applications*. Wiley, New Jersey.

Fedorov, V. (1972). *Theory of Optimal Experiments*. Academic Press, New York.

Kuhn, E. and M. Lavielle (2001). Maximum likelihood estimation in nonlinear mixed effects models. *Computational Statistics & Data Analysis 49*, 1020–1038.

Mielke, T. (2011). Approximation of the Fisher information and design in nonlinear mixed effects models. Preprint 22-11, Fakultät für Mathematik, Otto-von-Guericke University Magdeburg.

Mielke, T. and R. Schwabe (2010). Some considerations on the Fisher information in nonlinear mixed effects models. In A. Giovagnoli, A. Atkinson, B. Torsney, and C. May (Eds.), *mODa9 - Advances in Model-Oriented Design and Analysis*, pp. 129–136. Physica, Heidelberg.

Pinheiro, J. and D. Bates (2000). *Mixed-Effects Models in S and S-Plus*. Springer, New York.

Retout, S. and F. Mentré (2003). Further developments of the Fisher information matrix in nonlinear mixed effects models. *Journal of Biopharmaceutical Statistics 13*, 209–227.

Schmelter, T. (2007). *Experimental Design For Mixed Models With Application to Population Pharmacokinetic Studies*. Ph. D. thesis, Otto-von-Guericke University Magdeburg.

Tierney, L. and J. Kadane (1986). Accurate approximations for posterior moments and marginal densities. *Journal of American Statistical Association 81*, 82–86.