Statistical Power of Approximate Tests in Two-Stage Nonlinear Mixed Models

Jeff H. Burton * Julia Volaufova † Robbie A. Beyl * William D. Johnson *

Abstract

Here, we investigate power of tests about fixed effects parameters in nonlinear mixed models. For obtaining estimates of unknown model parameters needed for construction of test statistics, a two-stage approach is implemented. The tests of interest that shall be compared are the large-sample Wald and likelihood ratio tests as well as several approximate *F*-tests. For the Wald and approximate *F*-tests, we modify the test statistics by developing an approximation to the large-sample covariance matrix of estimates of fixed population parameters based on the Fisher information matrix. In contrast to previous research in which null properties of tests were investigated (Burton & Volaufova, 2014), our focus here is on comparing power of tests using various configurations of fixed model parameters under the alternative hypothesis. A simulation study using a one-compartment pharmacokinetic model is used for illustration.

Key Words: Nonlinear, mixed model, two-stage, power, testing, Fisher information matrix

1. Introduction

With respect to properties of hypothesis tests about fixed effects parameters in nonlinear mixed models, the literature focuses primarily on tests which are constructed from estimates based on some approximation of the marginal likelihood. These tests have been shown to have poor properties (e.g., Bertrand, Comets, & Mentré, 2008; Bertrand, Comets, Chenel, & Mentré, 2012; Volaufova, 2014). We focus instead on using a two-stage estimation approach that does not rely on the marginal likelihood. We have shown previously that certain tests, when constructed using two-stage estimates, have favorable properties under the null hypothesis (Volaufova & Burton, 2012; Burton, 2013; Burton & Volaufova, 2014). The goal here is to extend this research to include properties of the tests under the alternative hypothesis. In particular, we will compare power between the approximate *F*-tests and large-sample Wald and likelihood ratio tests.

2. Model and Methods

The nonlinear mixed model considered in this context is a random coefficients model for longitudinal data. Such a model can be defined as a hierarchical model with two levels (e.g., Davidian & Giltinan, 2003). The first is the *individual* level and is defined in terms of a single subject i for $i = 1, \ldots, N$ with responses observed at times t_{ij} for $j = 1, \ldots, T_i$.

$$Y_i = f(\theta_i, X_i) + \varepsilon_i, \quad i = 1, \ldots, N. \tag{1}$$

In model (1), Y_i is a T_i-dimensional vector of responses, θ_i is a p-dimensional vector of subject-specific parameters, X_i is a matrix containing within-subject covariates including observation times, $f(\cdot)$ is a function that is nonlinear and twice differentiable with repect to θ_i, and ε_i is a normally distributed T_i-dimensional random error vector with mean zero and covariance matrix $\sigma^2 R(\phi)$.

*Pennington Biomedical Research Center, 6400 Perkins Road, Baton Rouge, LA 70808
†LSU Health Sciences Center, 2020 Gravier Street, New Orleans, LA 70112

The second level of the model is the *population* level. Here, the subject-specific parameters θ_i are defined in terms of variation about a set of fixed population parameters.

$$\theta_i = g(a_i, \beta) + Bb_i, \quad i = 1, \ldots, N, \tag{2}$$

Here, a_i is a s-dimensional design vector, β is a r-dimensional vector of unknown fixed population parameters, $g(\cdot)$ is a twice differentiable function of β, B is a $p \times q$ matrix used to indicate which components of θ_i are random, and b_i is a q-vector of normally distributed random effects with mean zero and covariance matrix $D(\lambda)$.

In order to obtain the general form of the nonlinear mixed model, these two levels can be combined by plugging (2) in for θ_i in (1).

2.1 Two-Stage Estimation

Rather than combining the two levels of the hierarchical model defined in the previous section, we utilize a two-stage approach to obtain estimates of the unknown model parameters. Before proceeding, we shall make some simplifications to the context in which we will investigate. First, we only consider the case in which $\text{Cov}(\varepsilon_i) = \sigma^2 I_{T_i}$. Second, we shall assume that the entire vector θ_i is random. That is, $q = p$ and $B = I_q = I_p$. Last, we shall assume that the relatioship between θ_i and β is linear (i.e., $g(a_i, \beta) = A_i\beta$, where $A_i = I \otimes a_i'$, (e.g., Vonesh & Carter, 1987)).

The two-stage estimation approach begins with obtaining ordinary nonlinear least squares (ONLS) estimates of θ_i from (1). For this approach to be viable, there must be a sufficient number of observations (T_i) per subject. If this condition is met for subject i, we get ONLS estimates via the objective funtion

$$Q(\theta_i) = [y_i - f(\theta_i, X_i)]'[y_i - f(\theta_i, X_i)], \tag{3}$$

where

$$\tilde{\theta}_i = \arg\min_{\theta_i} Q(\theta_i).$$

Under certain regularity conditions (see Burton (2013) and the references therein), we can assume asymptotic normality and approximate the marginal mean and covariance matrix of $\tilde{\theta}_i$ (Burton, 2013) as

$$E(\tilde{\theta}_i) \approx A_i\beta \tag{4}$$

and

$$\text{Cov}(\tilde{\theta}_i) \approx D(\lambda) + \sigma^2 \left[Z_i(\beta)' Z_i(\beta)\right]^{-1} \equiv W_i(\beta, \vartheta). \tag{5}$$

In (5), the covariance matrix W_i is a function of both fixed effects parameters β and variance-covariance parameters $\vartheta = (\lambda, \sigma^2)'$. From this marginal distribution of $\tilde{\theta}_i$, then, we can define the second-stage model as a generalized linear mixed model with form

$$\tilde{\theta}_i = A_i\beta + u_i, \quad i = 1, \ldots, N, \tag{6}$$

where u_i is a normally distributed error term with mean zero and covariance matrix $W_i(\beta, \vartheta)$. Using model (6), we can utilize maximum likelihood methodology to estimate the unknown parameters (see Burton (2013) for details).

2.2 Hypothesis Testing

The focus of this research is investigation of properties of tests of linear combinations of fixed effects parameters. The hypotheses of interest are of form:

$$H_0 : L'\beta = h_0 \text{ vs. } H_1 : L'\beta \neq h_0, \tag{7}$$

where L is a known $r \times c$ matrix of full column rank. Commonly used tests are the large-sample Wald test and likelihood ratio test (LRT), both of which are asymptotically χ^2 distributed with degrees of freedom c. Additionally, approximate F-tests can be used, with test statistic

$$F^* = (1/c)(L'\hat{\boldsymbol{\beta}} - h_0)'[\hat{\text{Cov}}(L'\hat{\boldsymbol{\beta}})]^{-1}(L'\hat{\boldsymbol{\beta}} - h_0). \tag{8}$$

Under the null hypothesis H_0, F^* is approximately F-distributed with numerator degrees of freedom c and denominator degrees of freedom (*ddf*) denoted ν. When constructing such tests, the key points are the covariance matrix of $L'\hat{\beta}$ (see Section 2.2.1) and the choice of ν. In the simulation study described in Section 3, *ddf* for the approximate F-tests investigated are either $N - q$, where N is number of subjects and q is number of random effects (Wolfinger, 2000), or Fai-Cornelius-Satterthwaite type approximate degrees of freedom (Fai & Cornelius, 1996).

The focus here, however, will be on power of the tests. Thus, we are investigating under the alternative hypothesis. In this nonlinear setting, the distributions of the test statistics under H_1 are unknown.

2.2.1 Approximate Covariance Matrix of $\hat{\beta}$

Assuming large-sample normality of $\tilde{\theta}_i$, we can define the Fisher information matrix as

$$\mathcal{I}(\varphi) = -\sum_{i=1}^{N} E \begin{pmatrix} \dfrac{\partial^2 \ell_i(\varphi|\tilde{\theta}_i)}{\partial\beta\partial\beta'} & \dfrac{\partial^2 \ell_i(\varphi|\tilde{\theta}_i)}{\partial\beta\partial\gamma'} \\ \dfrac{\partial^2 \ell_i(\varphi|\tilde{\theta}_i)}{\partial\gamma\partial\beta'} & \dfrac{\partial^2 \ell_i(\varphi|\tilde{\theta}_i)}{\partial\gamma\partial\gamma'} \end{pmatrix} = \begin{pmatrix} J(\varphi) & C(\varphi) \\ C'(\varphi) & B(\varphi) \end{pmatrix}, \tag{9}$$

where $\ell_i(\cdot)$ is the normal log-likelihood, φ is the vector of all unknown model parameters, and $J(\varphi)$, $B(\varphi)$, and $C(\varphi)$ represent blocks of $\mathcal{I}(\varphi)$.

Using the Schur complement to invert (9), we get the following expression for the approximate covariance matrix of $\hat{\beta}$ (e.g., Retout, Mentré, & Bruno, 2002; Retout & Mentré, 2003; Volaufova & Burton, 2012; Burton, 2013; Burton & Volaufova, 2014):

$$\text{Cov}_S(\hat{\beta}) \approx J(\varphi)^{-1} + J(\varphi)^{-1}C(\varphi)\text{Cov}_S(\hat{\vartheta})C'(\beta,\vartheta)J(\varphi)^{-1}, \tag{10}$$

where $\text{Cov}_S(\hat{\vartheta}) \approx [B(\varphi) - C'(\varphi)J(\varphi)^{-1}C(\varphi)]^{-1}$ and the subscript S refers to the Schur complement based approach. The result is similar in nature to what was done in the linear mixed model by Kenward and Roger (1997), such that the covariance matrix of $\hat{\beta}$ consists of two terms, the second of which accounts for variability in $\hat{\beta}$ due to estimating ϑ.

A simpler approach to finding an expression for the covariance matrix of $\hat{\beta}$ is to ignore the dependence of (5) on β (e.g., Mielke & Schwabe, 2010; Mielke, 2011). Doing so results in a second-stage model that has the form of a typical linear mixed model:

$$\tilde{\theta}_i = A_i\beta + u_i^*, \quad u_i^* \sim N(0, D^*), \quad i = 1, \ldots, N, \tag{11}$$

where D^* is some positive definite matrix that is a function only of variance-covariance parameters. From (11) we get

$$\text{Cov}_A(\hat{\beta}) \approx \left[\sum_{i=1}^{N} A_i' D^{*-1} A_i\right]^{-1}, \tag{12}$$

where subscript A is short for *alternative*.

Estimates of both matrices (10) and (12) needed for calculating Wald and approximate F-test statistics can be obtained by plugging in maximum likelihood estimates $\hat{\beta}$ and $\hat{\vartheta}$.

Table 1: Descriptions and legend labels of tests investigated.

Label	Cov. Approx.	Statistic	df	Distribution
Wald	$\hat{\text{Cov}}_S(\hat{\beta})$	Wald$(= cF_S^*)$	c	χ^2
LRT	N/A	LRT	c	χ^2
Wolf	$\hat{\text{Cov}}_S(\hat{\beta})$	F_S^*	(Wolfinger) $N-q$	$F(c, N-q)$
Satt	$\hat{\text{Cov}}_S(\hat{\beta})$	F_S^*	(Satterthwaite) ν_S	$F(c, \nu_S)$
A	$\hat{\text{Cov}}_A(\hat{\beta})$	F_A^*	(Satterthwaite) ν_A	$F(c, \nu_A)$

3. Simulation Study

In order to investigate the power of the tests of interest constructed under the two-stage estimation scheme, we use a simulation study. We first define our null and alternative hypotheses:

$$H_0 : L'\beta = 0 \text{ vs. } H_1 : L'\beta \neq 0. \tag{13}$$

In this study, data are generated under various parameter configurations under the alternative hypothesis, empirical power is estimated for each test, and results are compared graphically.

3.1 One-Compartment Pharmacokinetic Model

For illustration of the methodology described here, we carry out a simulation study in which data are generated from a one-compartment pharmacokinetic (PK) model for an intravenous (IV) bolus. The *individual* level of the two-stage model is defined

$$Y_{ij} = \frac{d}{V_i} e^{-\left(\frac{Cl_i}{V_i}\right)t_{ij}} + \varepsilon_{ij},\ i = 1, \ldots, N,\ j = 1, \ldots, T_i. \tag{14}$$

In this model, the subject-specific parameters are $Cl_i = e^{\theta_{1i}}$ and $V_i = e^{\theta_{2i}}$, which are PK parameters representing clearance and volume of distribution. They are exponentiated in order to ensure positive values. Other model components are drug dose d, time points t_{ij}, and mutually independent error terms $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$.

For this simulation study, we are comparing two populations. We can define the *population* level of the two-stage model, for the first population:

$$\theta_{1i} = \beta_1 + b_{1i}; \quad \theta_{2i} = \beta_2 + b_{2i} \tag{15}$$

and, for the second population:

$$\theta_{1i} = \beta_1 + \delta_1 + b_{1i}; \quad \theta_{2i} = \beta_2 + \delta_2 + b_{2i}. \tag{16}$$

Using these definitions for our two populations, we can set $\delta_1 = \delta_2 = 0$ to generate data under the null hypothesis. Values for β_1 and β_2 are the same in both populations and take values $\beta_1 = \log(1.25)$ and $\beta_2 = \log(10)$ as in Hartford and Davidian (2000). For the covariance matrix of b_i,

$$D(\lambda) = \begin{pmatrix} 0.086 & 0.052 \\ 0.052 & 0.086 \end{pmatrix}.$$

The dose $d = 100$ is the same for all subjects, and two values of σ^2 are used, 1.0 and 2.0.

In order to generate data under the alternative hypothesis, several non-zero values of δ_1 are specified. Assigning a non-zero value to this parameter represents a difference between

Table 2: Adjusted empirical power of investigated tests.

(n_1, n_2)	σ^2	δ_1	Wald	LRT	Wolf	Satt	A
$(10, 8)$	1.0	0.1	0.092	0.111	0.080	0.082	0.094
		0.2	0.196	0.260	0.155	0.161	0.211
		0.3	0.399	0.495	0.327	0.337	0.416
		0.4	0.615	0.714	0.537	0.548	0.632
		0.5	0.824	0.879	0.766	0.776	0.811
		0.6	0.934	0.960	0.916	0.919	0.916
		0.7	0.984	0.986	0.991	0.991	0.971
		0.8	0.999	0.990	1.000	1.000	0.991
	2.0	0.1	0.069	0.091	0.059	0.062	0.093
		0.2	0.133	0.194	0.107	0.113	0.176
		0.3	0.254	0.363	0.201	0.208	0.321
		0.4	0.417	0.553	0.337	0.348	0.495
		0.5	0.618	0.743	0.536	0.546	0.654
		0.6	0.768	0.858	0.713	0.722	0.807
		0.7	0.883	0.932	0.853	0.857	0.902
		0.8	0.950	0.967	0.946	0.947	0.951
$(20, 25)$	1.0	0.1	0.107	0.148	0.100	0.101	0.156
		0.2	0.388	0.497	0.358	0.358	0.474
		0.3	0.746	0.852	0.726	0.727	0.808
		0.4	0.953	0.976	0.951	0.951	0.963
		0.5	0.999	0.986	1.000	1.000	0.996
		0.6	1.000	0.987	1.000	1.000	1.000
		0.7	1.000	0.987	1.000	1.000	1.000
		0.8	1.000	0.987	1.000	1.000	1.000
	2.0	0.1	0.086	0.124	0.082	0.083	0.115
		0.2	0.252	0.376	0.226	0.228	0.357
		0.3	0.535	0.694	0.508	0.509	0.638
		0.4	0.799	0.912	0.780	0.780	0.864
		0.5	0.960	0.982	0.959	0.959	0.967
		0.6	1.000	0.991	1.000	1.000	0.994
		0.7	1.000	0.992	1.000	1.000	0.998
		0.8	1.000	0.993	1.000	1.000	1.000

the two populations in the fixed parameter β_1. For each alternative hypothesis configuration, the value assigned to δ_1 ranges from 0.1 to 0.8. Samples are generated with equal observation times per subject using the vector $t_i = (0.25, 0.5, 1, 2, 4, 8, 12, 24)'$. Two sample size options $(n_1, n_2) = (10, 8)$ and $(n_1, n_2) = (20, 25)$ are used. For each model configuration, 5000 replications are generated.

Table 1 details each test investigated and assigns a corresponding label to be used for identification in the figures.

3.1.1 Power

For each model configuration, power was defined as the proportion of p-values from each test less than the nominal significance level 0.05. Because these tests have different em-

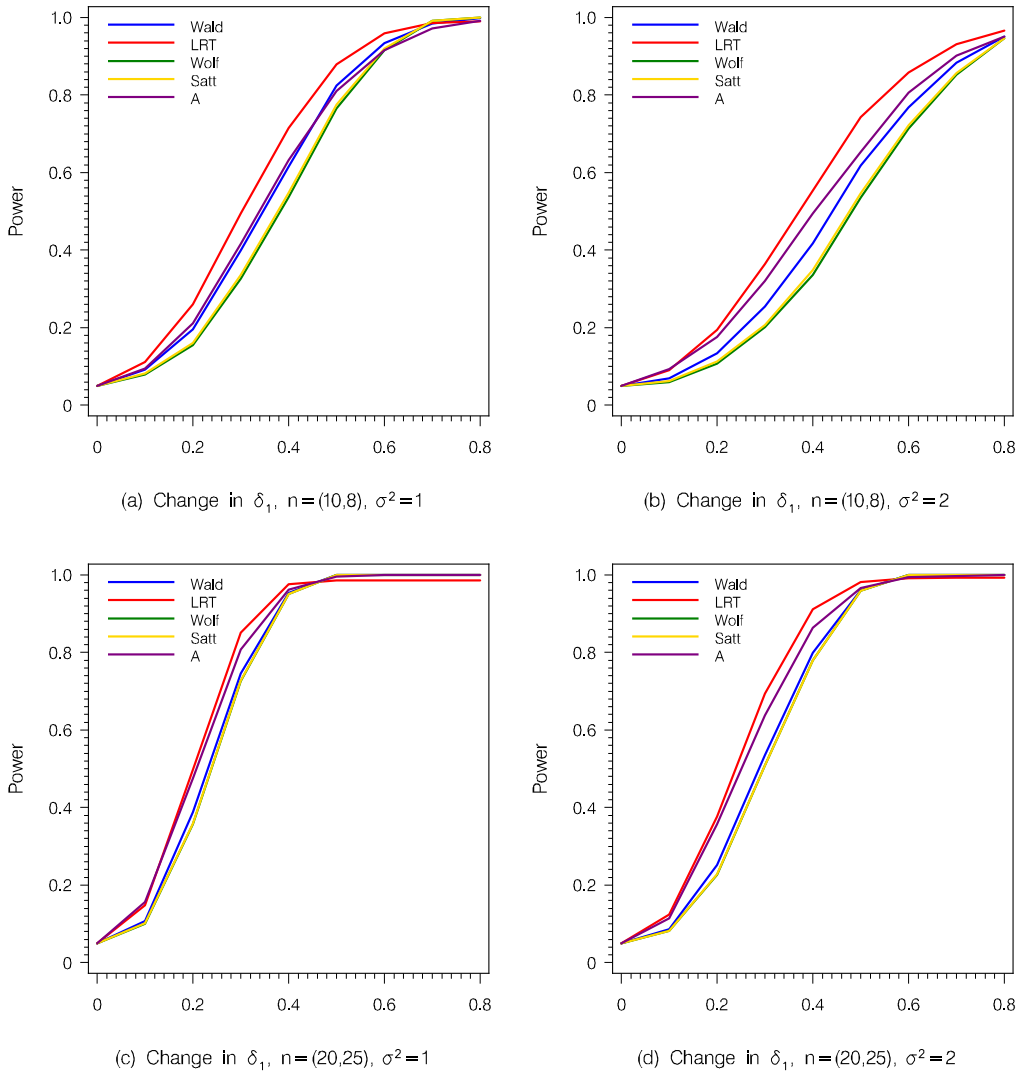

Figure 1: Adjusted power of tests constructed using parameter estimates obtained via two-stage estimation scheme.

pirical type I error rates, estimated power was adjusted such that each test has power of 0.05 under H_0, making the comparisons more meaningful. For each test, this was done by adjusting estimated power for all model configuration under H_1 by the difference between estimated power under H_0 and 0.05.

3.2 Results

In Figure 1, subfigures (a) and (b) represent adjusted empirical power curves for tests under the smallest sample size setting $(n_1, n_2) = (10, 8)$. For subfigure (a), data was generated with error variance $\sigma^2 = 1$. Here, the LRT has the highest adjusted power. The two approximate *F*-tests, *Wolf* and *Satt*, have the lowest adjusted power among the tests investigated and perform very similar to one another. Between are the Wald test and the aprroximate *F*-test *A*, which also perform very similarly in terms of adjusted power. Looking at subfigure (b) in which error variance $\sigma^2 = 2$, all of the curves are shifted to the right. This indicates that the power is lower for the same changes in δ_1 when σ^2 is increased from 1 to 2.

In the two lower subfigures (c) and (d), adjusted power curves are displayed for the

larger sample size setting $(n_1, n_2) = (20, 25)$. In comparison to the empirical adjusted power curves for the tests under the smaller sample size setting, the curves here approach 1.0 (100% power) more rapidly. That is, all of the tests are able to detect changes in δ_1 with higher power when sample size is increased. The ordering of the tests from highest power to lowest power remains the same as observed in the smaller sample size setting. Here, however, the Wald test performs more similarly to approximate *F*-tests *Wolf* and *Satt*, and the approximate *F*-test *A* is more similar in terms of adjusted power to the LRT.

Estimates of adjusted power under each model configuration investigated can be found in Table 2.

4. Discussion

The research presented here is an extension of Burton and Volaufova (2014), in which properties of the Wald test, likelihood ratio test, and approximate *F*-tests were investigated under the null hypothesis. The conslusions about power of these tests is limited to a single simulation study (see Section 3). In order to draw more concrete conclusions, additional simulation studies are required in which a broader range of model configurations and additional models are explored. Based solely on the model and model configurations explored here, the likelihood ratio test appears to be most powerful in all settings. Currently investigations are underway to determine properties of these tests under the null (type I error rate, accuracy of p-values) and alternative (power) hypotheses in a more general framework. Here, we make simplifications with respect to the covariance matrix of the error vector ε_i. It may be more meaningful to determine these properties under a general structure for this matrix which will require a different estimation method for developing the second stage model (6).

Acknowledgements

This research was supported in part (JHB, RAB, WDJ) by 1 U54 GM104940 from the National Institute of General Medical Sciences of the National Institutes of Health, which funds the Louisiana Clinical and Translational Science Center. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

Bertrand, J., Comets, E., & Mentré, F. (2008). Comparison of model-based tests and selection strategies to detect genetic polymorphisms influencing pharmacokinetic parameters. *Journal of Biopharmaceutical Statistics, 18*, 1084–1102.

Bertrand, J., Comets, E., Chenel, M., & Mentré, F. (2012). Some alternatives to asymptotic tests for the analysis of pharmacogenetic data using nonlinear mixed effects models. *Biometrics, 68*, 146–155.

Burton, J.H. (2013). *Improved hypothesis tests of linear combinations of fixed effects parameters in nonlinear random coefficients models* (Unpublished doctoral dissertation). Louisiana State University Health Sciences Center, New Orleans, LA, USA.

Burton, J.H. & Volaufova, J. (2014). Approximate testing in two-stage nonlinear mixed models. *Journal of Statistical Computation and Simulation*. Advance online publication. doi:10.1080/00949655.2014.948442

Davidian, M. & Giltinan, D. (2003). Nonlinear models for repeated measurement data: An overview and update. *Journal of Agricultural, Biological, and Environmental Statistics, 8*(4), 387–419.

Fai, A. & Cornelius, P. (1996). Approximate F-tests of multiple degree of freedom hypotheses in generalized least squares analyses of unbalanced split-plot experiments. *Journal of Statistical Computation and Simulation, 54*, 363–378.

Hartford, A. & Davidian, M. (2000). Consequences of misspecifying assumptions in nonlinear mixed effects models. *Computational Statistics and Data Analysis, 34*, 139–164.

Kenward, M. & Roger, J. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics, 53*(3), 983–997.

Mielke, T. & Schwabe, R. (2010). Some considerations on the Fisher information in nonlinear mixed effects models. In A. Giovagnoli, A.C. Atkinson, B. Torsney, & C. May (Eds.) *Proceedings of the 9th International Workshop in Model-Oriented Design and Analysis* (129–136). Bertinoro, Italy: Springer, Physica Verlag Heidelberg.

Mielke, T. (2011). *Approximation of the fisher information and design in nonlinear mixed effects models*. Retrieved from: http://www.math.uni-magdeburg.de/institute/imst/ag_schwabe/preprints/2011_22.pdf.

Retout, S., Mentré, F., & Bruno, R. (2002). Fisher information matrix for nonlinear mixed effects models: Evaluation and application for optimal design of Enoxaparin population pharmacokinetics. *Statistics in Medicine, 21*, 2623–2639.

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

Volaufova, J. & Burton, J.H. (2012). *Note on hypothesis testing in mixed models*.Paper presented at LINSTAT 2012/21st IWMS, Bedlewo, Poland.

Volaufova, J. (2014). Approximate small sample tests in nonlinear mixed models. *Communication in Statistics - Simulation and Computation, 43*(9), 2117–2137.

Vonesh, E., Carter, R. (1987). Efficient inference for random-coefficient growth curve models with unbalanced data. *Biometrics, 43*(3), 617–628.

Wolfinger, R.D. (2000). *Fitting nonlinear mixed models with the new NLMIXED procedure* (Technical report 287). Cary, NC: SAS Institute.