Design evaluation and optimisation in crossover pharmacokinetic studies analysed by nonlinear mixed effects models

Thu Thuy Nguyen, Caroline Bazzoli, France Mentré

Design evaluation and optimisation in crossover pharmacokinetic studies analysed by nonlinear mixed effects models

Thu Thuy Nguyen*, Caroline Bazzoli, France Mentré

UMR738 INSERM and University Paris Diderot, Paris, France

Bioequivalence or interaction trials are commonly studied in crossover design and can be analysed by nonlinear mixed effects models as an alternative to noncompartmental approach. We propose an extension of the population Fisher information matrix in nonlinear mixed effects models to design crossover pharmacokinetic trials,using a linearisation of the model around the random effect expectation, including within-subject variability and discrete covariates fixed or changing between periods. We use the expected standard errors of treatment effect to compute the power for the Wald test of comparison or equivalence and the number of subjects needed for a given power. We perform various simulations mimicking crossover two-period trials to show the relevance of these developments. We then apply these developments to design a crossover pharmacokinetic study of amoxicillin in piglets and implement them in the new version 3.2 of the R function PFIM.

Keywords: ***crossover; design; Fisher information matrix; nonlinear mixed effects models; PFIM; power; sample size***

*Correspondence to: Thu Thuy Nguyen, UMR738 INSERM and University Paris Diderot, 75018 Paris, France; E-mail: thu-thuy.nguyen@inserm.fr

1 Introduction

Pharmacokinetics (PK) is the science of the kinetics of drug in an organism, generally based on drug plasma concentrations. Bioequivalence trials are performed to compare the PK of two drug formulations and interaction trials are performed to study the effect on the PK of another drug. The most commonly used design for these trials is the crossover design, where measurements are made at several occasions. Bioequivalence is tested from the log ratio of the geometric means of two PK parameters: the area under the curve (AUC) of the concentrations and the maximal concentrations (C_{max}) [1, 2]. Linear mixed effects models (LMEM) including treatment effect are usually used to analyse the log-transformed AUC and C_{max} [3]. Bioequivalence tests are then performed on the estimates of the treatment effect. The same testing method is used for an assessment of the absence of interaction. The tested end-points AUC and C_{max} are usually estimated by non-compartmental analysis (NCA) using the trapezoidal rule to evaluate AUC [4]. NCA needs few hypotheses but has several limitations. Although using sparse sampling in NCA is possible [5, 6, 7], a large number of samples per subject (usually between 10 and 20, as in trials on healthy volunteers) is however needed when using this approach to be able to estimate mean and variability of AUC and C_{max} with a good precision. NCA also does not take into account nonlinear pharmacokinetics, which can bias the bioavailability estimation [8] and may amplify small availability differences between drugs [9]. Alternatively, another approach to analyse PK data is the population approach based on nonlinear mixed effects models (NLMEM) [10, 11], which is more complex than NCA but benefits from the knowledge accumulated about the drug and can characterise the PK with few samples per subject, enabling us to analyse studies in patients with more power than does NCA. These models can also lead to a better understanding of the biological system and help to interpret ambiguous results. However, the use of NLMEM is rather recent and still rare in early phases of drug development and/or to analyse crossover studies in spite of its several advantages [12, 13, 14].

The work presented here focusses on crossover PK trials analysed by NLMEM. Before the modelling step, data needs to be collected and we have consequently to define an appropriate design, which consists of determining a balance between the number of subjects and the number of samples per subject as well as the allocation of sampling times according to experimental conditions. The choice of design has an important impact on the study results, on the precision of the parameter estimates and on the power of the tests [15, 16, 17]. Indeed, a bad choice of design can lead to results which are difficult to interpret and minimise the interest of the study. The main approach for design evaluation has been for a long time based on simulations but it is a cumbersome method, and thus the number of designs which can be evaluated is limited. An alternative approach has been described in the general theory of optimum experimental design used for classical nonlinear models [18, 19], relying on the inequality of Rao-Cramer which states that the inverse of the Fisher information matrix ($\boldsymbol{M}_F$) is the lower bound of the variance-covariance matrix of any unbiased estimate of the parameters and its diagonal elements are the expected standard errors (SE) of the parameters. Several criteria based on $\boldsymbol{M}_F$ have been developed to evaluate designs. One of the criteria widely used is the criterion of D-optimality, consisting of maximising the determinant of $\boldsymbol{M}_F$. But since there is no analytical form of the likelihood in our framework, the determination of the exact analytical expression of $\boldsymbol{M}_F$ is not possible. An approximation of $\boldsymbol{M}_F$ was first proposed by Mentré *et al* [20] for NLMEM. Its evaluation was implemented in the R function PFIM [21] using first order linearisation of the model around the random effect expectation. Evaluation and optimisation of design have become a large field of research with more and more publications on several extensions of $\boldsymbol{M}_F$ [22, 23, 24, 25, 26] as well as the creation of several software besides PFIM by different research teams, implemented in MATLAB (MathWorks Inc., Natick,

MA, USA) (PopDes, PopED, WinPOPT). These softwares dedicated to evaluating and optimising designs were compared in an oral presentation by Mentré *et al* [27] at the meeting of Population Approach Group in Europe (www.page-meeting.org). However, these expressions of $\boldsymbol{M}_F$ were not yet applicable to crossover trials with models including within subject variability (WSV) and discrete covariates changing between periods. It is important to model not only between subject variability (BSV) but also WSV when subjects have measurements at several periods, as accentuated by Karlsson and Sheiner [28] and Panhard *et al* [12]. The development of $\boldsymbol{M}_F$ in NLMEM under uni-response models with additional terms for WSV has been implemented and illustrated on a specific example of two occasions [29]. However, this has not yet been generalised to crossover designs for any number of periods and sequences, neither extended to include discrete covariates which can change between periods. The prediction of power for the Wald test to detect a treatment effect has been previously described by Retout *et al* [30] using the expected SE of the treatment effect but only for an interaction study in a parallel group design, not yet for an equivalence study in a crossover design. These developments would be very useful since to our knowledge, this is the first proposed statistical method to predict power of test and to choose optimal sampling times for crossover bioequivalence or absence of interaction trials analysed by NLMEM.

In this context, the first objective of this work is to extend the expression for $\boldsymbol{M}_F$ in NLMEM including WSV in addition to BSV and discrete covariates fixed or changing between periods in crossover trials. Then, we also propose a computation of the expected power for the Wald test of comparison or equivalence and the number of subjects needed (NSN) for a given power. The relevance of these developments is evaluated by simulations mimicking a PK crossover two-period study of amoxicillin (an antibiotic widely used in bacterial infection treatment) in piglets. Finally, we apply these extensions to design a future PK crossover study in piglets, based on the results of a previous study inspiring our simulations, aiming to show the absence of interaction of a compound X on the PK of amoxicillin. We wish, through this motivating example, to show how our statistical methodology can be easily applied in practice, to choose optimal sparse sampling times with almost no loss of power and to compute the NSN for a given power. The example also brings us to discuss how different elements in a study design (number of subjects, of model parameters, of periods, etc.) could influence the power and to investigate/confirm the advantage of crossover design compared to parallel group design in bioequivalence trials.

We introduce some notation and present the new extension of $\boldsymbol{M}_F$ as well as the computation of the power and NSN in Section 2. By simulation, we evaluate the relevance of these developments in Section 3. An application of this extension to a real example is presented in Section 4, followed by a discussion in Section 5.

2 Extension of population Fisher information matrix

2.1 Notations

2.1.1 Design

The elementary design $\boldsymbol{\xi}_{ih}$ of individual i ($i = 1, \ldots, N$) at period h ($h = 1, \ldots, H$) is defined by the number n_{ih} of samples and their allocation in time $(t_{ih1}, \ldots, t_{ihn_{ih}})$. The global elementary design ξ_i of individual i for H periods is defined as $\boldsymbol{\xi}_i = (\boldsymbol{\xi}_{i1}, \ldots, \boldsymbol{\xi}_{iH})$, with a total number of samples $n_i = \sum_{h=1}^{H} n_{ih}$. Consequently, the population design for N individuals and for H periods can be defined as $\boldsymbol{\Xi} = \{\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_N\}$. Usually, population designs are composed of a limited number Q of groups of individuals with identical designs in each group. Each of these groups is

composed of a global elementary design $\boldsymbol{\xi}_q$ and is performed in a number N_q of subjects. The population design can thus be written as $\boldsymbol{\Xi} = \{[\boldsymbol{\xi}_1, N_1]; \ldots; [\boldsymbol{\xi}_Q, N_Q]\}$.

2.1.2 Models

We denote by $\boldsymbol{y}_{ih}$ the n_{ih}-vector of observations for the individual i measured at period h and by f the known nonlinear function describing the nonlinear structural model. The NLMEM linking $\boldsymbol{y}_{ih}$ to the sampling times $\xi_{ih} = (t_{ih1}, \ldots, t_{ihn_{ih}})$ can be written as

$$\boldsymbol{y}_{ih} = f(\boldsymbol{\phi}_{ih}, \boldsymbol{\xi}_{ih}) + \boldsymbol{\epsilon}_{ih} \tag{1}$$

f maps $\boldsymbol{\phi}_{ih}$, the P-vector of parameters for individual i at period h and the set ξ_{ih} of n_{ih} sampling times to a vector of which each element is a scalar $f(\boldsymbol{\phi}_{ih}, t_{ihj})$, with $j = 1, \ldots, n_{ih}$. ϵ_{ih} is the vector of random error which follows a normal distribution $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{ih})$ with $\boldsymbol{\Sigma}_{ih}$ a $n_{ih} \times n_{ih}$ diagonal matrix expressed as a function of $(\boldsymbol{\phi}_{ih}, \sigma_{inter}, \sigma_{slope}, \boldsymbol{\xi}_{ih})$ such as each diagonal element of $\boldsymbol{\Sigma}_{ih}$ is $(\sigma_{inter} + \sigma_{slope} f(\boldsymbol{\phi}_{ih}, t_{ihj}))^2$, with $j = 1, \ldots, n_{ih}$. σ_{inter} and σ_{slope} are additional and proportional parameters in the variance model of the random error. The case of $\sigma_{slope} = 0$ returns a homoscedastic error model, whereas the case of $\sigma_{inter} = 0$ returns a constant coefficient of variation error model. The general case where the two parameters differ from 0 is called a combined error model.

We denote by C the set of discrete covariates changing or not between periods, K_c is the set of all possible categories k of the covariate c in C. We assume that the vector $\boldsymbol{\phi}_{ih}$ of parameters for an individual i at period h can be expressed as a function of the covariate c_{ih} of individual i at period h, which can be additive (for normal parameters):

$$\boldsymbol{\phi}_{ih} = \boldsymbol{\mu} + \sum_{c \in C} \sum_{k \in K_c} 1_{c_{ih}=k} \boldsymbol{\beta}_{c_k} + \boldsymbol{b}_i + \boldsymbol{\kappa}_{ih} \tag{2}$$

or exponential (for lognormal parameters):

$$\boldsymbol{\phi}_{ih} = \boldsymbol{\mu} \otimes \exp\Big(\sum_{c \in C} \sum_{k \in K_c} 1_{c_{ih}=k} \boldsymbol{\beta}_{c_k}\Big) \otimes \exp(\boldsymbol{b}_i + \boldsymbol{\kappa}_{ih}) \tag{3}$$

Here, $\boldsymbol{\mu}$ is the P-vector of fixed effects, $\boldsymbol{\beta}_{c_k}$ is the P-vector of fixed effects for category k of covariate c ($\boldsymbol{\beta}_{c_k} = \mathbf{0}$ if k is the category of reference) on the P parameters of the structural model. $\boldsymbol{b}_i$ is the P-vector of random effects of individual i, $\boldsymbol{\kappa}_{ih}$ is the P-vector of random effects of individual i for period h.

c_{ih} can take any value k in the set K_c. $1_{c_{ih}=k}$ is the indicator function of the category k in the set K_c, having the value 1 if c_{ih} is in the category k, and having the value 0 if not. Consequently, $1_{c_{ih}=k}\boldsymbol{\beta}_{c_k}$ is equal to $\boldsymbol{\beta}_{c_k}$ if $1_{c_{ih}=k} = 1$ and to $\mathbf{0}$ if $1_{c_{ih}=k} = 0$. exp with a vector $\boldsymbol{v}$ as its argument is the vector such as each element is the exponential of the corresponding element of $\boldsymbol{v}$. In (2), the P-vector $\boldsymbol{\phi}_{ih}$ is defined by addition of the vectors on the right side of the equation. In (3), the symbol $\otimes$ here means that each element p ($p = 1, \ldots, P$) of $\boldsymbol{\phi}_{ih}$ is obtained by multiplication of the p^{th} elements ($p = 1, \ldots, P$) of the vectors $\boldsymbol{\mu}$ and $\exp(\sum_{c \in C} \sum_{k \in K_c} 1_{c_{ih}=k} \boldsymbol{\beta}_{c_k})$ and $\exp(\boldsymbol{b}_i + \boldsymbol{\kappa}_{ih})$.

It is assumed that $\boldsymbol{b}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$ and $\boldsymbol{\kappa}_{ih} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma})$, with $\boldsymbol{\Omega}$ and $\boldsymbol{\Gamma}$ defined as the diagonal variance-covariance matrices of size $P \times P$. Each element ω_p of $\boldsymbol{\Omega}$ and γ_p of $\boldsymbol{\Gamma}$ ($p = 1, \ldots, P$) represents, respectively, the variance of BSV for the p^{th} component of $\boldsymbol{b}_i$ and the variance of WSV

for the p^{th} component of $\boldsymbol{\kappa}_{ih}$. $\boldsymbol{b}_i$ and $\boldsymbol{\kappa}_{ih}$ are supposed to be independent. We denote by $\boldsymbol{\Omega}^*$ the diagonal matrix

$$\boldsymbol{\Omega}^* = \begin{pmatrix} \boldsymbol{\Omega} & \\ & \boldsymbol{\Gamma}^* \end{pmatrix} \tag{4}$$

where $\boldsymbol{\Gamma}^*$ is a block diagonal matrix composed of H blocks $\boldsymbol{\Gamma}$. The relation between $\boldsymbol{\phi}_{ih}$ and $(\boldsymbol{\mu}, \boldsymbol{\beta}, \boldsymbol{b}_i, \boldsymbol{\kappa}_{ih})$, with $\boldsymbol{\beta}$ regrouping all the fixed effects for discrete covariates, is summarised in a function g such as $\boldsymbol{\phi}_i = g(\boldsymbol{\theta}, \boldsymbol{\nu}_i)$ where $\boldsymbol{\phi}_i = (\boldsymbol{\phi}'_{i1}, ..., \boldsymbol{\phi}'_{iH})'$ is the vector of parameters for individual i for H periods, $\boldsymbol{\theta} = (\boldsymbol{\mu}', \boldsymbol{\beta}')'$ is the vector of all fixed effects, and $\boldsymbol{\nu}_i = (\boldsymbol{b}'_i, \boldsymbol{\kappa}'_{i1}, ..., \boldsymbol{\kappa}'_{iH})'$ is the vector of all random effects for H periods. The vector of observations of individual i for H periods, $\boldsymbol{y}_i = (\boldsymbol{y}'_{i1}, ..., \boldsymbol{y}'_{iH})'$ can thus be written

$$\boldsymbol{y}_i = f(g(\boldsymbol{\theta}, \boldsymbol{\nu}_i), \boldsymbol{\xi}_i) + \boldsymbol{\epsilon}_i \tag{5}$$

where $\xi_i = (\xi_{i1}, ..., \xi_{iH})$ and $\boldsymbol{\epsilon}_i = (\boldsymbol{\epsilon}'_{i1}, ..., \boldsymbol{\epsilon}'_{iH})' \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_i(\boldsymbol{\theta}, \boldsymbol{\nu}_i, \sigma_{inter}, \sigma_{slope}, \xi_i))$. $\boldsymbol{\Sigma}_i$ is an $n_i \times n_i$ diagonal matrix composed of each diagonal element of $\boldsymbol{\Sigma}_{ih}$ ($h = 1, \ldots, H$). The following assumptions are made: $\boldsymbol{\epsilon}_i|\boldsymbol{\nu}_i$ are independent between subjects, $\boldsymbol{\epsilon}_{ih}|\boldsymbol{\kappa}_{ih}$ are independent between periods, and $\boldsymbol{\epsilon}_i$ and $\boldsymbol{\nu}_i$ are independent for each subject.

Let $\boldsymbol{\lambda} = (\omega_1, ..., \omega_P, \gamma_1, ..., \gamma_P, \sigma_{inter}, \sigma_{slope})'$ be the vector of variance terms, and $\boldsymbol{\Psi}$ be the vector of all population parameters to be estimated, so that $\boldsymbol{\Psi} = (\boldsymbol{\theta}', \boldsymbol{\lambda}')'$. We also denote by $dim(\boldsymbol{\theta})$ the number of fixed effects and $dim(\boldsymbol{\lambda})$ the number of variance terms.

2.2 Population Fisher information matrix

We extend the expression of $\boldsymbol{M}_F$ in NLMEM for crossover trials using the first order Taylor expansion of the model as in Mentré *et al* [20]. For simplicity, we omit the index i for the individual in this section. The elementary $\boldsymbol{M}_F(\boldsymbol{\Psi}, \boldsymbol{\xi})$ for an individual with elementary design $\boldsymbol{\xi}$ composed of one or several periods is calculated by

$$\boldsymbol{M}_F(\boldsymbol{\Psi}, \boldsymbol{\xi}) = E\left(\frac{-\partial^2 l(\boldsymbol{\Psi}, \boldsymbol{y})}{\partial \boldsymbol{\Psi} \partial \boldsymbol{\Psi}'}\right) \tag{6}$$

where $l(\boldsymbol{\Psi}, \boldsymbol{y})$ is the log-likelihood of the vector of observations $\boldsymbol{y}$ of that individual for the population parameters $\boldsymbol{\Psi}$. There is no analytical expression for $l(\boldsymbol{\Psi}, \boldsymbol{y})$ because of the nonlinearity of the structural model f. An approximation is possible using the first-order Taylor expansion of $f(g(\boldsymbol{\theta}, \boldsymbol{\nu}), \boldsymbol{\xi})$ around the expectation of the random effects $\boldsymbol{\nu}$, i.e. $\boldsymbol{0}$. The statistical model can be written as

$$\boldsymbol{y} \cong f(g(\boldsymbol{\theta}, \boldsymbol{0}), \boldsymbol{\xi}) + \left(\frac{\partial f'(g(\boldsymbol{\theta}, \boldsymbol{0}), \boldsymbol{\xi})}{\partial \boldsymbol{\nu}}\right) \boldsymbol{\nu} + \boldsymbol{\epsilon} \tag{7}$$

and the log-likelihood l is then approximated by

$$-2l(\boldsymbol{\Psi}, \boldsymbol{y}) \cong nln2\pi + ln|\boldsymbol{V}| + (\boldsymbol{y} - \boldsymbol{E})'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{E}) \tag{8}$$

where n is the total number of observations of the individual; $\boldsymbol{E}$ and $\boldsymbol{V}$ are the approximated marginal expectation and variance of $\boldsymbol{y}$ given by

$$\boldsymbol{E} = E(\boldsymbol{y}) \cong f(g(\boldsymbol{\theta}, \boldsymbol{0}), \boldsymbol{\xi}) \tag{9}$$

$$\boldsymbol{V} = Var(\boldsymbol{y}) \cong \left(\frac{\partial f'(g(\boldsymbol{\theta},\boldsymbol{0}),\boldsymbol{\xi})}{\partial \boldsymbol{\nu}}\right) \boldsymbol{\Omega}^* \left(\frac{\partial f(g(\boldsymbol{\theta},\boldsymbol{0}),\boldsymbol{\xi})}{\partial \boldsymbol{\nu}'}\right) + \boldsymbol{\Sigma}(\boldsymbol{\theta},\boldsymbol{0},\sigma_{inter},\sigma_{slope},\boldsymbol{\xi}) \quad (10)$$

We derive the expression of the log-likelihood l, taking into account WSV in addition to BSV and fixed effects for discretes covariates changing between periods. We obtain the elementary $\boldsymbol{M}_F$ depending on $\boldsymbol{E}$ and $\boldsymbol{V}$. Assuming that $\boldsymbol{V}$ does not or does depend on the fixed effects, the elementary $\boldsymbol{M}_F$ is a block diagonal matrix or a full matrix. There is no clear consensus on what is the best approximation but here in our approach by linearisation, we assume the choice of the block diagonal expression [31]:

$$\boldsymbol{M}_F(\boldsymbol{\Psi},\boldsymbol{\xi}) = \frac{1}{2}\begin{pmatrix} \boldsymbol{A}(\boldsymbol{E},\boldsymbol{V}) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{B}(\boldsymbol{E},\boldsymbol{V}) \end{pmatrix} \quad (11)$$

where

$$(\boldsymbol{A}(\boldsymbol{E},\boldsymbol{V}))_{rs} = 2\frac{\partial \boldsymbol{E}'}{\partial \boldsymbol{\theta}_r}\boldsymbol{V}^{-1}\frac{\partial \boldsymbol{E}'}{\partial \boldsymbol{\theta}_s} \quad (12)$$

$$(\boldsymbol{B}(\boldsymbol{E},\boldsymbol{V}))_{ml} = tr\left(\frac{\partial \boldsymbol{V}}{\partial \boldsymbol{\lambda}_m}\boldsymbol{V}^{-1}\frac{\partial \boldsymbol{V}}{\partial \boldsymbol{\lambda}_l}\boldsymbol{V}^{-1}\right) \quad (13)$$

with r and $s = 1, ..., dim(\boldsymbol{\theta})$; m and $l = 1, ..., dim(\boldsymbol{\lambda})$.

To define the population matrix, we use the fact that $\boldsymbol{M}_F(\boldsymbol{\Psi},\boldsymbol{\Xi})$ for a population design $\boldsymbol{\Xi}$ is defined by the sum of the N elementary matrices $\boldsymbol{M}_F(\boldsymbol{\Psi},\boldsymbol{\xi}_i)$, so that

$$\boldsymbol{M}_F(\boldsymbol{\Psi},\boldsymbol{\Xi}) = \sum_{i=1}^{N} \boldsymbol{M}_F(\boldsymbol{\Psi},\boldsymbol{\xi}_i) \quad (14)$$

In the case of a limited number Q of elementary designs, we have

$$\boldsymbol{M}_F(\boldsymbol{\Psi},\boldsymbol{\Xi}) = \sum_{q=1}^{Q} N_q \boldsymbol{M}_F(\boldsymbol{\Psi},\boldsymbol{\xi}_q). \quad (15)$$

2.3 Computation of power and number of subjects needed

In this section, we consider the Wald tests of comparison and equivalence on a discrete covariate effect fixed or changing between periods for the case of a crossover trial. Here, these tests are performed on the estimator of the effect $\beta_{c_k p}$ of each category k of each covariate c for parameter p, noted $\widehat{\beta_{c_k p}}$. Given a value β_1 of $\widehat{\beta_{c_k p}}$, and from the square root of the diagonal terms of the inverse of $\boldsymbol{M}_F$, we calculate the standard error of β_1. Then, we predict the power of the tests and also compute the NSN to achieve a given power. Note that for simplicity, we omit index $c_k p$ for β in this section.

2.3.1 Comparison test

The null hypothesis for the Wald test of comparison is $H_0 : \{\beta = 0\}$ while the alternative hypothesis is $H_1 : \{\beta \neq 0\}$. Putting $\widehat{\beta}$ as the estimator of β, then the statistic of the Wald test under the

null hypothesis H_0 is $W = \widehat{\beta}/\mathrm{SE}(\widehat{\beta})$. With a type I error α, H_0 is rejected if $|W| > z_{1-\alpha/2}$, where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution.

We compute the power of the Wald test of comparison under H_1, when $\widehat{\beta} = \beta_1 \neq 0$. With a given design and values of population parameters, we predict $SE(\beta_1)$ using the extension of $\boldsymbol{M}_F$. Under H_1, W is asymptotically distributed with a normal distribution centred at $\beta_1/\mathrm{SE}(\beta_1)$. Therefore, the power of the Wald test is

$$P_{\text{comp}} = 1 - \Phi\left(z_{1-\alpha/2} - \frac{\beta_1}{\mathrm{SE}(\beta_1)}\right) + \Phi\left(-z_{1-\alpha/2} - \frac{\beta_1}{\mathrm{SE}(\beta_1)}\right) \tag{16}$$

where Φ is the cumulative distribution function of the standard normal distribution.

Using the predicted SE by the extension of $\boldsymbol{M}_F$, we can also derive the NSN to achieve a power P_{comp} to detect a covariate effect as in [30]. For that, we first compute the SE needed on β_1 to obtain the power of P_{comp}, called $\mathrm{SEN}(P_{\text{comp}})$, using the following relation

$$\mathrm{SEN}(P_{\text{comp}}) = \frac{\beta_1}{z_{1-\alpha/2} - \Phi^{-1}(1 - P_{\text{comp}})} \tag{17}$$

We then compute the NSN to obtain a power of P_{comp}, called $\mathrm{NSN}(P_{\text{comp}})$ using

$$\mathrm{NSN}(P_{\text{comp}}) = N \times \left(\frac{\mathrm{SE}(\beta_1)}{\mathrm{SEN}(P_{\text{comp}})}\right)^2 \tag{18}$$

2.3.2 Equivalence test

The null hypothesis for the Wald test of equivalence is $H_0 : \{\beta \leq -\delta \text{ or } \beta \geq +\delta\}$ while the alternative hypothesis is $H_1 : \{-\delta < \beta < +\delta\}$, where δ is the equivalence limit. In the guidelines for equivalence assessment [1, 2], it is recommended that $\delta = 0.2$. H_0 is composed of two unilateral hypotheses $H_{0,-\delta} : \{\beta \leq -\delta\}$ and $H_{0,+\delta} : \{\beta \geq +\delta\}$. Equivalence between two groups on covariate effect β can be concluded if and only if the two hypotheses $H_{0,-\delta}$ and $H_{0,+\delta}$ are rejected. The equivalence Wald test using NLMEM has already been developed [32, 12], based on the Schuirmann's two one-sided tests (TOST) [33]. Denoting the estimator of β by $\widehat{\beta}$, the following formulas define $W_{-\delta}$, the statistic of the unilateral test under the null hypothesis $H_{0,-\delta}$, and $W_{+\delta}$, the statistic of the unilateral test under the null hypothesis $H_{0,+\delta}$: $W_{-\delta} = (\widehat{\beta} + \delta)/\mathrm{SE}(\widehat{\beta})$ and $W_{+\delta} = (\widehat{\beta} - \delta)/\mathrm{SE}(\widehat{\beta})$. With a type I error α, H_0 is rejected if $W_{-\delta} \geq z_{1-\alpha}$ and $W_{+\delta} \leq -z_{1-\alpha}$, where $z_{1-\alpha}$ is the $(1-\alpha)$ quantile of the standard normal distribution.

We compute the power of the Wald test of equivalence under H_1, when $\widehat{\beta} = \beta_1 \in [-\delta, +\delta]$. Usually we choose β_1 = 0. From a given design and values of population parameters, we predict $SE(\beta_1)$ using the extension of $\boldsymbol{M}_F$. The power of the Wald test is then

$$P_{\text{equi}} = 1 - \Phi\left(z_{1-\alpha} - \frac{\beta_1 + \delta}{\mathrm{SE}(\beta_1)}\right) \qquad \text{if } \beta_1 \in [-\delta, 0] \tag{19}$$

$$P_{\text{equi}} = \Phi\left(-z_{1-\alpha} - \frac{\beta_1 - \delta}{\mathrm{SE}(\beta_1)}\right) \qquad \text{if } \beta_1 \in [0, +\delta] \tag{20}$$

where Φ is defined as in Section 2.3.1. As expressed in (19) and (20), the power of the equivalence test depends on the sign of β_1. When $\beta_1 = 0$, we can use any of the two equations to obtain the power because in that case:

$$P_{\text{equi}} = 1 - \Phi\left(z_{1-\alpha} - \frac{\delta}{\mathrm{SE}(\beta_1)}\right) = \Phi\left(-z_{1-\alpha} - \frac{(-\delta)}{\mathrm{SE}(\beta_1)}\right) \tag{21}$$

Using the predicted SE by the extension of $\boldsymbol{M}_F$, we can also derive the NSN to achieve the power P_{equi} of showing the equivalence between two groups on covariate effect β_1. For that, we first compute the SE needed on β_1 to obtain the power of P_{equi}, called SEN(P_{equi}), using

$$\text{SEN}(P_{\text{equi}}) = \frac{-\beta_1 - \delta}{-z_{1-\alpha} + \Phi^{-1}(1 - P_{\text{equi}})} \qquad \text{if } \beta_1 \in [-\delta, 0] \tag{22}$$

$$\text{SEN}(P_{\text{equi}}) = \frac{-\beta_1 + \delta}{z_{1-\alpha} + \Phi^{-1}(P_{\text{equi}})} \qquad \text{if } \beta_1 \in [0, +\delta] \tag{23}$$

We then compute the NSN to obtain the power of P_{equi}, called NSN(P_{equi}) using (18) with SEN(P_{equi}) instead of SEN(P_{comp}). If $\beta_1 < 0$, we substitute (22) into (18); if $\beta_1 > 0$, we substitute (23) into (18); if $\beta_1 = 0$, we can substitute either of (22) or (23) into (18) because in that case:

$$\text{SEN}(P_{\text{equi}}) = \frac{-\delta}{-z_{1-\alpha} + \Phi^{-1}(1 - P_{\text{equi}})} = \frac{\delta}{z_{1-\alpha} + \Phi^{-1}(P_{\text{equi}})} \tag{24}$$

3 Evaluation by simulations

The simulation model and design are based on a previous PK study DAV1 of amoxicillin in piglets, conducted by the biotechnology company Da Volterra (http://www.davolterra.com) which will be given in detail in Section 4.1.

3.1 Simulations

The PK model is a one compartment oral model with first order absorption and first order linear elimination. The observed amoxicillin concentration of subject i at sampling time j of period h is modelled by a NLMEM as described in Section 2.1.2:

$$y_{ihj} = f(\boldsymbol{\phi}_{ih}, t_{ihj}) + \epsilon_{ihj} = \frac{F\ D\ ka_{ih}}{Cl_{ih} - V_{ih}\ ka_{ih}}[e^{-ka_{ih}\ t_{ihj}} - e^{-(Cl_{ih}/V_{ih})\ t_{ihj}}] + \epsilon_{ihj} \tag{25}$$

In (25), D is the dose, F the bioavailability, ka the absorption rate constant, Cl the clearance of elimination of the drug and V the volume of distribution. After oral administration only, the bioavailability can not be estimated and the vector $\boldsymbol{\phi}$ of PK parameters is $(ka, V/F, Cl/F)$. For simplicity, this will be denoted subsequently by (ka, V, Cl) where Cl and V are apparent parameters.

We simulate crossover PK trials with two periods, one sequence as in DAV1 study. In each dataset, N = 40 subjects are allocated at period 1 to treatment 1, with drug A plus placebo and then, at period 2, to treatment 2, with the same drug A plus drug B. All compounds are administrated orally once a day. The dose is fixed at 30 mg.kg^{-1} for all subjects. The sampling times for all subjects and both periods are similar and are taken from among the times of the DAV1 study: 0.5, 1, 1.5, 2, 4, 6, and 8 h after dosing. The vector of fixed effects $\boldsymbol{\mu}$ is composed of (μ_{ka} = 1.00 h^{-1}, μ_V = 3.5 L.kg^{-1}, μ_{Cl} = 2 $\text{L.h}^{-1}\text{.kg}^{-1}$) for treatment 1 (the treatment of reference). We take into account a treatment effect β_{Cl} on Cl, characterising the interaction between A and B. The relation

between these parameters is expressed as in (3), so that at period 1,

$$\boldsymbol{\phi}_{i1} = \begin{pmatrix} \mu_{ka} \times \exp(b_i^{ka} + \kappa_{i1}^{ka}) \\ \mu_V \times \exp(b_i^V + \kappa_{i1}^V) \\ \mu_{Cl} \times \exp(b_i^{Cl} + \kappa_{i1}^{Cl}) \end{pmatrix} \tag{26}$$

and at period 2,

$$\boldsymbol{\phi}_{i2} = \begin{pmatrix} \mu_{ka} \times \exp(b_i^{ka} + \kappa_{i2}^{ka}) \\ \mu_V \times \exp(b_i^V + \kappa_{i2}^V) \\ \mu_{Cl} \times \exp(\beta_{Cl}) \times \exp(b_i^{Cl} + \kappa_{i2}^{Cl}) \end{pmatrix} \tag{27}$$

We fix the variance of BSV $\omega = (0.3)^2 = 0.09$ and of WSV $\gamma = (0.15)^2 = 0.0225$ for the three parameters. We assume no correlation between PK parameters, so the variance-covariance matrices $\boldsymbol{\Omega}$ and $\boldsymbol{\Gamma}$ are diagonal. The parameters of the random error model are σ_{inter} = 0.1 $\mathrm{ng.L^{-1}}$ and σ_{slope} = 0. For each subject i of each simulated trial at period h, we simulate a vector of random effects $\boldsymbol{b}_i$ in $\mathcal{N}(0, \boldsymbol{\Omega})$ and two vectors of random effects $\boldsymbol{\kappa}_{ih}$ in $\mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma})$, one for each period and then compute $\boldsymbol{\phi}_{ih}$ while taking into account the treatment effect β_{Cl} using (26, 27). We calculate the concentrations predicted by the PK model at each sampling time. For each predicted concentration, we generate the vector of the random error using the value of σ_{inter}. To get the simulated concentrations, these errors are added to the predicted concentrations. Then, we left-censored them at 0.05 $\mathrm{ng.L^{-1}}$, the limit of quantification (LOQ) in the DAV1 study.

We consider in the above simulation procedure different values of the treatment effect parameter β_{Cl} such as $\log(0.8), \log(1), \log(1.1), \log(1.2), \log(1.25)$ and $\log(1.5)$. The case $\beta_{Cl} = \log(1) = 0$ corresponds to the null hypothesis of the comparison test (no treatment effect) while the cases $\beta_{Cl} = \log(0.8) \approx -0.2$ or $\log(1.25) \approx 0.2$ correspond to the null hypothesis of the equivalence test (the two limits of equivalence). We also derive and study an optimal sparse design of four samples by optimising the rich one using the new extension of $\boldsymbol{M}_F$ with the Fedorov-Wynn algorithm [34, 35]. For that, we use the same simulated values above for a model without treatment effect then with different values of treatment effect parameter. The optimal sparse design obtained is the same with or without treatment effect (0.5, 2, 6, 8 h after dosing). For each simulated scenario, we simulate 1000 trials, using R 2.4.0.

3.2 Evaluation methods

Our aim is to evaluate and compare the predictions of the standard errors and test powers computed by the extension of $\boldsymbol{M}_F$ to the empirical ones obtained from simulations. For that, first, to obtain these predictions, with the design and the population parameter values of each simulated scenario above, we compute the standard errors denoted by SE of each population parameter from the square root of the diagonal terms of the extended $\boldsymbol{M}_F^{-1}$ or the relative SE denoted by RSE, defined as the ratio of SE by the parameter simulated value. Then, from the SE of treatment effect parameter, we also predict the power of the Wald test of comparison and equivalence. We consider a type I error $\alpha = 0.05$ and an equivalence limit $\delta = 0.2$.

In parallel, the empirical values are calculated as follows. Each simulated data file of each scenario with different values of β_{Cl} is analysed in MONOLIX 2.4 (http://www.monolix.org) by an NLMEM including discrete covariate. We use the SAEM algorithm [36, 37, 38, 39] implemented in this software to estimate parameters. The data under LOQ are taken into account in the analysis using the extension of the SAEM algorithm proposed by Samson *et al* [40], which allows of

handling the left-censored data in NLMEM as an exact Maximum Likelihood estimation method. We then calculate the empirical SE, defined as the sample estimate of the standard deviation from the parameter estimates obtained with SAEM. Concerning the powers, the observed power of the Wald test of comparison and equivalence are calculated as the proportion of trials for which the null hypothesis is rejected. Finally, we compare the results obtained from simulations to the predictions.

We present in Section 3.3.1 the comparison of the predicted RSE to the empirical ones in each simulated scenario. As the treatment effect is added on Cl, here are presented only the results for the fixed effect, variances of BSV and WSV on Cl, as well as the for the parameter σ_{inter} of the error model. Then, we focus on the comparison of the SE of treatment effect β_{Cl}. The relevance of this power computation using the extension of $\boldsymbol{M}_F$ will be investigated in Section 3.3.2.

3.3 Results

3.3.1 Evaluation of the predicted standard errors

The RSE predicted by the extension of $\boldsymbol{M}_F$ and the empirical ones calculated from simulations are displayed in Figure 1. The predicted RSE are very close to the empirical ones in each simulated scenario. Some RSE are slightly underestimated by $\boldsymbol{M}_F$, especially for γ_{Cl}, but the difference is not clinically relevant ($< 10\%$). As expected, the RSE are slightly higher in the sparse design than in the rich design, but the differences are very small, except for σ_{inter}.

We now focus on the SE of treatment effect β_{Cl}. The distributions of the observed SE from each simulated scenario are reported as boxplots in Figure 2. The SE(β_{Cl}) predicted by the extension of $\boldsymbol{M}_F$ are very close to the empirical ones. The SE(β_{Cl}) are slightly higher in the rich design compared to the sparse one but they remain very close. These values are reported in Table I. There is no important change of SE(β_{Cl}) when varying β_{Cl}.

Although not all the predicted SE and RSE are reported here, some of these values are present in the online supplementary evaluation output of the rich design [1] with $\beta_{Cl} = log(1.1)$. In this example and for other scenarios including the ones with optimal designs as well, the RSE predicted for the fixed effects are very close to each other and close to 5%. The minimum values of RSE of the log parameters can be approximated by the SE, in the case of a rich design, as $\sqrt{\omega}/\sqrt{N}$ (ω being here the variance of BSV). Here, $\sqrt{\omega}/\sqrt{N}$ are roughly 4.9% since we have fixed the same ω for the three PK parameters (ka, V, Cl). This means that the original design (seven samples) is quite rich and the derived optimal design (four samples) is a good design, allowing of estimating well the parameters with good precision.

3.3.2 Evaluation of the predicted powers

With each scenario, the predicted power by the extension of $\boldsymbol{M}_F$ and the observed power from simulations for the comparison test as well as for the equivalence test are reported in Table II and also as barplots in Figure 3. Note that when $\exp(\beta_{Cl})$ = 1 for the comparison test and when $\exp(\beta_{Cl})$ = 0.8 or $\exp(\beta_{Cl})$ = 1.25 for the equivalence test, we are under the null hypothesis so the probability computed is not the test power but the type I error. The predicted powers by the extension of $\boldsymbol{M}_F$ and the observed powers from simulations are very close for both comparison and equivalence tests. The powers obtained with the optimal sparse design are also close to the ones obtained with the rich design.

[1] Supporting information may be found in the online version of this article.

4 Application to a pharmacokinetic study in piglets

In this section, we apply the extension of $\boldsymbol{M}_F$ to design a subsequent study DAV2 of amoxicillin in piglets based on the results of the previous study DAV1 conducted by the company Da Volterra that inspired the simulations. DAV2 will be a one-way crossover trial. The piglets receive at period 1 amoxicillin plus placebo, at period 2 amoxicillin plus a compound X developed by Da Volterra. The objective of DAV2 is to show the absence of interaction of X on the AUC of amoxicillin concentrations. The direct relation $AUC = dose/Cl$ implies the equality between the log parameter of the treatment effect on AUC and the one on Cl [12, 13]. That is why we are interested particularly in Cl. To design DAV2, we analyse the data of DAV1 then with the results obtained, we use the extension of $\boldsymbol{M}_F$ to evaluate/optimise designs and to compute the NSN to show the absence of interaction of X on Cl of amoxicillin with a power of 90%.

4.1 Analysis of DAV1 study

4.1.1 Materials and methods

The DAV1 study is a one-way, two-period crossover trial with 16 piglets receiving amoxicillin and placebo at period 1 and only amoxicillin at period 2. Amoxicillin was administrated orally with a single dose of 30 $\mathrm{mg.kg}^{-1}$. Plasma amoxicillin concentrations were collected at same sampling times at each period: 0.5, 1, 1.5, 2, 4, 6, 8, 10 and 12 h. Figure 4 represents the individual amoxicillin concentrations ($\mathrm{ng.L}^{-1}$) versus time of each period.

To describe the pharmacokinetics of amoxicillin, we analyse jointly the data of both periods for all piglets using MONOLIX 2.4 with an NLMEM and the SAEM algorithm for estimating parameters. At 8, 10 and 12 h, respectively, 21.9%, 68.8% and 87.55% of observations of both periods are below LOQ (0.05 $ng.L^{-1}$). The very high proportions of data below LOQ especially at 10 h and 12 h justify the choice of excluding these two samples in the simulations previously. Here, MONOLIX 2.4 allows of taking into account the data below LOQ [40]. A one compartment oral model with first order absorption, first order linear elimination and the presence of a lag time adequately describes the amoxicillin concentrations in this study. The fixed effects are then $(tlag, ka, CL/F, V/F)$ but denoted for simplicity by $(tlag, ka, Cl, V)$. A treatment effect β_{Cl} is added on $\log(Cl)$ and is estimated. We consider log parameters and assume the choice of diagonal matrices $\boldsymbol{\Omega}$ and $\boldsymbol{\Gamma}$. To choose the error model, we compare the Bayesian Information Criteria (BIC) between the homoscedastic (σ_{slope} = 0), proportional (σ_{inter} = 0), and combined models. We examine the SAEM convergence graph and the goodness-of-fit plots to evaluate the chosen model.

4.1.2 Results

We choose a homoscedastic error model with σ_{slope} = 0. With respect to the model of random effects, when estimating ω_{tlag} and ω_{Cl}, the obtained variabilities are very low and their RSE very high. Consequently, we fix $\omega_{tlag} = \omega_{Cl}$ = 0 for $\boldsymbol{\Omega}$ and do not estimate these two variances. With the chosen model, the algorithm converges and the goodness-of-fit plots are satisfactory. The population parameter estimates with this model are reported in Table III. Because the dose are in $\mathrm{mg.kg}^{-1}$, the unit of the parameters Cl and V are also expressed by their usual units per kg. The estimation of most of these parameters is good because the RSE are satisfactory except for ω_{ka} and γ_{ka} of which the RSE are, respectively, 144% and 121%. We note that these variabilities are low and are consequently difficult to be precisely estimated. With $\widehat{\beta_{Cl}}$ = 0.06 and $SE(\widehat{\beta_{Cl}})$ = 0.16, we perform,

using the formulas in Section 2.3, the Wald tests of equivalence with the equivalence limit δ = 0.2 and the type I error α = 0.05. The 90% confidence interval for treatment effect on Cl is [-0.198; 0.318] and not included in [-0.2, 0.2]. Consequently, we can not show a significant equivalence between the two treatment groups.

4.2 Designing of DAV2 study

4.2.1 Methods

We evaluate and optimise the design for the DAV2 study, using the model selected and the values of population parameters estimated from the DAV1 study, with the estimated treatment effect = 0.06 (cf. Table III). We first evaluate for the next study a rich design with two periods which is the original design of the DAV1 study (0.5, 1, 1.5, 2, 4, 6, 8 h after dosing) with same sampling times at each period. We do not consider the samples 10 h and 12 h because of the high proportions of data below LOQ as there is no way presently to incorporate this information in the computation of $\boldsymbol{M}_F$. Consequently, this population design involves a total number of 224 samples to be performed in 16 piglets. We omit subsequently the parameter $tlag$ because its estimated value is small (roughly 0.37 h) whereas the first sampling time of the study is given to be after 0.5 h. We also derive a two-period optimal design with a constraint of four points for each period (same samples at each period) from the rich one using the extension of $\boldsymbol{M}_F$ and the Fedorov-Wynn algorithm as in the simulations. Thus, the optimal design involves a total number of 128 samples to be performed in 16 piglets. We evaluate the values of SE and RSE for all parameters and examine the critetion value computed from the extension of $\boldsymbol{M}_F$, defined as the determinant of $\boldsymbol{M}_F$ normalised by the inverse of the total number of parameters to be estimated with NLMEM (the size of $\boldsymbol{M}_F$). The ratio of the two criteria expresses the relative efficiency between the two designs and evaluates the factor of the mean estimation variance decrease when changing the design.

Then, we compute, with the extension of $\boldsymbol{M}_F$, the predicted power of the Wald test assessing the equivalence on the parameter Cl between the two groups: amoxicillin + placebo (AP) group (treatment 1) versus amoxicillin + compound X (AX) group (treatment 2) from the expected SE of β_{Cl} and with the same number of piglets (N= 16) as in DAV1. We examine different given values of β_{Cl}: 0 as is usually done in equivalence assessment, 0.06 as in DAV1 and 0.1, for the rich design as well as for the sparse design. We predict the NSN to achieve a power of 90% for the equivalence test with an equivalence limit of 0.2. The computations are first performed for a design with two periods AP/AX then with four periods AP/AX/AP/AX.

4.2.2 Results

The evaluation results for the two-period design of DAV2 study are reported in Table IV with the predictions of the RSE (or SE for β_{Cl}) for each parameter and the criterion value of both the original and optimal design composed of (0.5, 2, 4, 6 h) at each period. The original design gives correct RSE and SE for all parameters except for ω_{ka} and γ_{ka} as in DAV1. We note also that the RSE and SE of the sparse design are slightly higher than those given by the rich design but they are still very close. The criterion value decreases slightly when there are less sampling times per subject. The efficiency of the sparse design compared to the rich one is calculated as the ratio of the two criterion values 32.71/39.04 = 0.82. Concequently, the choice of similar samples as DAV1 is acceptable for DAV2, giving the correct SE for most of the population parameters. The sparse design is efficient while allowing us to perform fewer samples per subject.

The results for power and NSN computation are reported in Table V. We observe a lack of power for the design with two periods and an increase of power for four periods to show equivalence between two treatment groups with 16 subjects per group. Indeed, with β_{Cl} = 0.06 as estimated in the DAV1 study and a rich design, we predict 128 subjects in a two-period trial (involving a total number of 1792 samples to be performed) and 67 subjects in a four-period trial (1876 samples) needed to achieve a power of 90%. With the sparse design, we predict 131 subjects in a two-period trial (1048 samples) and 69 subjects in a four-period trial (1104 samples) needed to achieve a power of 90%.

5 Discussion

In this paper, we proposed the expression of $\boldsymbol{M}_F$ for crossover trials analysed with NLMEM including WSV, in addition to BSV, and with discrete covariates which can change between periods. It is an extension of the developments proposed in [20], [29] and [21], using a linearisation of the model around the expectation of random effects. Simulations with various scenarios showed the correctness of the predictions of the SE and power for the comparison or equivalence Wald tests computed by $\boldsymbol{M}_F$. In spite of a slight underestimation of RSE by $\boldsymbol{M}_F$ for some parameters, which respects the Rao-Cramer inequality, the discrepancy is not important ($< 10\%$); indeed, when experimenters design population studies, it is more the amplitude of RSE than the exact value that they are looking for. This extension of $\boldsymbol{M}_F$ can be concluded as relevant while saving lots of computation time compared to extensive simulations. We also derived an optimal sparse design and showed that it had power very close to that of the rich design, even with fewer samples per subject. This can be explained by the fact that there are only three parameters in the PK model, so a lower number of parameters compared to the number of samples in the sparse design (four samples per period). It is also what we obtained in our application example, which showed the importance of the number of subjects for the power of test. We notice that in the application, the covariate effect β_{Cl} was estimated from the data of the DAV1 study between two periods although in that design, the piglets received amoxicillin plus placebo at the first period and amoxicillin alone at the second period. So β_{Cl} expressed something other than a treatment effect, perhaps a period or a placebo effect. We chose however to include this discrete covariate in the statistical model as a treatment effect and to estimate this parameter because our objective in this application is to design the subsequent study DAV2 where there will really be a treatment effect from the compound X co-administrated with amoxicillin at the second period and similarly we assume that there will be no effect of drug X (not more than placebo). Then as presented in section 4.2, we examined not only the estimated value of β_{Cl} as in DAV1 but also different ones to predict the power and NSN for DAV2. In this application, more than 16 piglets were needed in a one-way, two-period crossover design to show an absence of interaction of compound X on the PK of amoxicillin, which can be explained by the important standard deviation of WSV on Cl ($\sqrt{\gamma} = 0.45$). Increasing the number of periods improved the power of the study because the variance of the treatment effect parameter was reduced.

The extension of $\boldsymbol{M}_F$ provides a useful tool to study the influence of the number of periods, of subjects or the effect size of a discrete covariate. We can also easily examine how the power would vary with different ratios between the BSV and WSV in a crossover design, and to compare these results to those obtained from a parallel group design. This emphasises the advantage of crossover design compared to parallel group design in bioequivalence trials except when WSV is the total variability and BSV = 0. It would also be interesting to include in the model a period or sequence effect and to consider different ratios between BSV and WSV to see its influence on the results.

These extensions were implemented in the new version 3.2 of PFIM, available freely since January 2010 at www.pfim.biostat.fr, with several new features in terms of model specification and of $\boldsymbol{M}_F$ expression. An output example of PFIM 3.2 for design evaluation of a two-way, two-period crossover trial is shown in the Supplementary Material [2]. This new mathematical development of $\boldsymbol{M}_F$ is applicable to single or multiple response models, for any number of periods and sequences, with different designs at each period. However, the present implementation of $\boldsymbol{M}_F$ in PFIM was performed only for the case of trials with the same sampling times in each period, which is the design of our motivating example.

In our present work, we omitted the parameter $tlag$ when designing the DAV2 study. But the fact that $tlag$ was not taken into account in design optimisation has no important influence on the optimal design, all the more justified as its value is very small ($\approx$ 0.37h), and before the first sampling time (0.5h) of the study. However, it would be useful, when the value of $tlag$ is important, to evaluate $\boldsymbol{M}_F$ in the presence of this parameter. In addition, data below LOQ were omitted for design evaluation in this extension but it would be interesting to derive an expression of $\boldsymbol{M}_F$ including these left-censored data. An ad hoc method would be to take into account the percentage of data below LOQ at each sampling time, evaluated by simulation, in the evaluation of $\boldsymbol{M}_F$. It would be also interesting to consider the contribution of this information to the likelihood as proposed in [40]. Also, here we considered only diagonal $\boldsymbol{\Omega}$ and $\boldsymbol{\Gamma}$ matrices but one may want to allow correlations between parameters. The expression of $\boldsymbol{M}_F$ for the full $\boldsymbol{\Omega}$ matrix has been developed by [20] but has not been yet implemented in PFIM.

In this work, the development of $\boldsymbol{M}_F$ was based on an approximation by linearisation of the model using the Taylor expansion. In our linearised calculation, we chose to assume independence between the variance of the observations and the fixed effects as in a linear mixed model, which leads to a block diagonal expression of $\boldsymbol{M}_F$. Retout and Mentré [29] and other authors [22, 24, 25, 41] have also proposed an expression of $\boldsymbol{M}_F$ taking into consideration the dependence of the observations on the parameters of the model, leading to a complete $\boldsymbol{M}_F$ with an additional off-diagonal bloc $\boldsymbol{C}$. There is no clear consensus on what is the best approximation but we think that the block diagonal expression is better. Indeed, presently, Mielke and Schwabe [31] in addition showed that an approach with block diagonal $\boldsymbol{M}_F$ was more reliable than the one with the full $\boldsymbol{M}_F$. In the software MONOLIX, the observed $\boldsymbol{M}_F$ calculated by linearisation of the model around the individual parameter estimate has also a block diagonal expression. Also, the method presented here was shown to be relevant by simulation for the present example and in others [26, 42]. This can be perhaps explained by the fact that $\boldsymbol{M}_F$ calculation is based on derivations of the model. The linearisation could perhaps introduce potential problems when the model is very nonlinear or very complex and written in differential equations. Consequently, one of the perspectives of this work would be to propose a computation of $\boldsymbol{M}_F$ without model linearisation using a stochastic approach.

Studies analysed through NLMEM can be perfomed with optimal sparse simpling times with almost no loss of power. This requires the knowledge of the model and its parameters, a limitation also present in the approach by simulation. Sensitivity analyses with respect to the model and the parameter values would be necessary and interesting to quantify how the results possibly vary. However, at the designing step in a PK equivalence or interaction study of a drug, one can usually use the results of a previous study on the same drug so that the PK of the drug is already well known, and the model has already been evaluated. Also, this method allows of reducing the number of samples per subject, which can be ethically and practically very important for performing studies

[2] Supporting information may be found in the online version of this article.

in patients. In conclusion, we have shown the relevance of the extension of $\boldsymbol{M}_F$ in NLMEM for crossover trials. The approach by NLMEM is an appropriate alternative to NCA in the case of trials in patients. The implementation of these new developments in PFIM provides a useful tool to design bioequivalence/interaction studies as well as other kinds of longitudinal studies, avoiding extensive simulations.

6 Acknowledgments

The authors would like to thank the biotechnology company Da Volterra, 172 rue de Charonne, 75011 Paris, France (http://www.davolterra.com) for providing the data of the pharmacokinetic study DAV1 of amoxicillin in piglets. *Conflict of Interest*: None declared.

References

1. FDA. Guidance for Industry - Statistical Aproaches to establishing bioequivalence. *Technical report*, FDA 2001.
2. EMEA. Note for guidance on the investigation of bioavailability and bioequivalence. *Technical report*, EMEA 2001.
3. Hauschke D, Steinijans V, Pigeot I. *Bioequivalence studies in drug development*. John Wiley & Sons: Chichester, 2007.
4. Gabrielson J, Weiner D. *Pharmacokinetic and pharmacodynamic data analysis: concepts and applications*. 4th edn., Swedish Pharmaceutical Press: Stockholm, 2006.
5. Yeh C. Estimation and significant tests of area under the curve derived from incomplete blood sampling. *American Statistical Association Proceedings of the Biopharmaceutical Section*, 1990; 74–81.
6. Jaki T, Wolfsegger MJ. A theoretical framework for estimation of aucs in complete and incomplete sampling designs. *Statistics in Biopharmaceutical Research* 2009; **1**:176–184, DOI: 10.1198/sbr.2009.0025.
7. Jaki T, Wolfsegger MJ, J-P L. Establishing bioequivalence in complete and incomplete data designs using aucs. *Journal of Biopharmaceutical Statistic* 2010; **20**:803–820.
8. Jusko WJ, Koup JR, Alvan G. Nonlinear assessment of phenytoin bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics* 1976; **4**:327–336, DOI:10.1007/BF01063122.
9. Hayashi N, Aso H, Higashida M, Kinoshita H, Ohdo S, Yukawa E, Hiquchi S. Estimation of rhG-CSF absorption kinetics after subcutaneous administration using a modified Wagner-Nelson method with a nonlinear elimination model. *European Journal of Pharmaceutical Sciences* 2001; **13**:151–158, DOI:10.1016/S0928-0987(00)00219-0.
10. Sheiner LB, Rosenberg B, Melmon KL. Modelling of individual pharmacokinetics for computer-aided drug dosage. *Computers and Biomedical Research* 1972; **5**:411–459, DOI: 10.1016/0010-4809(72)90051-1.
11. Sheiner LB, Steimer JL. Pharmacokinetic/pharmacodynamic modelling in drug development. *Annual Review of Pharmacology and Toxicology* 2000; **40**:67–95, DOI: 10.1146/annurev.pharmtox.40.1.67.
12. Panhard X, Taburet AM, Piketti C, Mentré F. Impact of modeling intra-subject variability on tests based on non-linear mixed-effects models in cross-over pharmacokinetic trials with ap-

plication to the interaction of tenofovir on atazanavir in HIV patients. *Statistics in Medicine* 2007; **26**:1268–1284, DOI:10.1002/sim.2622.
13. Dubois A, Gsteiger S, Pigeolet E, Mentré F. Bioequivalence tests based on individual estimates using non-compartmental or model-based analyses: evaluation of estimates of sample means and type I error for different designs. *Pharmaceutical Research* 2010; **27**:92–110, DOI: 10.1007/s11095-009-9980-5.
14. Dubois A, Lavielle M, Gsteiger S, Pigeolet E, Mentré F. Model-based analyses of bioequivalence crossover trials using the SAEM algorithm. *Statistics in Medicine* 2011; DOI: 10.1002/sim.4286.
15. Al-Banna MK, Kelman AW, Whiting B. Experimental design and efficient parameter estimation in population pharmacokinetics. *Journal of Pharmacokinetics and Biopharmaceutics* 1990; **18**:347–360, DOI:10.1007/BF01062273.
16. Hashimoto Y, Sheiner LB. Designs for population pharmacodynamics: value of pharmacokinetic data and population analysis. *Journal of Pharmacokinetics and Biopharmaceutics* 1991; **19**:333–353, DOI:10.1007/BF03036255.
17. Jonsson EN, Wade JR, Karlsson MO. Comparison of some practical sampling strategies for population pharmacokinetic studies. *Journal of Pharmacokinetics and Biopharmaceutics* 1996; **24**:245–263, DOI:10.1007/BF02353491.
18. Atkinson AC, Donev AN. *Optimum Experimental Designs*. Oxford Statistical Science, Clarendon Press: Oxford, 1992.
19. Walter E, Pronzato L. *Identification of Parametric Models from Experimental Data*. Springer: New York, 2007.
20. Mentré F, Mallet A, Baccar D. Optimal design in random effect regression models. *Biometrika* 1997; **84**:429–442, DOI:10.1093/biomet/84.2.429.
21. Bazzoli C, Retout S, Mentré F. Design evaluation and optimisation in multiple response nonlinear mixed effect models: PFIM 3.0. *Computer Methods and Programs in Biomedicine* 2010; **98**:55–65, DOI:10.1016/j.cmpb.2009.09.012.
22. Hooker A, Vicini P. Simultaneous population optimal design for pharmacokinetic-pharmacodynamic experiments. *American Association of Pharmaceutical Scientists Journal* 2005; **7**:759–785, DOI:10.1208/aapsj070476.
23. Gueorguieva I, Aarons L, Ogungbenro K, Jorga KM, Rodgers T, Rowland M. Optimal design for multivariate response pharmacokinetic models. *Journal of Pharmacokinetics and Pharmacodynamics* 2006; **33**:97–124, DOI:10.1007/s10928-006-9009-1.
24. Ogungbenro K, Gueorguieva I, Majid O, Graham G, Aarons L. Optimal design for multiresponse pharmacokinetic-pharmacodynamic models-dealing with unbalanced designs. *Journal of Pharmacokinetics and Pharmacodynamics* 2007; **34**:313–331, DOI:0.1007/s10928-006-9048-7.
25. Ogungbenro K, Graham G, Gueorguieva I, Aarons L. Incorporating correlation in interindividual variability for the optimal design of multiresponse pharmacokinetic experiments. *Journal of Biopharmaceutical Statistics* 2008; **18**:342–358, DOI:10.1080/10543400701697208.
26. Bazzoli C, Retout S, Mentré F. Fisher information matrix for nonlinear mixed effects multiple response models: Evaluation of the appropriateness of the first order linearization using a pharmacokinetic/pharmacodynamic model. *Statistics in Medicine* 2009; **28**:1940–1956, DOI: 10.1002/sim.3573.
27. Mentré F, Duffull S, Gueorguieva I, Hooker A, Leonov S, Ogungbenro K, Retout S. Software for optimal design in population pharmacokinetics and pharmacodynamics: a comparison. 16th Population Approach Group in Europe, Copenhagen, Denmark, 2007; Abstract 1179.

`http://www.page-meeting.org/?abstract=1179`.
28. Karlsson MO, Sheiner LB. The importance of modelling the inter-occasion variability in population pharmacokinetic studies. *Journal of Pharmacokinetics and Biopharmaceutics* 1993; **21**:735–750, DOI:10.1007/BF01113502.
29. Retout S, Mentré F. Further developments of the Fisher information matrix in nonlinear mixed effects models with evaluation in population pharmacokinetics. *Journal of Biopharmaceutical Statistics* 2003; **13**:209–227, DOI:10.1081/BIP-120019267.
30. Retout S, Comets E, Samson A, Mentré F. Design in nonlinear mixed effects models: Optimisation using the Fedorov-Wynn algorithm and power of the wald test for binary covariates. *Statistics in Medicine* 2007; **26**:5162–5179, DOI:10.1002/sim.2910.
31. Mielke T, Schwabe R. Some considerations on the Fisher information in nonlinear mixed effects models. *Proceedings of the 9th International Workshop in Model-Oriented Design and Analysis*, Bertinoro, Italy, 2010.
32. Panhard X, Mentré F. Evaluation by simulation of tests based on non-linear mixed effects models in pharmacokinetic interaction and bioequivalence cross-over trials. *Statistics in Medicine* 2005; **24**:1509–1524, DOI:10.1002/sim.2047.
33. Schuirmann DJ. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics* 1987; **15**:657–680, DOI:10.1007/BF01068419.
34. Fedorov V. *Theory of Optimal Experiments*. Academic Press: New York, 1972.
35. Wynn HP. Results in the theory and construction of D-optimum designs. *Journal of the Royal Statistical Society B* 1972; **34**:133–147.
36. Delyon B, Lavielle M, Moulines E. Convergence of a stochastic approximation version of EM algorithm. *The Annals of Statistics* 1999; **27**:94–128, DOI:10.1214/aos/1018031103.
37. Kuhn E, Lavielle M. Maximum likelihood estimation in nonlinear mixed effects model. *Computational Statistics and Data Analysis* 2004; **49**:1020–1038, DOI: 10.1016/j.csda.2004.07.002.
38. Samson A, Lavielle M, Mentré F. The SAEM algorithm for group comparison tests in longitudinal data analysis based on non-linear mixed effects model. *Statistics in Medecine* 2007; **26**:4860–4875, DOI:10.1002/sim.2950.
39. Panhard X, Samson A. Extension of the SAEM algorithm for nonlinear mixed effects models with two levels of random effects. *Biostatistics* 2009; **10**:121–135, DOI: 10.1093/biostatistics/kxn020.
40. Samson A, Lavielle M, Mentré F. Extension of the SAEM algorithm to left-censored data in nonlinear mixed-effects model: application to HIV dynamics model. *Computational Statistics and Data Analysis* 2006; **51**:1562–1574, DOI:10.1016/j.csda.2006.05.007.
41. Gagnon R, Leonov S. Optimal population designs for PK models with serial sampling. *Journal of Biopharmaceutical Statistics* 2005; **15**:143–163, DOI:10.1081/BIP-200040853.
42. Guedj J, Bazzoli C, Neumann AU, Mentré F. Design evaluation and optimisation for models of hepatitis C viral dynamics. *Statistics in Medicine* 2011; **30**:1045–1056, DOI: 10.1002/sim.4191.

Table I – Predicted and empirical standard error SE(β_{Cl}) for each design and value of β_{Cl}.

Design	$\exp(\beta_{Cl})$	Predicted SE(β_{Cl}) [$\times 10^2$]	Empirical SE(β_{Cl}) [$\times 10^2$]
rich	0.8	3.401	3.290
(0.5, 1, 1.5, 2, 4, 6, 8 h)	1	3.404	3.297
	1.1	3.405	3.300
	1.2	3.406	3.305
	1.25	3.407	3.308
	1.5	3.410	3.320
sparse	0.8	3.443	3.357
(0.5, 2, 6, 8 h)	1	3.454	3.372
	1.1	3.459	3.377
	1.2	3.462	3.390
	1.25	3.463	3.397
	1.5	3.467	3.423

Table II – Predicted powers (%) and observed powers (%) for each design and value of β_{Cl}.

Design	$\exp(\beta_{Cl})$	Comparison test		Equivalence test	
		Predicted power	Observed power	Predicted power	Observed power
rich	0.8	100	100	5.00	5.60
(0.5, 1, 1.5, 2, 4, 6, 8 h)	1	5.00	5.30	100.00	100.00
	1.1	79.92	82.80	98.25	98.90
	1.2	99.97	99.90	32.76	32.00
	1.25	100	100	5.00	4.30
	1.5	100	100	NC	NC
sparse	0.8	100	100	5.00	5.60
(0.5, 2, 6, 8 h)	1	5.00	5.10	100.00	100.00
	1.1	78.69	79.80	97.99	98.20
	1.2	99.95	99.80	32.08	34.10
	1.25	100	100	5.00	4.70
	1.5	100	100	NC	NC

(NC: not computed because β_{Cl} is not under H_1.)

Table III – Population pharmacokinetic parameters of amoxicillin and and standard errors (SE) as well as relative standard errors (RSE) of parameter estimates by MONOLIX 2.4 for DAV1 study.

Parameters	Estimates	SE	RSE (%)
$tlag$ (h)	0.37	0.03	7
k_a (h^{-1})	0.81	0.13	16
V ($L.kg^{-1}$)	2.86	0.85	30
Cl ($L.h^{-1}.kg^{-1}$)	2.99	0.35	12
β_{Cl}	0.06	0.16	265
ω_{tlag}	0	-	-
ω_{ka}	0.10	0.14	144
ω_V	0.79	0.51	64
ω_{Cl}	0	-	-
γ_{tlag}	0.083	0.04	42
γ_{ka}	0.10	0.12	121
γ_V	0.73	0.34	46
γ_{Cl}	0.19	0.06	28
σ_{inter} ($ng.L^{-1}$)	0.31	19.00	6

Table IV – Relative standard error (RSE; %) and standard error (SE) predicted by new extension ofMF for all the population parameters and criterion value according to each two-period crossover design of DAV2 study.

	Rich design	Sparse design
RSE(ka)	23.7	26.4
RSE(V)	36.0	37.9
RSE(Cl)	11.2	11.2
SE(β_{Cl})	0.157	0.158
RSE(ω_{ka})	230.2	282.5
RSE(ω_V)	65.6	68.5
RSE(ω_{Cl})	-	-
RSE(γ_{ka})	177.7	205.7
RSE(γ_V)	40.9	43.7
RSE(γ_{Cl})	18.4	18.6
RSE(σ_{inter})	5.9	10.2
Criterion value	36.55	30.01

Table V – Prediction of power for the Wald test of equivalence and of NSN to achieve a power of 90% computed by PFIM 3.2 for each design of DAV2 study.

Design	β_{Cl}	Two periods AP/AX		Four periods AP/AX/AP/AX	
		Predicted power (%) with N= 16 piglets	NSN for a power of 90%	Predicted power (%) with N= 16 piglets	NSN for a power of 90%
rich	0	41.0	68	64.3	34
(0.5, 1, 1.5, 2, 4, 6, 8 h)	0.06	27.0	128	41.3	67
	0.10	19.3	209	27.7	124
sparse	0	40.5	70	63.7	35
(0.5, 2, 4, 6 h)	0.06	26.7	131	40.8	69
	0.10	19.1	213	27.4	126

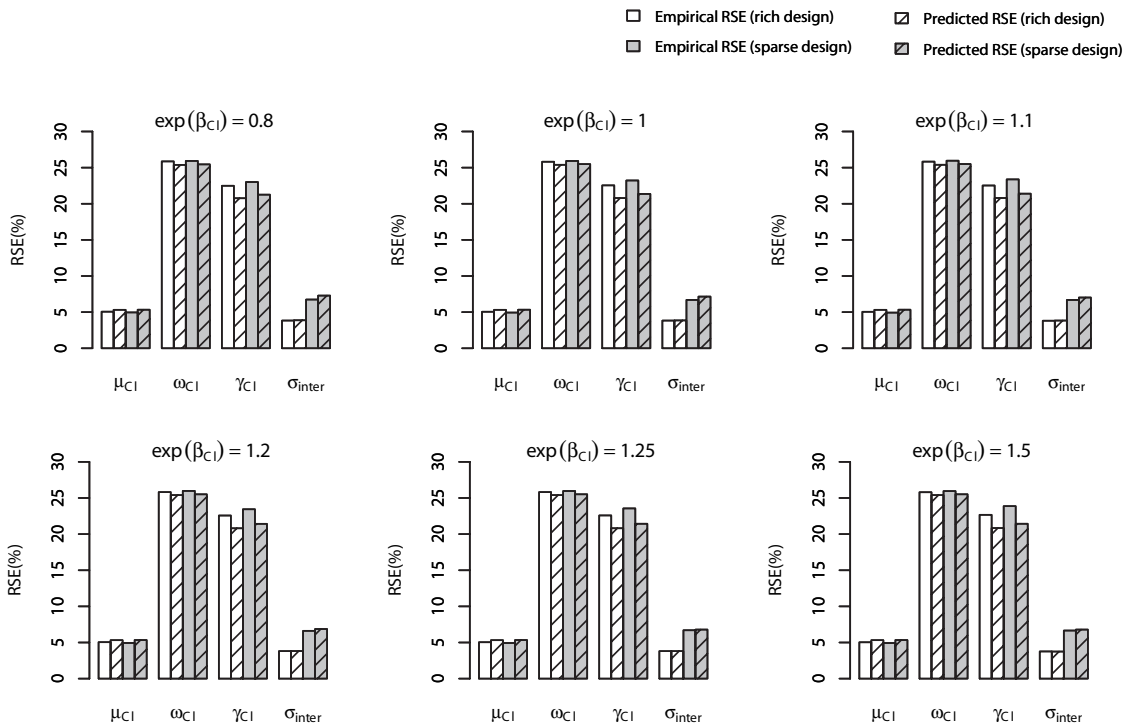

Figure 1 – Barplots of predicted (hatched bar) and empirical (plain bar) RSE (%) for the fixed effect, the between subject variability, the within subject variability on Cl and for the random error parameter σ_{inter} for different values of β for the rich (white bar) or the sparse (grey bar) design

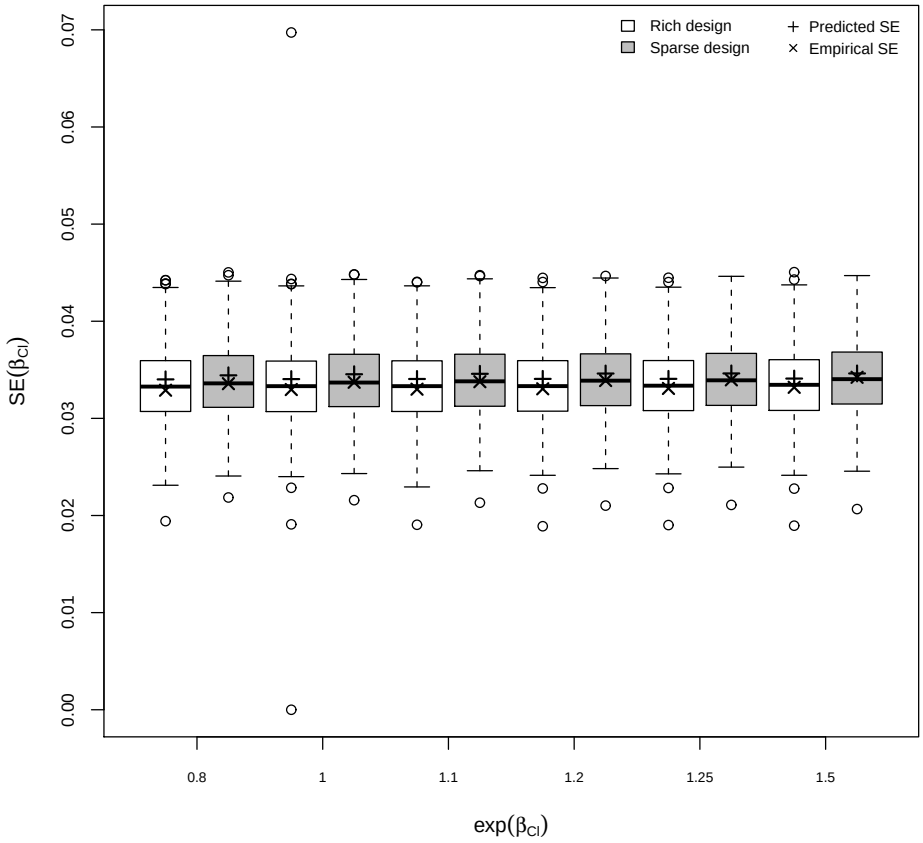

Figure 2 – Boxplots of the observed SE for treatment effect parameter in each simulated scenario. $(+)$ represents the value of the SE predicted by the extension of $\boldsymbol{M}_F$ and $(\times)$ represents the empirical SE for the rich (white bar) or the sparse (grey bar) design.

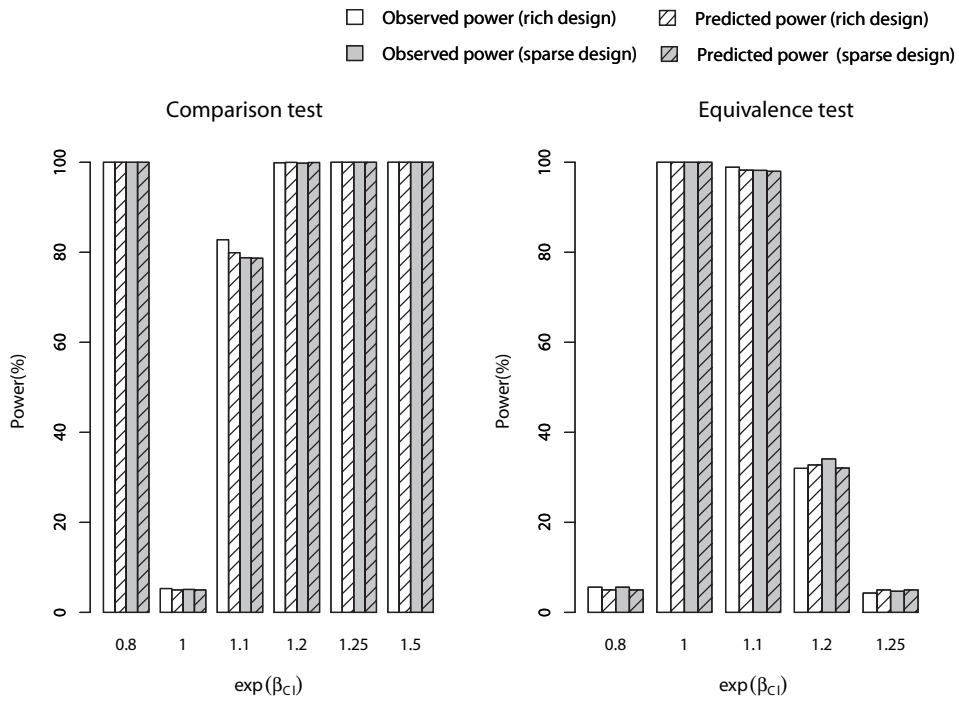

Figure 3 – Barplots of predicted (hatched bar) and observed (plain bar) power (%) for different values of β_{Cl} for the rich (white bar) or the sparse (grey bar) design for the comparison test (left) and for the equivalence test (right)

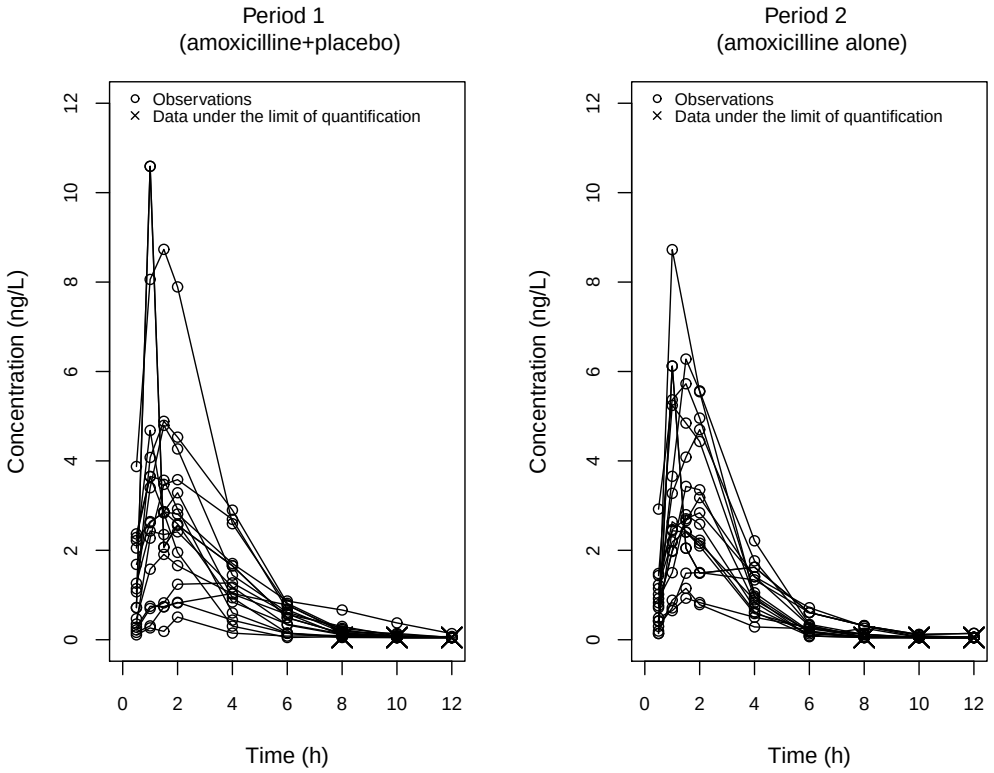

Figure 4 – Spaghetti plots of plasma amoxicillin concentrations of DAV1 study.