

Using Fisher Information Matrix to predict uncertainty in covariate effects and power to detect their relevance in Non-Linear Mixed Effect Models in pharmacometrics

Lucie Fayette^{1,2}, Karl Brendel², and France Mentré¹

¹*Université Paris Cité, Inserm, IAME, F-75018 Paris, France*

²*Pharmacometrics, Ipsen Innovation, Les Ulis, France*

lucie.fayette@inserm.fr

Abstract

This work focuses on design of experiments for Pharmacokinetic (PK) and Pharmacodynamic (PD) studies. Non-Linear Mixed effects Models (NLMEM) modeling allows the identification and quantification of covariates that explain inter-individual variability (IIV). The Fisher Information Matrix (FIM), computed by linearization, has already been used to predict uncertainty on covariate parameters and the power of a test to detect statistical significance. A covariate effect on a parameter is deemed statistically significant if it is different from 0 according to a Wald comparison test and clinically relevant if the ratio of change it causes in the parameter is relevant according to two one-sided tests (TOST) as in bioequivalence studies. FIM calculation was extended by computing its expectation on the joint distribution of the covariates, discrete and continuous. Three methods were proposed: using a provided sample of covariate vectors, simulating covariate vectors, based on provided independent distributions or on estimated copulas. Thereafter, CI of ratios, power of tests and number of subjects needed to achieve desired confidence were derived. Methods were implemented in a working version of the R package *PFIM*. A simulation study was conducted under various scenarios, including different sample sizes, sampling points, and IIV. Overall, uncertainty on covariate effects and power of tests were accurately predicted. The method was applied to a population PK model of the drug cabozantinib including 27 covariate relationships. Despite numerous relationships, limited representation of certain covariates, PFIM correctly predicted uncertainty, and is therefore suitable for rapidly computing number of subjects needed to achieve given powers.

Keywords: Design; Fisher information matrix; Nonlinear mixed effects models; Covariate effect; Power of relevance; Pharmacometrics

1 Introduction

One of the main aims of pharmacokinetic (PK) and pharmacodynamic (PD) analysis is to identify and quantify the covariates that attempt to explain inter-individual variability (IIV). In this context Non-Linear Mixed effects Models (NLMEM) are an appropriate tool as population modeling allows to quantitatively describe relationships between covariates and parameters, and consequently to explain some of the IIV. [1] The list of covariates of interest can be represented on a forest plot where the 90% confidence interval (CI) of the ratio of change in the value of the parameter is expressed for given values, or categories, of the covariate, and relative to a reference value [2]. These plots are in line with the FDA Guidance for Industry Population Pharmacokinetics (2022) [3] and are essential to support clinicians' decisions on drug dosing [4]. Therefore, regarding the pharmaceutical development of a new drug, it is essential to be able to detect covariates that may explain some of the PKPD variability, and to distinguish between a statistically significant and a clinically relevant effect. A covariate effect on a parameter is considered statistically significant if it is different from 0 according to a Wald test at level 95%, clinically relevant if the ratio of change it causes in the parameter is relevant according to a Two One-Sided Tests Procedure (TOST) at level 95% . The latter is equivalent to showing that the 90% CI of the ratio is outside a predetermined interval (often [0.80; 1.25]) [5]. The ratio is said irrelevant if this CI is included in the range. In the case of partial overlap, there is insufficient data to determine clinical relevance.

For a given model and parameters, the Standard Error (SE) of the covariate parameters and the width of CI of the ratio are influenced by the design of the study and by the covariate characteristics of the included patients in the analysis. The design refers to the number of subjects and their elementary designs, which means their dosage regimens and the allocation of their measurement times. Being able to predict the SE of the covariate effect estimates, in order to predict the ability to conclude on the significance and relevance of an effect is a key issue for the design of a clinical trial. For this reason, at design step, it is recommended to perform clinical trial simulation (CTS) to determine an adequate sample size or an **NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.** can be computationally expensive, and one aim of this work was to develop a method avoiding these computationally intense CTS

for design evaluation through the use of the Fisher Information Matrix (FIM). This type of methodology is described in the general theory of optimal experimental design, which is applied to traditional non-linear models [7]. Indeed, according to the inequality of Rao-Cramer the variance of any unbiased estimator is bounded by the inverse of the FIM. Therefore, expected SE resulting from a design can be computed from it. In NLMEM framework there is no analytical solution for FIM computation, and the usual approach consist in a first order linearization of the structural model, around a mean of 0 for the random effects (FO method) or around individual realizations of the random effects (FOCE method) and then the FIM is computed as the FIM of the obtained Gaussian approximation [8]. This methodology has been shown to be appropriate for evaluating and optimizing design of experiments [9]. In this computation, the covariates are generally considered as being known for all the subjects, although accounting for their distributions, especially for continuous covariates is a priority for pharmaceutical companies in terms of clinical trial design and tools [10]. From the SE predicted by the FIM, the power of a parametric test to detect whether a covariate effect is statistically significant can be derived. Previous works have explored this computation for example including one discrete covariate and have shown that the trial design can be optimized to achieve a desired confidence level [11] [12]. These solutions have been implemented in *PFIM*, which was the first software tool proposed in 2001 for evaluating a design without using simulations. Especially, *PFIM* version 4.0 [13] includes evaluation methods for discrete covariates only while the latest version, the R package *PFIM 6.03* [14] (CRAN release March 2024) does not include covariates. Other software tools relying on this FIM based approach have been developed and handle both discrete and continuous covariates, such as PopED [15, 16] and NONMEM\$DESIGN [17]. In PopED, covariates values are considered as known in FIM computation. Otherwise, to improve the computation, it is suggested in the user-guide to assume a distribution for the covariates and sample from it for each subject, then repeat the process and compute the average standard errors across the repetitions. NONMEM\$DESIGN handles covariate similarly. The drawback of this assumption is the need to know the covariate values for each subject.

Furthermore, while the FIM gives asymptotic SE for covariate parameters, the quantity of interest is the ratio of change in PK parameters when covariate values change relative to a reference value. For classical covariate relationship models, an analytical solution allows to compute the CI on these ratios from the value of the covariate parameters and its SE derived from the FIM, otherwise approximation such as the Delta method can be used. Thereafter, it is possible to predict the power of relevance test and to predict the CI on a forest plot, and thus evaluate whether the design leads to sufficient information to conclude to the relevance of covariates.

The goals of this work were thus to extend *PFIM6* to include both discrete and continuous covariates, to explore different approaches to handle continuous covariates and to predict SE of ratio, power of significance and relevance tests. To that purpose the computation of the linearized FIM was extended to the case with continuous covariates, comparing three methods to account for covariate distribution based on Monte-Carlo computation as already proposed [18]. The first one uses covariate vectors data, the second one independent distributions for the covariates, and the third one copulas, as the benefit of using copulas in pharmacometrics has already been demonstrated [[zwep2022virtual](#)]. The power of a clinical relevance test on covariate effect was computed analytically. These methods were implemented in a working version of *PFIM 6* and are available in the Zenodo repository <https://doi.org/10.5281/zenodo.13692989>. Based on a simple PK model, simulations were performed to assess the accuracy of the proposed approaches. CTS and PFIM were compared in 4 scenarios, varying the sample size, the observation schedule and the IIV. Thereafter, this methodology was applied to a real example inspired from a population PK analysis of the drug cabozantinib [19] performed from 10 clinical trials, in healthy volunteers and patients with various cancer types The original Population PK model included 27 covariate relationships. The theoretical design was evaluated using the proposed methodology and the power of tests and the number of subjects that would have been required to achieve 80% power were calculated.

In Section 2, the notations and methods are detailed. In Section 3 we go through the simulation study, while section 4 describes the application to the real example. Finally global results are discussed in section 5

2 Methods

In this section we detail notations and methods on study design as well as NLMEM and on the FIM computation accounting for covariates. After that, derivation of SE and CI on ratios from the FIM are shown. Then computation for power of significance and relevance tests, and for number of subjects needed to achieve given levels of power are detailed.

2.1 Design and NLMEM

For one subject indexed by i , the elementary design ξ_i is defined by its number of observations n_i , some designs variables $\xi_{i,1}, \dots, \xi_{i,n_i}$ as observations time, doses, dosage regimens. The population design $\Xi = \{N, (\xi_1, \dots, \xi_N)\}$ is composed of the number of subjects N and their respective elementary designs.

We denote y_i the n_i -vector of observations for the i^{th} subject. Assuming Gaussian distribution for the observations, these can be modeled as given in equation (1), where f denotes the structural non-linear model. Thus the structural response depends on ξ_i , the elementary design for subject i and on $\theta_i = u(\mu, \beta, z_i, \eta_i)$, the p -vector of individual parameters. The latter is a function of $\nu = (\mu, \beta)$ the vector of fixed effects, with μ the typical values and β the covariate effects, and of the vector of possibly transformed individual covariates z_i and individual random effects η_i , where $\eta_i \sim \mathcal{N}(0, \Omega)$.

$$y_{ij} = f(\xi_i, \theta_i) + g(\xi_i, \theta_i, \sigma)\epsilon_{ij} \quad (1)$$

It is commonly assumed that the individual parameter can be transformed by a function h to a normal variable as given in equation (2).

$$\phi_i = h(\theta_i) = h(\mu, \beta, z_i) + \eta_i \sim \mathcal{N}(h(\mu, \beta, z_i), \Omega) \quad (2)$$

Moreover, it is usually assumed that each individual parameter ϕ_{il} , $l = 1, \dots, p$, is a linear function of the possibly transformed covariates as given in equation (3), where β_l is a vector of size C , the number of covariates, and z_{pop} denotes the typical value of the (transformed) covariates in the population.

$$\phi_{il} = h_l(\mu_l) + \beta_l(z_i - z_{pop}) + \eta_{il} \quad (3)$$

In PKPD, most of the parameters are modeled as log-normally distributed with additive covariate relationships on the log scale, i.e. $\theta_{il} = \mu_l e^{\beta_l(z_i - z_{pop}) + \eta_{il}}$ and $\phi_{il} = \log \mu_l + \beta_l(z_i - z_{pop}) + \eta_{il}$, where z_i is the vector of (transformed) covariates.

The residual error $g(\xi_i, \theta_i, \sigma)\epsilon_i$ is normally distributed with $\epsilon_i \sim \mathcal{N}(0, I_{n_i})$. The function g models the residual error, and also depend on ξ_i and θ_i , and on some parameters denoted σ .

The P -vector of population parameters is denoted $\Psi = \{\nu, \lambda\}$, where λ contains the variance parameters (elements of Ω and of the residual error model σ).

2.2 Population Fisher Information Matrix computation with covariates

2.2.1 Fisher Information Matrix

For a given elementary design ξ_i and a given vector of covariates z_i , the expected elementary Fisher information is defined as the covariance matrix of the Fisher score (equation (4)) where $l(y_i; \Psi, z_i, \xi_i)$ is the likelihood of the vector of observations y_i for the population parameters Ψ , given the individual vector of covariates z_i and the elementary design ξ_i (equation (5)).

$$M_F(\xi_i, \Psi, z_i) = \mathbb{E}_\Psi \left(\frac{\partial \log l(y_i; \Psi, z_i, \xi_i)}{\partial \Psi} \frac{\partial \log l(y_i; \Psi, z_i, \xi_i)}{\partial \Psi^T} \right) \quad (4)$$

$$l(y_i; \Psi, z_i, \xi_i) = \int p(y_i | \eta_i; \Psi, z_i, \xi_i) p(\eta_i; \Psi, z_i) d\eta_i \quad (5)$$

If the log-likelihood is twice differentiable, the FIM can be computed as minus the expectation of the Hessian of the log-likelihood (see equation (6)).

$$M_F(\xi_i, \Psi, z_i) = -\mathbb{E}_\Psi \left(\frac{\partial^2 \log l(y_i; \Psi, z_i, \xi_i)}{\partial \Psi^2} \right) \quad (6)$$

To calculate the standard error of a vector of population parameters, the standard method in NLMEM is to use the square root of the diagonal elements of the inverse of the FIM. According to the Rao-Cramer inequality, these values constitute the lower bound of the standard errors of any unbiased estimator of the parameters. We therefore assume in what follows that we are in asymptotic conditions and with an unbiased estimator such that: $\hat{\Psi} \sim \mathcal{N}(\Psi, M_F(\xi_i, \Psi, z_i)^{-1})$.

2.2.2 Calculation by linearization

Due to the non-linearity of the structural model f with respect to θ , there is no analytical expression for $l(y_i; \Psi, z_i, \xi_i)$. The expression for $M_F(\xi_i, \Psi, z_i)$ are usually developed using a first-order Taylor expansion of the structural model around $\phi_{i,0} = h(\mu_0, \beta_0, z_i)$, a guess value of the fixed effects associated with covariates z_i , and a zero-order expansion of g [8] as given in equation (7), where $\mathbf{J}_\theta f$ and $\mathbf{J}_\phi h^{-1}$ respectively denote the Jacobian matrix of f and h^{-1} .

$$y_i \approx f(\xi_i, h^{-1}(\phi_{i,0})) + \mathbf{J}_\theta f(\xi_i, h^{-1}(\phi_{i,0})) \mathbf{J}_\phi h^{-1}(\phi_{i,0})(h(\mu, \beta, z_i) - \phi_{i,0}) + \mathbf{J}_\theta f(\xi_i, h^{-1}(\phi_{i,0})) \mathbf{J}_\phi h^{-1}(\phi_{i,0}) \eta_i + \text{diag}(g(\xi_i, h^{-1}(\phi_{i,0}), \sigma)) \epsilon_i \quad (7)$$

Observations are thus approximated to normal variables, with mean $E_i(\xi_i, \phi_{i,0}, \nu, z_i)$ and variance-covariance matrix $V_i(\xi_i, \phi_{i,0}, \Omega, \sigma)$ given in equation (8).

$$\begin{cases} E_i(\xi_i, \phi_{i,0}, \nu, z_i) = f(\xi_i, h^{-1}(\phi_{i,0})) + \mathbf{J}_\theta f(\xi_i, h^{-1}(\phi_{i,0})) \mathbf{J}_\phi h^{-1}(\phi_{i,0})(h(\mu, \beta, z_i) - \phi_{i,0}) \\ V_i(\xi_i, \phi_{i,0}, \Omega, \sigma) = (\mathbf{J}_\theta f(\xi_i, h^{-1}(\phi_{i,0})) \mathbf{J}_\phi h^{-1}(\phi_{i,0}))^T \Omega (\mathbf{J}_\theta f(\xi_i, h^{-1}(\phi_{i,0})) \mathbf{J}_\phi h^{-1}(\phi_{i,0})) + \text{diag}(g(\xi_i, h^{-1}(\phi_{i,0}), \sigma))^2 \end{cases} \quad (8)$$

From this, the FIM is calculated as the FIM of the Gaussian vector obtained by approximation (see for example [18]).

2.2.3 Calculation with covariates

In previous expressions, $M_F(\xi_i, \Psi, z_i)$ is a function of z_i . There are two main cases. First, if the covariate vector is known for each subject, $M_F(\xi_i, \Psi, z_i)$ can be computed directly. Otherwise, z_i must be considered as a random variable and $M_F(\xi_i, \Psi)$ is compute as an integral with respect to p_z , the distribution of the covariate vector as given in equation (9).

$$M_F(\xi_i, \Psi) = \int M_F(\xi_i, \Psi, z_i) p_z(z_i) dz_i \quad (9)$$

The population FIM of the design $\Xi = \{\xi_1, \dots, \xi_N\}$ writes as the sum of the FIM of the elementary designs (see equation (10)).

$$M_F(\Xi, \Psi) = \sum_{i=1}^N \int M_F(\xi_i, \Psi, z_i) p_z(z_i) dz_i \quad (10)$$

To compute $M_F(\Xi, \Psi)$, a solution is to approximate it through a Monte Carlo integration of the covariates as given in equation (11).

$$M_F(\Xi, \Psi) \approx \sum_{i=1}^N \frac{1}{S} \sum_{s=1}^S M_F(\xi_i, \Psi, z_{i,s}) \quad \text{where } z_{i,s} \text{ is drawn from } p_z \quad (11)$$

This method allows to account for the dependence structure between the covariates. However, it raises the question of how to draw the $z_{i,s}$ used to calculate the integral. In this work, three methods were explored. The first one is non-parametric and consists in using an available dataset of covariate vectors, directly as a sample from the covariate distribution or as a discrete distribution and sample from it. This approach allows to account for the correlations between the covariates. Otherwise, vectors of covariates can be simulated using either a given probability distribution or a copula.

Copulas are cumulative distribution functions that capture the dependence structure between random variables and are therefore of interest to simulate correlated covariates. The benefits of using copulas for patients simulation in the field of pharmacometrics have already been demonstrated [20]. However, while the theory of copulas is well established for continuous variables, the same cannot be said for discrete variables. For a discrete vector, a copula cannot be uniquely identified and measures of dependence between variables become dependent on marginal laws. Methods for constructing copulas with discrete variables have been proposed (see for instance Geenens [21] for a review), but are beyond the scope of this work. In pharmacometrics, discrete covariates being in the vast majority of cases categorical covariates, we will be able to circumvent this problem by working with a copula by categories/combination of categories in the presence of several categorical covariates.

2.3 SE and ratio CI derivation from the FIM

Regarding covariate effects, the quantity of interest is often the ratio of change in primary parameters when covariate values change relative to a reference value z_{ref} .

For categorical covariates, the expected ratio is calculated for the non-reference categories in relation to the reference category. For instance, for a binary covariate z_{ic} with reference $z_{ic} = z_{pop,c} = 0$, and non-reference $z_{ic} = 1$, the general formula for the ratio on the l^{th} parameter is given in equation (12).

$$r_{l,c} = \frac{h_l^{-1}(h_l(\mu_l) + \beta_{l,c})}{\mu_l} \quad (12)$$

If $h_l(\mu_l) = \mu_l$, the ratio writes $r_{l,c} = \frac{\mu_l + \beta_{l,c}}{\mu_l}$ and if $h_l(\mu_l) = \log \mu_l$, it writes $r_{l,c} = e^{\beta_{l,c}}$.

For continuous covariates, the expected ratio is usually computed in the forest plot for the 10th and 90th percentiles of the covariate distribution, respectively denoted $P90$ and $P10$, in relation to its median denoted $P50$, as given in equation (13);

$$r_{l,c(PX)} = \frac{h_l^{-1}(h_l(\mu_l) + \beta_{l,c}(PX - z_{pop,c}))}{h_l^{-1}(h_l(\mu_l) + \beta_{l,c}(P50 - z_{pop,c}))} \quad \text{where } PX = P10 \text{ or } PX = P90 \quad (13)$$

For instance, if $h_l(\mu_l) = \mu_l$, the ratio writes $r_{l,c(PX)} = \frac{\mu_l + \beta_{l,c}(PX - z_{pop,c})}{\mu_l + \beta_{l,c}(P50 - z_{pop,c})}$ and if $h_l(\mu_l) = \log \mu_l$, it writes $r_{l,c(PX)} = e^{\beta_{l,c}(PX - P50)}$.

The predicted CI at level $1 - \alpha$ on a covariate parameter, assuming unbiased estimator, is $\beta \pm q_{1-\alpha/2} SE_{\beta}$, where $q_{1-\alpha/2}$ denotes the $1 - \alpha/2$ -quantile of the normal distribution. Therefore, when the ratio is a monotonic transformation of the covariate parameter only, the CI on the ratio can be easily derived. Otherwise, an approximation by the Delta method can be considered to also take into account the uncertainty on the typical value μ . For more complex cases, in particular to calculate the uncertainty of the ratio on secondary parameters such as AUC or Cmax, which depend on several model parameters, stochastic simulations can be carried out by sampling from a multivariate-normal centered on the parameter values and with variance-covariance matrix derived from the FIM. Details are for instance provided in Philipp et al. [22] and Guhl et al. [23].

In case of log-normally distributed parameter with additive covariate relationships on the log-scale, if $z_{i,c}$ is binary: $CI(r_{l,c}) = e^{\beta_{l,c} \pm q_{1-\alpha/2} SE_{\beta_{l,c}}}$, and if $z_{i,c}$ is continuous: $CI(r_{l,c(PX)}) = e^{(\beta_{l,c} \pm q_{1-\alpha/2} SE_{\beta_{l,c}})(PX - P50)}$.

2.4 Power of tests and Number of subjects needed

2.4.1 Significance test

The statistical significance of a covariate parameter is assessed by a Wald test of comparison with null hypothesis $H_0 : \beta_{l,c} = 0$ and alternative hypothesis $H_1 : \beta_{l,c} \neq 0$. The power of the significance test, P_{sign} , at level $1 - \alpha$ is given in

equation (14). The detailed power calculation is presented in the Supplementary A.1.

$$P_{sign} = 1 - \Phi\left(q_{1-\alpha/2} - \frac{\beta_{l,c}}{SE_{\beta_{l,c}}}\right) + \Phi\left(-q_{1-\alpha/2} - \frac{\beta_{l,c}}{SE_{\beta_{l,c}}}\right) \quad (14)$$

The number of subjects needed, NSN , to achieve a given power P_{sign} in significance test is given in equation (15).

$$NSN(P_{sign}) = N \times \left(\frac{SE_{\beta_{l,c}}}{SEN_{sign}(P_{sign})}\right)^2 \quad \text{where} \quad SEN_{sign}(P_{sign}) \begin{cases} \frac{\beta_{l,c}}{q_{1-\alpha/2} - \Phi^{-1}(1 - P_{sign})} & \text{if } \beta_{l,c} > 0 \\ -\frac{\beta_{l,c}}{q_{1-\alpha/2} + \Phi^{-1}(P_{sign})} & \text{if } \beta_{l,c} < 0 \end{cases} \quad (15)$$

2.4.2 Relevance test

The relevance of a covariate effect is assessed by a TOST procedure at level α , which is equivalent to the confidence interval approach with a $1 - 2\alpha$ CI [5].

The null hypothesis is H_0 : "the covariate effect is not relevant", i.e. $r_{l,c(PX)} \in [R_{inf}; R_{sup}]$ while the alternative hypothesis is H_1 : "the covariate is relevant", i.e. $r_{l,c(PX)} \notin [R_{inf}; R_{sup}]$. This two sided null hypothesis can be split into two, respectively $H_{0,inf}$ and $H_{0,sup}$:

$$\begin{cases} H_{0,inf} : r_{l,c(PX)} \geq R_{inf} & \text{and} & H_{0,sup} : r_{l,c(PX)} \leq R_{sup} \\ H_1 : r_{l,c(PX)} < R_{inf} & \text{or} & r_{l,c(PX)} > R_{sup} \end{cases}$$

Thus H_0 is not rejected unless neither $H_{0,inf}$ nor $H_{0,sup}$ is rejected.

For log-normally distributed parameter with additive covariate relationships on the log-scale, the ratio writes $r_{l,c(PX)} = e^{\beta_{l,c}(z - z_{ref})}$, and at level $1 - \alpha$, the null hypothesis is rejected if $\widehat{\beta_{l,c}} + q_{1-\alpha}SE_{\beta_{l,c}} < B_{inf}$ or if $\widehat{\beta_{l,c}} - q_{1-\alpha}SE_{\beta_{l,c}} > B_{sup}$ where depending on the sign of $z - z_{ref}$, B_{sup} and B_{inf} equal $\frac{\log R_{sup}}{z - z_{ref}}$ or $\frac{\log R_{inf}}{z - z_{ref}}$.

The power is the probability under H_1 to reject H_0 , it is given by the equation (16). The detailed calculations are presented in the Supplementary A.1.

$$P_{relev} = \Phi\left(-q_{1-\alpha} + \frac{B_{inf} - \beta_{l,c}}{SE_{\beta_{l,c}}}\right) + 1 - \Phi\left(q_{1-\alpha} + \frac{B_{sup} - \beta_{l,c}}{SE_{\beta_{l,c}}}\right) \quad (16)$$

The number of subjects needed NSN to achieve a given power P_{relev} in relevance test is given in equation (17).

$$NSN(P_{relev}) = N \times \left(\frac{SE_{\beta}}{SEN_{relev}(P_{relev})}\right)^2 \quad \text{where} \quad SEN_{relev}(P_{relev}) \begin{cases} \frac{B_{sup} - \beta_{l,c}}{\Phi^{-1}(1 - P_{relev}) - q_{1-\alpha}} & \text{if } B_{inf} < \beta_{l,c} \\ \frac{B_{inf} - \beta_{l,c}}{\Phi^{-1}(P_{relev}) + q_{1-\alpha}} & \text{if } B_{sup} > \beta_{l,c} \end{cases} \quad (17)$$

3 Evaluation by simulations

In this section, we implemented the methodology for computing the FIM using a simple PK model. Subsequently, we compared the predictions derived from this method with results from CTS. This comparison focuses on the uncertainty associated with covariate effects and their corresponding ratios, as well as the power of tests for both statistical significance and relevance.

3.1 Settings

3.1.1 Data

The evaluation framework was inspired by a case study from the University of Maryland [24]. The data are from a double blind, parallel, single dose intravenous (IV) bolus trial, in which $N = 100$ patients were randomized equally in two arms receiving respectively 100 mg and 250 mg of drug. Concentration measurements from blood samples were collected at 16 times points (0.25, 0.5, 0.75, 1, 1.5, 2, 2.5, 3, 4, 6, 8, 10, 12, 16, 20 and 24 hours post-dose). This design is hereafter referred as Design 1. Covariates were Age, *Weight*, creatinine clearance (*CLCR*) and *Sex*, with a 0.82 correlation between *Weight* and *CLCR*. Only the covariate data were used in the following.

3.1.2 Model

The PK model was a one compartment model with IV bolus and linear elimination, $f(t) = \frac{Dose}{V}e^{-\frac{Cl}{V}t}$, with two parameters: the clearance (Cl) and the volume of distribution (V). The residual error was modeled as a combined error model: $g = a + bf$. Continuous covariates *CLCR* and *Weight* were first log-transformed, and effects of log *CLCR* on Cl and

Table 1: Evaluation - Population PK parameter values

Parameter	μ_{Cl} (L/h)	μ_V (L)	ω_{Cl}	ω_V	a (mg/L)	b	$\beta_{Cl,CLCR}$	$\beta_{V,Weight}$	$\beta_{V,Sex}$
Value	0.43	7.13	0.30	0.24	2.39	0.08	0.87	1.0	-0.35
Ratio	$r_{Cl,CLCR}$ (P10)	$r_{Cl,CLCR}$ (P90)	$r_{V,Weight}$ (P10)	$r_{V,Weight}$ (P90)	$r_{V,Sex}$				
Value	0.62	1.66	0.65	1.44	0.70				

$\mu_.$: refers to the typical value, $\beta_.$: covariate parameter, $\omega_.$: standard deviation of the random effect, $r_{l,c}$: ratio on the parameter l for the covariate c

$\log Weight$ and Sex on V were modeled as additive on the log-scale, as shown in equation (18) where $\eta_{Cl,i} \sim \mathcal{N}(0, \omega_{Cl}^2)$ and $\eta_{V,i} \sim \mathcal{N}(0, \omega_V^2)$ and with $Sex_i = 0$ if the i^{th} subject is a Male and 1 otherwise.

$$\begin{cases} \log Cl_i = \log \mu_{Cl} + \beta_{Cl,CLCR} \log \left(\frac{CLCR_i}{CLCR_{pop}} \right) + \eta_{Cl,i} & \text{where } CLCR_{pop} = \left(\prod_i CLCR_i \right)^{\frac{1}{N}} \\ \log V_i = \log \mu_V + \beta_{V,Weight} \log \left(\frac{Weight_i}{Weight_{pop}} \right) + \beta_{V,Sex} Sex_i + \eta_{V,i} & \text{where } Weight_{pop} = \left(\prod_i Weight_i \right)^{\frac{1}{N}} \end{cases} \quad (18)$$

The PK parameter value used were the estimated values for a model without Sex effect on the initial dataset and are given in Table 1. Sex effect on V was arbitrarily set to $\beta_{V,Sex} = -0.35$. The ratio values corresponding to the covariate parameters are also given in Table 1. It should be noted that all the covariate effects considered were statistically significant and clinically relevant.

3.1.3 Scenarios

Concentration evolution was studied in the reference scenario and in three other scenarios chosen to be more challenging. Those 4 different scenarios are obtained combining two design options and two variances options for random effects. Design 1 is the initial design of the study, and is composed of $N = 100$ subjects, divided equally between the 100 mg and 250 mg arms, with 16 sampling time points from 0.25 to 24h post dose. Design 2 is composed of $N = 24$ subjects only, also divided equally between the 100 mg and 250 mg arms, with 3 time points at 1, 4 and 12h post dose. For the variance, the first option, called True Omega keep the values ω_{Cl} and ω_V given in Table 1, while the second, called High Omega, doubled these standard deviations.

3.2 Prediction from FIM

The *PFIM6* source code has been extended in a working version to take account of both discrete and continuous covariates.

The FIM was computed by linearization of the structural model with first-order approximation and the three methods to handle covariates were considered. The first one used the 100 covariate vectors from the data: in the 4 scenarios, the FIM is computed with $S = 100$ by computing the elementary FIM for each of the vectors. The other two methods simulated Z_1, \dots, Z_S with $S = 1000$ covariate vectors, using either independent distributions fitted on the data, namely independent Gaussian distributions for continuous covariates and Bernoulli distribution for discrete covariates, or two copulas fitted on the covariate table, one for each sex. Vine copula were fitted using the R package *rvinecopulib* 0.6.3.1.1 [25] and the code made available by other researchers at https://github.com/vanhassel1tab/copula_vps [zwep2022virtual].

The Figure 4 given in appendix shows the fitted joint distribution of $\log \left(\frac{CLCR_i}{CLCR_{pop}} \right)$ and $\log \left(\frac{Weight_i}{Weight_{pop}} \right)$ for Males and Females, with the Independent Gaussian approach and with Copulas. As expected, regarding marginal densities, the two methods give similar distributions while only Copulas can capture the correlation between the two covariates very precisely.

The predicted SE were computed as the square root of the diagonal elements of the FIM and used to compute the predicted NSE=RSE on the parameters: $NSE^{PFIM}(\psi_0) = 100 \frac{SE^{PFIM}}{\psi_0}$. The predicted 90% CI on the ratios were also derived, as described in 2.3. The predicted power of significance test at level 95% and power of relevance test for the interval [0.80; 1.25] were calculated, as given in equations (14) and (16).

All other things being identical (i.e., same sampling scheme, same dosing and same covariate distributions), *PFIM* was also used to compute the power of comparison and relevance tests, for the simulation value $\beta_{Cl,CLCR} = 0.87$ (corresponding to the ratio $r_{Cl,CLCR} (P10) = 0.62$) as a function of N , for a smaller effect $\beta_{Cl,CLCR} = 0.66$ ($r_{Cl,CLCR} (P10) = 0.70$) and for a stronger effect $\beta_{Cl,CLCR} = 1.3$ ($r_{Cl,CLCR} (P10) = 0.50$).

3.3 Simulation

The same scenarios were explored by CTS in order to compare the estimates with the FIM predictions.

3.3.1 Implementation

$S = 200$ datasets were simulated with R. For each subject, covariates were kept, and random effects on CI and V , namely $\eta_{CI,i}$, $\eta_{V,i}$, and the residual error ϵ_i were simulated.

3.3.2 Estimations

Estimations were performed using SAEM algorithm implemented in Monolix 2023R1 [26], with default settings regarding auto-stop criteria, number of iterations, number of chains, and simulated annealing. Starting points were set to the simulation values for the base parameters and to 0 for the covariate effects. The expected FIM was computed within Monolix, by linearization around the Empirical Bayes Estimates (EBE) of the individual parameters.

The accuracy of parameter estimates was assessed by computing the relative estimation errors (REE) and by evaluating the Relative Bias (RB) and the Relative Root Mean Square Error (RRMSE) according to equation (19), where $\hat{\psi}$ and ψ_0 are the estimated and simulated parameters, respectively.

$$RB = \frac{1}{S} \sum_{s=1}^S REE_s \quad \text{where } REE_s = \frac{\hat{\psi}_s - \psi_0}{\psi_0} \times 100 \quad RRMSE = \sqrt{\frac{1}{S} \sum_{s=1}^S (REE_s)^2} \quad (19)$$

For each dataset, estimated SE, SE_s^{estim} , were calculated using the inverse of the expected FIM obtained by linearization. SE_s^{estim} were normalized with respect to the true value ψ_0 used for simulation: $NSE_s^{estim}(\psi_0) = 100 \frac{SE_s^{estim}}{\psi_0}$.

The Empirical normalized SE across the datasets, $NSE(\psi_0)^{emp}$ were also computed as the mean of the normalized SE.

Estimated RSE, RSE_s^{estim} , were computed: $RSE_s^{estim} = 100 \frac{SE_s^{estim}}{\hat{\psi}_s}$ and the Empirical RSE RSE_ψ^{emp} across the datasets were computed as the mean of the estimated RSE.

Ratio and their 90% CI were computed using the point estimates $\hat{\beta}_s$ and the SE_s^{estim} .

Comparison test and relevance test were performed for each of the covariate effect. For the comparison test, the null hypothesis was rejected if $|\frac{\hat{\beta}}{SE_s^{estim}}| \geq q_{1-\alpha/2}$ with $\alpha = 0.05$. For the relevance test, the null hypothesis was rejected if the 90%CI on the ratio was outside [0.80; 1.25].

3.4 Evaluation methods

RSE predicted with PFIM were compared to median and quantiles of estimated NSE and RSE. Predicted CI on ratios were compared to median and quantiles of estimated bounds of CI on ratio. Predicted Powers were compared to mean and 95% CI on mean of estimated powers.

3.5 Results

3.5.1 Estimation accuracy

In the 4 scenarios, all covariate parameter estimates were unbiased, see Figure 5 in the Supplementary, where REE boxplots were centered on 0. With Design 2 (i.e., fewer subjects and sparser observations), variance parameters had small negative bias (with True Omega, $RB(\omega_{CI}) = -20.1\%$ and $RB(\omega_V) = -9.99\%$; with High Omega, $RB(\omega_{CI}) = -11.5\%$ and $RB(\omega_V) = -9.31\%$).

3.5.2 Uncertainty evaluation

Uncertainty was well predicted with PFIM, see Figure 6 in the Supplementary. Indeed, the three methods for handling covariates in the FIM gave similar results, as in all the scenarios and for almost all parameters, PFIM predictions were very close to empirical NSE and RSE .

With Design 1, for both True Omega and High Omega, SE were very well predicted by PFIM. Of note, with High Omega the empirical RSE for $\beta_{V,SEX}$ was slightly larger than PFIM prediction and empirical NSE .

With Design 2, estimated RSE showed a large variability and because of few extreme values, the empirical RSE for some parameters were quite larger than the median RSE which remained close to the empirical NSE . With Design 2 and both True Omega and High Omega, SE were overpredicted for the residual error parameters compared to the empirical NSE but were smaller than the empirical RSE . For covariate parameters, PFIM predictions were in accordance with empirical NSE for the covariate effects. We can note that PFIM prediction for SE on $\beta_{V,SEX}$ were slightly higher than the empirical NSE and the median RSE .

3.5.3 Forest plots

Forest plots are shown on Figure 1, with PFIM predictions according to the three methods for handling covariates in green. The distributions of the estimated ratios are shown in red boxplots while the estimated bounds of the 90% CI on the ratios are shown in purple boxplots.

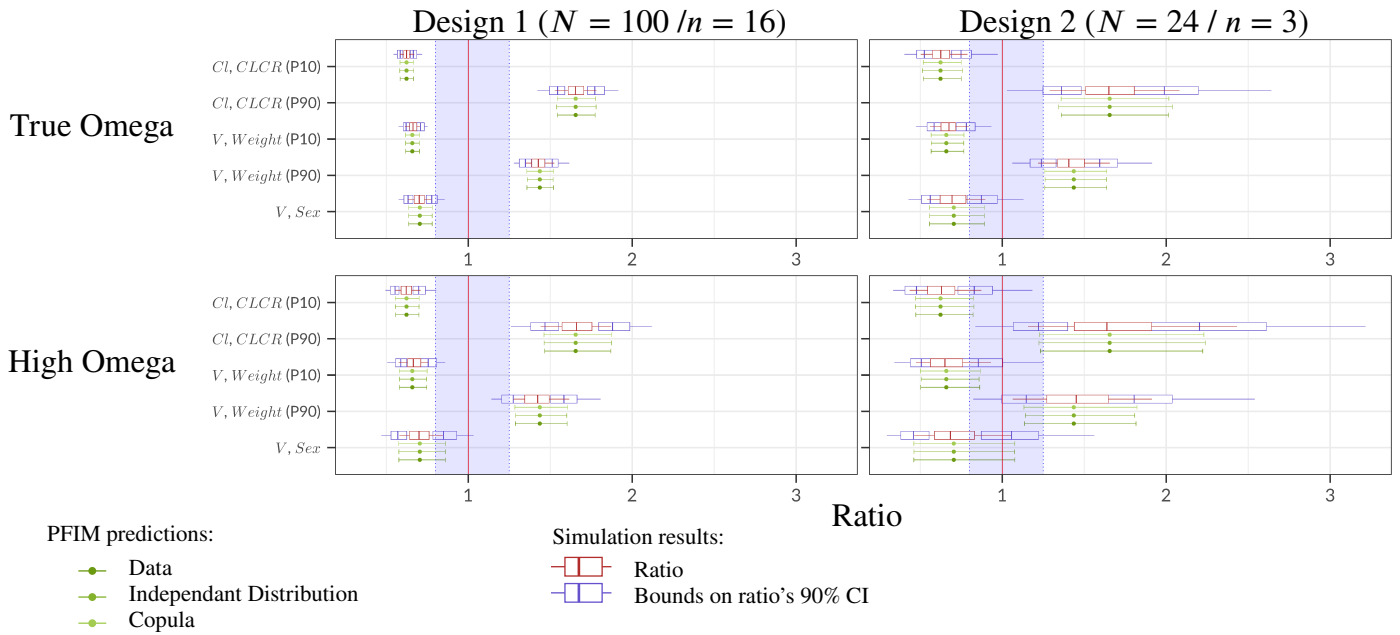


Figure 1: Evaluation - Forest Plots for the 4 scenarios: PFIM predictions using the three methods for handling covariates and simulation results across 200 datasets

The red line at 1 corresponds to the reference line i.e., no change from the typical individual; the shaded area in blue represents the reference area of [0.80, 1.25].

The boxplot displays the median, the 25th and 75th percentiles, while the whiskers are 5th and 95th percentiles.

The confidence intervals for the ratios were overall well predicted for all the ratios, as the bounds of predicted CI were always very close to the median of the estimated bounds.

With Design 1 and True Omega, all the ratios were predicted relevant. When increasing IIV to High Omega, the CI of *Sex* effect on *V* crossed the [0.80; 1.25] area, so one cannot conclude to relevance.

In accordance with the slightly overprediction of SE on $\beta_{V,Sex}$, it can be noted that with Design 2, the predicted CI on the ratio of *Sex* effect on *V* was larger than the CI formed by the respective median of the estimated lower and higher bounds of the 90%CI. In Design 2, the CI of the effect of the 90th percentile of *CLCR* and *Weight*, respectively on *Cl* and *V* were slightly larger than the estimated CI.

With Design 2 and True Omega, only the CI of *Sex* effect on *V* crossed the [0.80; 1.25], therefore the effects of *CLCR* on *Cl* and *Weight* on *V* were predicted relevant for the two percentiles. When increasing IIV to High Omega, none of the parameter was predicted relevant.

3.5.4 Power

Significance test For the 4 scenarios, the predicted power for significance test derived from the SE computed from the FIM and the estimated power computed as the mean of the rejection of the null hypothesis across the datasets are shown in Table 2.

First, as the SE were very close between the three methods for handling the covariates in the FIM computation, the powers were also the same for the three approaches.

With Design 1 and True Omega, comparison test on all the β parameters were predicted to 1 and estimated to 1 by CTS. Thus PFIM predictions were right.

With Design 1 and High Omega, power were very slightly overpredicted for $\beta_{Cl,CLCR}$ and $\beta_{V,Weight}$, with estimated 95% CI being respectively [0.90; 0.97] and [0.94; 0.99] and predictions being 0.99 and 1. For $\beta_{V,Sex}$, the estimated power was 0.72 and FIM prediction was 0.69, thus lying in the estimated power's 95% CI which was [0.65; 0.78].

With Design 2 (i.e., fewer subjects and sparser observations), PFIM prediction were very close to the estimated power and always within its 95% CI in case of both True Omega and High Omega.

Relevance test For the 4 scenarios, the predicted power for relevance test derived from the SE computed from the FIM and the estimated power computed as the mean of the rejection of the null hypothesis across the datasets are shown in Table 3.

As for significance test, with Design 1 and True Omega, power of relevance tests were close to 1 and well predicted for effects of *CLCR* on *Cl* and *Weight* on *V*. For *Sex* effect on *V*, PFIM prediction 0.66 lied within the 95% CI of the estimated power [0.61; 0.74].

Table 2: Evaluation - Power of significance test for the 4 scenarios: estimated power and its 95% CI across 200 datasets and PFIM prediction with the three methods for handling covariates

	Parameter	Value	Design 1 ($N = 100 / n = 16$)				Design 2 ($N = 24 / n = 3$)			
			Estimated	Data	PFIM		Estimated	Data	PFIM	
					Distributions	Copula			Distributions	Copula
True Omega	$\beta_{Cl,CLCR}$	0.87	1.00 [0.98; 1.00]	1.00	1.00	1.00	0.94 [0.90; 0.97]	0.99	0.98	0.99
	$\beta_{V,Weight}$	1.0	1.00 [0.98; 1.00]	1.00	1.00	1.00	0.98 [0.94; 0.99]	1.00	1.00	1.00
	$\beta_{V,Sex}$	-0.35	1.00 [0.98; 1.00]	1.00	1.00	1.00	0.72 [0.65; 0.78]	0.69	0.69	0.69
High Omega	$\beta_{Cl,CLCR}$	0.87	1.00 [0.98; 1.00]	1.00	1.00	1.00	0.74 [0.68; 0.80]	0.80	0.78	0.79
	$\beta_{V,Weight}$	1.0	1.00 [0.98; 1.00]	1.00	1.00	1.00	0.69 [0.62; 0.75]	0.72	0.74	0.71
	$\beta_{V,Sex}$	-0.35	0.84 [0.78; 0.89]	0.82	0.82	0.82	0.32 [0.26; 0.39]	0.27	0.28	0.27

Table 3: Evaluation - Power of relevance test for the interval [0.8, 1.25] for the 4 scenarios: estimated power and its 95% CI across 200 datasets and PFIM prediction with the three methods for handling covariates

	Parameter	Ratio	Design 1 ($N = 100 / n = 16$)				Design 2 ($N = 24 / n = 3$)			
			Estimated	Data	PFIM		Estimated	Data	PFIM	
					Distributions	Copula			Distributions	Copula
True Omega	$Cl, CLCR$ (P10)	0.62	1.00 [0.98; 1.00]	1.00	1.00	1.00	0.71 [0.64; 0.77]	0.72	0.68	0.72
	$Cl, CLCR$ (P90)	1.66	1.00 [0.98; 1.00]	1.00	1.00	1.00	0.74 [0.68; 0.80]	0.76	0.72	0.76
	$V, Weight$ (P10)	0.65	1.00 [0.98; 1.00]	1.00	1.00	1.00	0.59 [0.52; 0.66]	0.68	0.69	0.68
	$V, Weight$ (P90)	1.44	0.99 [0.96; 1.00]	0.99	0.99	0.99	0.46 [0.39; 0.53]	0.54	0.55	0.54
	V, Sex	0.70	0.68 [0.61; 0.74]	0.65	0.66	0.65	0.30 [0.23; 0.36]	0.22	0.23	0.22
High Omega	$Cl, CLCR$ (P10)	0.62	0.94 [0.90; 0.97]	0.97	0.97	0.97	0.44 [0.37; 0.51]	0.44	0.42	0.43
	$Cl, CLCR$ (P90)	1.66	0.98 [0.94; 0.99]	0.98	0.98	0.98	0.46 [0.39; 0.53]	0.47	0.45	0.46
	$V, Weight$ (P10)	0.65	0.73 [0.66; 0.79]	0.81	0.82	0.80	0.38 [0.32; 0.46]	0.32	0.33	0.32
	$V, Weight$ (P90)	1.44	0.60 [0.52; 0.66]	0.67	0.68	0.66	0.32 [0.26; 0.39]	0.25	0.26	0.25
	V, Sex	0.70	0.30 [0.24; 0.37]	0.27	0.27	0.27	0.15 [0.10; 0.21]	0.12	0.12	0.12

With Design 1 and High Omega, power of relevance tests were well predicted for effects of $CLCR$ on Cl and slightly overpredicted for effect of $Weight$ on V (respectively predicted at 0.68 with upper bound of the 95% CI at 0.66 for P10 and predicted at 0.54 with upper bound at 0.53 for P90). For Sex effect on V , PFIM prediction 0.22 was slightly below the 95% CI of the estimated power [0.23; 0.36].

Similarly, with Design 2 and True Omega, power of relevance tests were well predicted for effects of $CLCR$ on Cl and Sex effect on V but slightly overpredicted for effect of $Weight$ on V (respectively predicted at 0.81 with upper bound of the 95% CI at 0.79 for P10 and predicted at 0.67 with upper bound at 0.66 for P90).

With Design 2 and High Omega, power of relevance test were well predicted for all the parameters, just slightly underpredicted at 0.25 for $Weight$ on V P90 while the lower bound of the observed CI was 0.26.

Overall, we can conclude that the FIM predictions were very satisfactory.

Evolution of power of test Another important result from this methodology is being able to compute the power of tests depending on covariate effect magnitude, sample size, design or IIV. Thus is given on Figure 2 the power of significance test on $\beta_{Cl,CLCR}$ and the power of relevance test for effect of $CLCR$ (P10) on Cl , as function of N for different values of $\beta_{Cl,CLCR}$, sampling schemes from Design 1 and 2 and IIV setting to True Omega or High Omega.

The first foreseeable result shown by this plot is that the smaller the covariate effect, the smaller the powers of tests, which led to a higher number of subjects needed to achieve a given level of power. In addition, with sparser design (shown in red) the power of tests were always smaller than with rich design (shown in blue). The larger the IIV on parameters (dashed line), the lower the power of tests. We can also note that in this example, the variability had a greater impact than design, as switching from rich design with True Omega (blue plain line) to rich design with High Omega (blue dashed line) decreased power of both tests more than switching to sparse design with True Omega (red plain line).

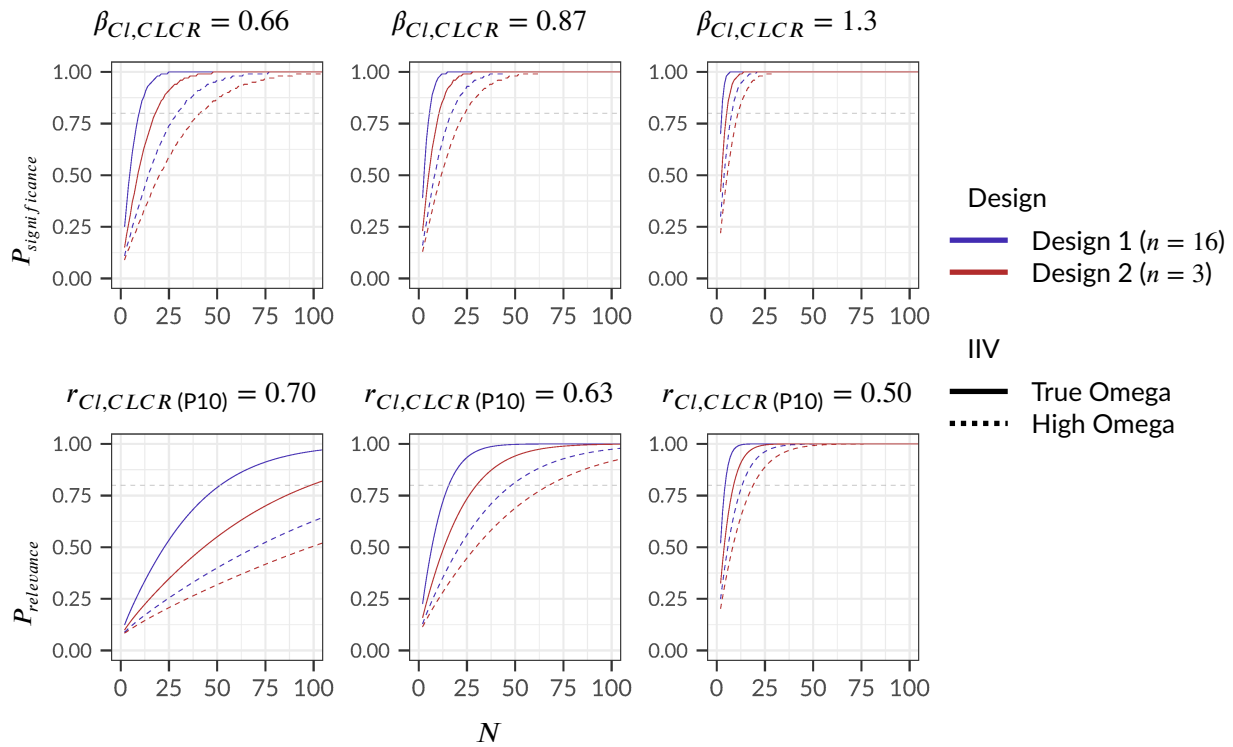


Figure 2: Evaluation - Predicted Power of significance test on $\beta_{CI,CLCR}$ and Power of relevance test for effect of $CLCR$ (P10) on CI for the interval $[0.80; 1.25]$, as function of N and for different values of $\beta_{CI,CLCR}$: PFIM predictions using Data method.

4 Application to a population PK example of cabozantinib

In this section, the aim was to validate the method to compute the FIM for model with continuous and/or discrete covariates, in a more complex set up in terms of PK model and covariate structure.

This study was inspired from a real population PK analysis [19] performed on 10 clinical trials in healthy volunteers and patients with various cancer types receiving cabozantinib daily. The original data included 9510 quantifiable cabozantinib concentrations from 2023 subjects. The table of covariate vectors from this analysis were used in the present work.

4.1 Settings

4.1.1 Covariate Data

The covariate data are pooled from the 10 clinical studies analysed in the original article. A summary is given in Table 4. Continuous covariates were Dose, Age, and Weight, while categorical covariates were Sex, Race, Population category and Formulation. It should be noted that certain categories are under-represented: for Race, only 10% of patients are Asian and only 3% Black and 2% from other race; and for pathology, only 10% of patients suffer from Metastatic medullary thyroid cancer (MTC) and 2% from Glioblastoma multiforme (GB). The baselines values from the 2023 subjects were used.

4.1.2 Design

The studied design, given in Table 8 in the Supplementary, corresponded to the theoretical sampling designs of the 10 clinical trials analysed in the original article, associated with the number of included subjects and their doses. Three trials were phase I studies, with therefore few subjects (between 40 and 70) but with rich observations (up to 43 sampling points per subjects). Data also included two phase II studies with more balance between subjects and observations, and finally five phase III studies, with much more subjects (up to 498) but much sparser design with generally two to three samplings per subjects. In total, the studied design included 11621 observations from 2023 subjects.

4.1.3 Model

For illustrative purposes and due to the current limitations of the *PFIM* package, the initial model [19] was simplified into a two-compartment model, with a first-order absorption discarding the dual absorption process, the presence of a lag time and a correlation between random effects. The analytical solution for the concentration in the central compartment

Table 4: Application - Covariates summary, data pooled from the 10 clinical studies analysed in Nguyen et al.[19] combining 2023 subjects

Continuous covariates		P50 [P10; P90]
	Age	64 [44; 76]
	Weight	78 [58; 101]
	Dose	60 [44; 140]
Categorical Covariates		N (%)
Sex	Male	1706 (84%)
	Female	317 (16%)
Race	White	1712 (85%)
	Black	53 (3%)
	Asian	211 (10%)
	OtherRace	47 (2%)
Pathology	Healthy volunteers	140 (7%)
	CRPC	823 (41%)
	RCC	282 (14%)
	MTC	210 (10%)
	GB	39 (2%)
	HCC	489 (24%)
	OtherMalign	40 (2%)
Formulation	Tablet	1375 (68%)
	CAP	648 (32%)

P50: indicates the median, P10: the 10th percentile, P90: the 90th percentile, N: the number of subjects

CRPC: Castration-resistant prostate cancer, RCC: Renal cell carcinoma, MTC: Metastatic medullary thyroid cancer, GB: Glioblastoma multiforme, HCC: Hepatocellular carcinoma, OtherMalign: other malignancy, CAP: Capsule formulation

is given in equation (20), where Ka is absorption rate constant, Q/F the apparent flow parameter between compartments, V_c/F the apparent distribution volume of the central compartment and V_p/F the apparent distribution volume of the peripheral compartment.

$$\begin{cases}
 f(t) = \frac{1}{\alpha - \beta} \frac{Ka}{V_c} F Dose \left(\left(\frac{Q/V_p - \alpha}{Ka - \alpha} + \frac{\beta - Q/V_p}{Ka - \beta} \right) e^{-Kat} - \frac{Q/V_p - \alpha}{Ka - \alpha} e^{-\alpha t} - \frac{\beta - Q/V_p}{Ka - \beta} e^{-\beta t} \right) \\
 \text{where:} \\
 \beta = \frac{1}{2} \left(\left(\frac{Q + Cl}{V_c} + \frac{Q}{V_p} \right) - \sqrt{\left(\frac{Q + Cl}{V_c} + \frac{Q}{V_p} \right)^2 - 4 \frac{ClQ}{V_p V_c}} \right) & \alpha = \frac{QCl}{V_p V_c}
 \end{cases} \quad (20)$$

The model included log-normally distributed IIV on all the PK parameters; i.e., on Ka , F , Cl , V_c , Q and V_p . The residual variability is modeled using a proportional error model with variance σ^2 i.e. $g = \sigma f$.

All the covariate relationships of the initial model have been reproduced. It should be noted, however, that the initial covariate model was built without selection, notably with the same covariates affecting both Cl and V_c , and that as a result, some covariate parameters values are very small and/or estimated with very wide confidence intervals leading to non statistically significant covariate effects. This is particularly true for the effects of categorical covariates for which the non-reference categories are very poorly represented (e.g. *MTC* patients represent 10% of the total population and *GB* only 2%).

Three continuous covariates were included in the model: *Dose*, *Age* and *Weight*. They were normalised by a reference value and log-transformed, as given in equation (21).

$$\log \left(\frac{Dose}{40} \right) \quad \log \left(\frac{Age}{60} \right) \quad \log \left(\frac{Weight}{80} \right) \quad (21)$$

Two binary covariates were included: *Sex* and Formulation, with the following definitions: *Sex* equals 0 if the subject is a Male (reference) and 1 otherwise and *CAP* equals 0 if the subject received a tablet formulation (reference), and 1 if the subject received a capsule formulation.

The categorical covariates included in the model were Race: White, Black, Asian or other race and Pathology: Healthy volunteers, Castration-resistant prostate cancer (*CRPC*), Renal cell carcinoma (*RCC*), Metastatic medullary thyroid cancer (*MTC*), Glioblastoma multiforme (*GB*), Hepatocellular carcinoma (*HCC*) and other malignancy (*OtherMalign*). These categorical covariates were converted into binary covariates by coding each category other than the reference as a separate binary variable, with 1 coding the corresponding category and 0 all others. Thus, for Race the reference was White and the categorical covariate was transformed to 3 binary covariates, respectively *Black*, *Asian* and *OtherRace*.

Table 5: Application - Parameter Values taken from Nguyen et al.[19] and used in the present work

Base parameters													
Parameter	μ_{Ka} (h ⁻¹)	$\mu_{Cl/F}$ (L/h)	$\mu_{Q/F}$ (L/h)	$\mu_{Vc/F}$ (L)	$\mu_{Vp/F}$ (L)	μ_F	ω_{ka}	ω_F	$\omega_{Cl/F}$	$\omega_{Vc/F}$	$\omega_{Q/F}$	$\omega_{Vp/F}$	σ
Value	1.24	2.48	30	213	178	1.00	1.42	0.840	0.462	0.666	0.141	0.141	0.356
Covariate parameters													
$\beta_{Ka,CAP}$	$\beta_{Ka,LDose}$	$\beta_{F,CAP}$											
-0.911	0.734	-0.166											
$\beta_{Cl/F,Sex}$	$\beta_{Cl/F,Black}$	$\beta_{Cl/F,Asian}$	$\beta_{Cl/F,OtherRace}$	$\beta_{Cl/F,CRPC}$	$\beta_{Cl/F,RCC}$	$\beta_{Cl/F,MTC}$	$\beta_{Cl/F,GB}$	$\beta_{Cl/F,HCC}$	$\beta_{Cl/F,OtherMalign}$				
-0.274	0.162	-0.0668	0.0279	-0.0115	-0.139	0.643	0.178	-0.130	0.171				
$\beta_{Cl/F,LAge}$	$\beta_{Cl/F,LWeight}$												
-0.157	-0.0393												
$\beta_{Vc/F,Sex}$	$\beta_{Vc/F,Black}$	$\beta_{Vc/F,Asian}$	$\beta_{Vc/F,OtherRace}$	$\beta_{Vc/F,CRPC}$	$\beta_{Vc/F,RCC}$	$\beta_{Vc/F,MTC}$	$\beta_{Vc/F,GB}$	$\beta_{Vc/F,HCC}$	$\beta_{Vc/F,OtherMalign}$				
0.0939	0.0441	-0.363	-0.126	-0.297	-0.422	-0.0657	-0.735	-0.166	-0.272				
$\beta_{Vc/F,LAge}$	$\beta_{Vc/F,LWeight}$												
0.0644	1.19												
Ratios of covariate effects													
$r_{Ka,CAP}$	$r_{Ka,Dose}$ (P10)		$r_{Ka,Dose}$ (P90)		$r_{F,CAP}$								
0.402	0.788		1.84		0.847								
$r_{Cl/F,Sex}$	$r_{Cl/F,Black}$	$r_{Cl/F,Asian}$	$r_{Cl/F,OtherRace}$		$r_{Cl/F,CRPC}$	$r_{Cl/F,RCC}$	$r_{Cl/F,MTC}$	$r_{Cl/F,GB}$	$r_{Cl/F,HCC}$	$r_{Cl/F,OtherMalign}$			
0.76	1.18	0.935	1.03		0.989	0.870	1.90	1.19	0.878	1.19			
$r_{Cl/F,Age}$ (P10)	$r_{Cl/F,Age}$ (P90)	$r_{Cl/F,Weight}$ (P10)		$r_{Cl/F,Weight}$ (P90)									
1.06	0.973	1.01		0.99									
$r_{Vc/F,Sex}$	$r_{Vc/F,Black}$	$r_{Vc/F,Asian}$	$r_{Vc/F,OtherRace}$		$r_{Vc/F,CRPC}$	$r_{Vc/F,RCC}$	$r_{Vc/F,MTC}$	$r_{Vc/F,GB}$	$r_{Vc/F,HCC}$	$r_{Vc/F,OtherMalign}$			
1.10	1.05	0.696	0.882		0.743	0.656	0.936	0.480	0.847	0.762			
$r_{Vc/F,Age}$ (P10)	$r_{Vc/F,Age}$ (P90)	$r_{Vc/F,Weight}$ (P10)		$r_{Vc/F,Weight}$ (P90)									
0.976	1.01	0.703		1.37									

Base parameters: refers to all model parameters except covariate effects

μ : refers to the typical value, β : covariate parameter, ω : standard deviation of the random effect

Ka : indicates absorption rate constant, Cl/F : the apparent clearance, Q/F : the apparent flow parameter between compartments, V_c/F : the apparent distribution volume of the central compartment, V_p/F : the apparent distribution volume of the peripheral compartment

$CRPC$: Castration-resistant prostate cancer, RCC : Renal cell carcinoma, MTC : Metastatic medullary thyroid cancer, GB : Glioblastoma multiforme, HCC : Hepatocellular carcinoma, $OtherMalign$: other malignancy, CAP : Capsule formulation

In the same way, for the Pathology, the reference was Healthy volunteers and 6 binary covariates were created: $CRPC$, RCC , MTC , GB , HCC , $OtherMalign$.

The reference was a healthy white Male subject with a body Weight of 80 kg, 60 years of Age, receiving a 40-mg tablet dose once daily.

The effects of the continuous covariates ($Dose$, Age and $Weight$) and of the binary covariates (Sex , CAP , $Black$, $Asian$, $OtherRace$, $CRPC$, RCC , MTC , GB , HCC and $OtherMalign$) were all additive on the logarithmic scale, as given by equation (22), where $\beta_{\theta,z}$ denotes the effect of covariate z on the parameter θ .

$$\log \theta_i = \log \mu_{\theta} + \sum_{z \in \text{binary}} (\beta_{\theta,z} \times z_i) + \sum_{z \in \text{continuous}} \left(\beta_{\theta,z} \times \log \left(\frac{z_i}{z_{pop}} \right) \right) + \eta_{\theta,i} \quad (22)$$

The model included effects of $Dose$ and CAP on Ka , CAP on F and Age , $Weight$, Sex , $Black$, $Asian$, $OtherRace$, $CRPC$, RCC , MTC , GB , HCC and $OtherMalign$ on both Cl/F and V_c/F , which corresponded to 27 covariate effect parameters.

The values used were taken from [19] and are given in Table 5. We can already notice that some of the covariate effects were very small (e.g. $\beta_{Cl/F,Asian} = -0.0668$ corresponding to $r_{Cl/F,Asian} = 0.94$; $\beta_{Cl/F,CRPC} = -0.0115$ corresponding to $r_{Cl/F,Asian} = 0.99$; $\beta_{Cl/F,Age} = -0.157$ corresponding to $r_{Cl/F,Age}$ (P10) = 1.06 and $r_{Cl/F,Age}$ (P90) = 0.97; $\beta_{Cl/F,Weight} = -0.0393$ corresponding to $r_{Cl/F,Weight}$ (P10) = 1.01 and $r_{Cl/F,Weight}$ (P90) = 0.99; $\beta_{Vc/F,Black} = 0.0441$ corresponding to $r_{Vc/F,Black} = 1.05$; $\beta_{Vc/F,MTC} = -0.0657$ corresponding to $r_{Vc/F,MTC} = 0.94$; and $\beta_{Vc/F,Age} = 0.0644$ corresponding to $r_{Vc/F,Age}$ (P10) = 0.98 and $r_{Vc/F,Age}$ (P90) = 1.01). In addition, only a few relationships were relevant, in the sense that their ratio was outside [0.80; 1.25], those relationships were the effects of CAP on Ka , both $Dose$ (P10) and (P90) on Ka , Sex on Cl/F , MTC on Cl/F , $Asian$ on V_c/F , $CRPC$ on V_c/F , GB on V_c/F , $OtherMalign$ on V_c/F and both $Weight$ (P10) and (P90) on V_c/F , i.e. 11 ratios out of 32.

4.2 Prediction from FIM

Because the three methods for handling covariates have shown similar performances in the first simulation study (see Section 3), only the Data method was kept. Thus, covariate were taken into account in the FIM computation through the covariate dataset: for each study and each dose (i.e., for each elementary design), the covariate vectors of the subjects corresponding to this elementary design were used.

As previously, from the predicted SE returned by $PFIM$, predicted NSE=RSE on the parameters were computed. Afterwards, predicted 90% CI on the ratios were derived, such as predicted powers of significance and relevance tests for the range [0.80; 1.25]. The number of subjects required to achieve 80% power, in the comparison test and the relevance test respectively, was calculated when the current predicted power was greater than 50%. These NSN hold as long as

all other things remains identical, i.e., with the same observation points, the same dosing schemes and the same covariate distributions within the trial population.

4.3 Simulation

The same scenarios was explored by CTS in order to compare the estimates with the FIM predictions. $S = 200$ datasets were simulated with the R API of Simulix [27] and using the initial covariate dataset.

4.3.1 Estimations

Estimations were performed using SAEM algorithm implemented in Monolix 2023R1 [26], with default settings regarding auto-stop criteria, number of iterations, and Simulated annealing; and 5 Markov chains. The starting point were set to the parameter values used for the simulation for the μ, ω and σ , and to 0 for the covariate effects β . μ_F was fixed to 1, and IIV on Q and V_p were fixed to $\omega = 0.141$.

The accuracy of parameter estimates was assessed by computing the REE, the RB and the RRMSE.

As detailed in the section 4, for each dataset, estimated NSE and RSE were computed for all the parameters and the empirical NSE and RSE were secondly derived.

For each covariate parameter, significance test was performed through a Wald test of comparison with $\alpha = 0.05$. Ratio and their 90% CI were computed and relevance tests are performed for the interval [0.80; 1.25].

4.3.2 Evaluation methods

RSE predicted with PFIM were compared to median and quantiles of estimated NSE and RSE. Predicted CI on ratios were compared to median and quantiles of estimated bounds of CI on ratio. Predicted Powers were compared to mean and 95% CI on mean of estimated powers.

4.4 Results

4.4.1 Estimation accuracy

The distribution of REE for base parameters and ratios on covariate effects are given in the Supplementary, respectively in Figures 7 and 8 and in Tables 9 and 10 assessing that parameters were overall accurately and precisely estimated.

Of note, some $RRMSE$ were very large for covariate parameters: over 20% and up to more than 800% for very small β (e.g. $\beta_{CL,CRPC} = -0.0115$, $RRMSE = 808\%$) and some biases over 10%, but one again for very small β (e.g. $\beta_{CL,CRPC} = -0.0115$, $RB = -54.6\%$).

4.4.2 Uncertainty evaluation

Uncertainty was well predicted with PFIM, see Figure 9 in the Supplementary for base parameters and 10 for covariate parameters. For base parameters, once again the Data method for handling covariates in the FIM gave SE predictions close to empirical NSE and RSE , except for μ_Q , for which the predicted SE was 13.4% while the empirical NSE and RSE were 10.3%.

For covariate parameters, we can first notice at first glance that the SE, both predicted and estimated, were on the whole much larger than in the previous example, particularly for the effects of discrete covariates, some categories of which were poorly represented in the dataset (e.g. Asian, OtherRace). Indeed, for 12 covariate parameters out of 27 SE were above 100%, even reaching more than 800% for $\beta_{V_c/F,Black}$ and $\beta_{CI/F,CRCP}$, recalling that those effects were very small, $\beta_{V_c/F,Black} = 0.0441$ and $\beta_{CI/F,CRCP} = -0.0115$ and that they were only 53 Black subjects (3%). In addition, there was a large uncertainty in empirical RSE and for 21 parameters out of 27, the empirical RSE was much larger than the median of estimated RSE . This is because the covariate parameters can be estimated very close to zero, which leads to a very high quantity when they are used as the denominator in a calculation. Nonetheless, for all the covariate parameters, PFIM prediction remained very close to the empirical NSE , and if there were differences, the orders of magnitude were always the same.

4.4.3 Forest plots

Forest plots are shown on Figure 3, with PFIM predictions according to the Data method for handling covariates in green. The distributions of the estimated ratios are shown in red boxplots while the estimated bounds of the 90% CI on the ratios are shown in purple boxplots.

As in the previous example, confidence intervals for the ratios were overall well predicted, even though there was an overall tendency for the upper bound of the interval to be slightly underpredicted. This can be explained if the covariate parameter was estimated with a slight bias leading to a slight bias on the ratio (which can be seen on the forest plot) and if $SE_\beta \approx S\hat{E}_\beta$. Then multiplying the ratio with $e^{q_{1-\alpha/2} SE_\beta} > 0$ for computing the upper bound of the confidence interval inflated the bias between the predicted bound and the estimated bound. For instance, for effect of GB on CI/F , bias on $\beta_{CI/F,GB}$ was 7.73% and FIM prediction for the SE was very close to the median RSE which led to a higher estimated bound for the CI on the ratio.

On the contrary, for V_c/F , RCC , the predicted upper bound on the CI was higher than the estimated median because of the difference between the predicted SE 242% higher than the median RSE : 192%. However, for such large uncertainties, this was still equivalent.

In this example, with the given design, very few covariate effects were relevant for the interval [0.80, 1.25]. According to PFIM prediction, the effects of CAP on Ka ; $Dose$ (P90) on Ka , and MTC on Cl/F were relevant. This also shows that the covariate may be relevant for certain percentiles only.

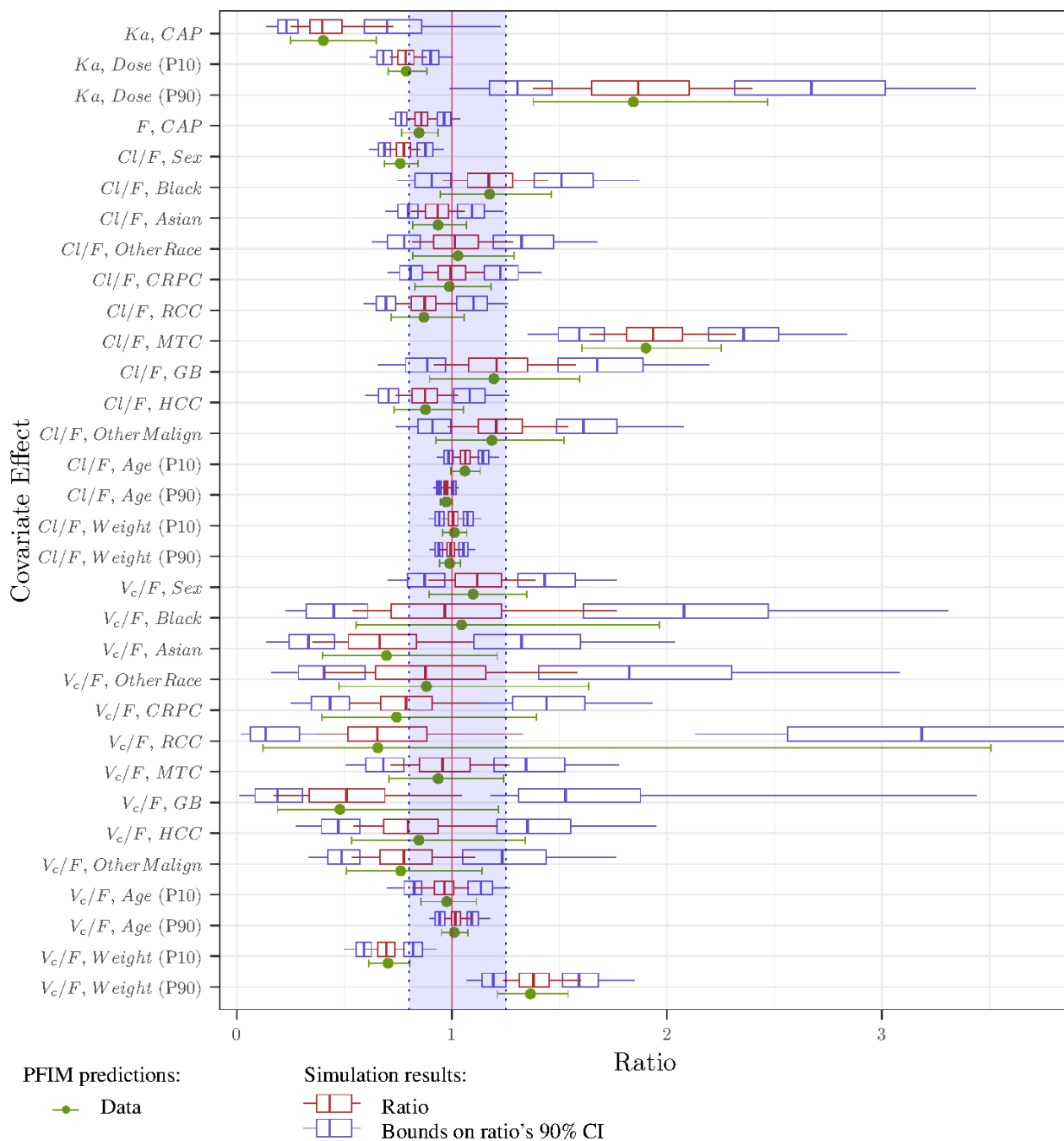


Figure 3: Application - Forest Plot: Ratios and their Confidence intervals: PFIM predictions using Data method for handling covariate and simulation results across 200 datasets

The red line at 1 corresponds to the reference line i.e., no change from the typical individual; the shaded area in blue represents the reference area of [0.80, 1.25].

The boxplot displays the median, the 25th and 75th percentiles, while the whiskers are 5th and 95th percentiles.

4.4.4 Power

Significance test The predicted power for significance test derived from the SE computed from the FIM and the estimated power are shown in Table 6. PFIM predictions were very good, as lying in the 95% CI of the estimated power for 22 parameters out of 27. For $\beta_{CI/F,OtherRace}$ and $\beta_{V_c/F,Sex}$, the prediction was slightly below the CI; for $\beta_{V_c/F,CRPC}$ and $\beta_{V_c/F,RCC}$ the prediction was slightly above while for $\beta_{V_c/F,GB}$ it was above: 0.25 against [0; 0.03]. This results comes from the bias in the ratio visible on the forest plot were the median of the estimated ratio is higher than the true value, and from the SE that was slightly under predicted: 77.0%, compared to empirical *NSE* 89.2% and median *RSE* 82.7%. However, it should be remembered that only 39 subjects (2%) were GB, which explains the difficulties in estimating this parameter.

The *NSN* to achieve a power of 80% was computed for covariate parameters for which the power of significance was already above 50% with the 2023 subjects. Only 6 covariate parameters matched this criteria: $\beta_{Ka,CAP}$, $\beta_{Ka,Dose}$, $\beta_{F,CAP}$, $\beta_{CI/F,Sex}$, $\beta_{CI,MTC}$ and $\beta_{V_c/F,Weight}$, with respective powers even above 80%, therefore the *NSN* to reach a power of 80% were smaller than 2023. All other things being identical (i.e., same sampling scheme, same dosing and same covariate distributions), for instance only 409 subjects would have been enough to conclude statistical significance for $\beta_{CI/F,MTC}$. As 10% of the initial population were MTC patients, this would roughly correspond to 41 MTC patients and 369 patients from other categories, respecting their initial proportions.

Relevance test The predicted power for relevance test derived from the SE computed from the FIM and the estimated power are shown in Table 7. PFIM predictions were overall slightly higher than what was really observed.

The *NSN* to achieve a power of 80% was computed for the ratios for which the power of relevance was already above 50% with the 2023 subjects, i.e., the effects of *CAP* on *Ka*; *Dose* (P90) on *Ka*, and *MTC* on *CI/F*.

For the effects of *CAP* and *Dose* (P90) on *Ka* relevance power were respectively 77% and 78% and the *NSN* to reach 80% were respectively 2 228 and 2 598. Thus, all other things being identical, just over 600 additional subjects would have been needed to conclude that *Ka*, *CAP* and *Ka* and *Dose* (P90) were relevant, with a power of 80%. *MTC* having a strong effect on *CI/F* ($r_{CI/F,MTC} = 1.90$), the power of relevance was 99% and all other things being identical the level 80% would have been reached for only 756 subjects (including 10% of *MTC*).

Table 6: Application - Significance test: estimated power (with 95% CI) across 200 datasets, each including 2023 subjects, PFIM predicted power and number of subjects needed NSN to achieve a 80% power using the Data method

Parameter	Value	Estimated	PFIM	NSN
$\beta_{Ka,CAP}$	-0.911	0.88 [0.83; 0.92]	0.88	1 612
$\beta_{Ka,Dose}$	0.734	0.94 [0.90; 0.97]	0.93	1 331
$\beta_{F,CAP}$	-0.166	0.78 [0.72; 0.84]	0.79	2 090
$\beta_{CI/F,Sex}$	-0.274	0.99 [0.96; 1.00]	0.99	845
$\beta_{CI/F,Black}$	0.162	0.23 [0.18; 0.30]	0.23	
$\beta_{CI/F,Asian}$	-0.0668	0.18 [0.13; 0.24]	0.13	
$\beta_{CI/F,OtherRace}$	0.0279	0.07 [0.04; 0.12]	0.05	
$\beta_{CI/F,CRPC}$	-0.0115	0.03 [0.01; 0.06]	0.05	
$\beta_{CI/F,RCC}$	-0.139	0.20 [0.14; 0.26]	0.22	
$\beta_{CI/F,MTC}$	0.643	1.00 [0.98; 1.00]	1.00	409
$\beta_{CI/F,GB}$	0.178	0.22 [0.17; 0.29]	0.17	
$\beta_{CI/F,HCC}$	-0.13	0.22 [0.16; 0.28]	0.22	
$\beta_{CI/F,OtherMalign}$	0.171	0.23 [0.18; 0.30]	0.20	
$\beta_{CI/F,Age}$	-0.157	0.30 [0.24; 0.37]	0.32	
$\beta_{CI/F,Weight}$	-0.0393	0.05 [0.02; 0.09]	0.06	
$\beta_{V_c/F,Sex}$	0.0939	0.18 [0.13; 0.25]	0.12	
$\beta_{V_c/F,Black}$	0.0441	0.06 [0.03; 0.10]	0.05	
$\beta_{V_c/F,Asian}$	-0.363	0.14 [0.09; 0.19]	0.19	
$\beta_{V_c/F,OtherRace}$	-0.126	0.09 [0.05; 0.14]	0.06	
$\beta_{V_c/F,CRPC}$	-0.297	0.05 [0.02; 0.09]	0.12	
$\beta_{V_c/F,RCC}$	-0.422	0.01 [0.00; 0.04]	0.07	
$\beta_{V_c/F,MTC}$	-0.0657	0.06 [0.03; 0.10]	0.07	
$\beta_{V_c/F,GB}$	-0.735	0.00 [0.00; 0.03]	0.25	
$\beta_{V_c/F,HCC}$	-0.166	0.06 [0.03; 0.10]	0.09	
$\beta_{V_c/F,OtherMalign}$	-0.272	0.18 [0.13; 0.24]	0.20	
$\beta_{V_c/F,Age}$	0.0644	0.08 [0.05; 0.13]	0.06	
$\beta_{V_c/F,Weight}$	1.19	1.00 [0.98; 1.00]	0.99	870

CRPC: indicates Castration-resistant prostate cancer, RCC: Renal cell carcinoma, MTC: Metastatic medullary thyroid cancer, GB: Glioblastoma multiforme, HCC: Hepatocellular carcinoma, OtherMalign: other malignancy, CAP: Capsule formulation

Table 7: Application - Relevance test: estimated power (with 95% CI) across 200 datasets, each including 2023 subjects, PFIM predicted power and number of subjects needed NSN to achieve a 80% power using the Data method

Parameter	Ratio	Estimated	PFIM	NSN
Ka, CAP	0.40	0.66 [0.58; 0.72]	0.77	2 228
$Ka, Dose$ (P10)	0.79	0.02 [0.01; 0.05]	0.08	
$Ka, Dose$ (P90)	1.84	0.62 [0.55; 0.69]	0.71	2 598
F, CAP	0.85	0.00 [0.00; 0.02]	0.00	
$Cl/F, Sex$	0.76	0.09 [0.05; 0.14]	0.20	
$Cl/F, Black$	1.18	0.01 [0.00; 0.04]	0.02	
$Cl/F, Asian$	0.94	0.00 [0.00; 0.02]	0.00	
$Cl/F, OtherRace$	1.03	0.01 [0.00; 0.04]	0.00	
$Cl/F, CRPC$	0.99	0.00 [0.00; 0.02]	0.00	
$Cl/F, RCC$	0.87	0.00 [0.00; 0.02]	0.01	
$Cl/F, MTC$	1.90	1.00 [0.97; 1.00]	0.99	756
$Cl/F, GB$	1.19	0.01 [0.00; 0.04]	0.03	
$Cl/F, HCC$	0.88	0.00 [0.00; 0.02]	0.01	
$Cl/F, OtherMalign$	1.19	0.01 [0.00; 0.04]	0.02	
$Cl/F, Age$ (P10)	1.06	0.00 [0.00; 0.02]	0.00	
$Cl/F, Age$ (P90)	0.97	0.00 [0.00; 0.02]	0.00	
$Cl/F, Weight$ (P10)	1.01	0.00 [0.00; 0.02]	0.00	
$Cl/F, Weight$ (P90)	0.99	0.00 [0.00; 0.02]	0.00	
$V_c/F, Sex$	1.10	0.00 [0.00; 0.03]	0.00	
$V_c/F, Black$	1.05	0.01 [0.00; 0.04]	0.03	
$V_c/F, Asian$	0.70	0.00 [0.00; 0.03]	0.11	
$V_c/F, OtherRace$	0.88	0.04 [0.02; 0.08]	0.03	
$V_c/F, CRPC$	0.74	0.00 [0.00; 0.03]	0.07	
$V_c/F, RCC$	0.66	0.00 [0.00; 0.02]	0.08	
$V_c/F, MTC$	0.94	0.00 [0.00; 0.03]	0.01	
$V_c/F, GB$	0.48	0.00 [0.00; 0.02]	0.23	
$V_c/F, HCC$	0.85	0.01 [0.00; 0.04]	0.03	
$V_c/F, OtherMalign$	0.76	0.03 [0.01; 0.06]	0.07	
$V_c/F, Age$ (P10)	0.98	0.00 [0.00; 0.02]	0.00	
$V_c/F, Age$ (P90)	1.01	0.00 [0.00; 0.02]	0.00	
$V_c/F, Weight$ (P10)	0.70	0.38 [0.31; 0.45]	0.47	
$V_c/F, Weight$ (P90)	1.37	0.26 [0.20; 0.32]	0.33	

P10: indicates the 10th percentile, P90: the 90th percentile

CRPC: Castration-resistant prostate cancer, RCC: Renal cell carcinoma, MTC: Metastatic medullary thyroid cancer, GB: Glioblastoma multiforme, HCC: Hepatocellular carcinoma, OtherMalign: other malignancy, CAP: Capsule formulation

5 Discussion

In this paper, we extended FIM computation accounting for continuous covariates. in the context of NLMEM. To that purpose, we introduced three methods for taking covariates into account when computing the FIM, and implemented them in a working version of the R package *PFIM*. This implementation is available (Zenodo repository <https://doi.org/10.5281/zenodo.13692989>) and could be used by others for further exploration. These methods are all based on Monte Carlo integration, but the first one relies on leveraging a covariate dataset directly, while the second one simulates from independent distribution and the last one simulates from copulas. For the latter, we used the code developed by other authors [[zwep2022virtual](#)] and made available online: https://github.com/vanhasseltlab/copula_vps. We showed that overall *PFIM* accurately predicts uncertainty on covariate effects and consequently the power of both significance and relevance tests. This method can then be used to display these predictions in the form of a forest plot for visual communication. We also used this approach to quantitatively describe that the power of both significance and relevance tests decrease when reducing the number of subjects or reducing the number of observations or increasing the IIV. The results of the three methods for handling covariates in FIM computation showed very similar performances in our simulation study, but their respective impact should be further explored in complex cases. Especially, if the covariates are highly correlated, we can expect the calculation using independent distributions to perform less well. In addition, we have chosen to propose an approach that fits independent covariates with Gaussian distributions rather than using a multivariate normal distribution to explore, using copulas, whether correlations play a role in this case. Nevertheless, we can imagine an intermediate method using a multivariate Gaussian, which are easier to interpret than Copulas.

The method was utilized in a practical example based on a population PK analysis of the drug cabozantinib, conducted across 10 clinical trials involving healthy volunteers and patients with various types of cancer. With its 27 relationships and many very low-value covariate parameters, this application represents, unlike the evaluation example, a case where there is a real challenge in determining the relevant effects. The proposed methodology was used to evaluate the theoretical design, computing the power of tests and the number of subjects needed to achieve 80% power. We found that despite the large number of relationships between covariates, some of which were not significant, and with certain covariates poorly represented in the data, *PFIM* was able to predict uncertainty very well and was very close to the CTS results. It is therefore an appropriate approach for rapidly computing the powers and the number of subjects required to achieve given powers, even in more complex contexts.

A limitation of the proposed methodology is that there is no consensus on how to take uncertainty into account when building forest plots, and simulations-based procedures often account for uncertainty in based parameters which is not the case here. Sometimes, the IIV is also considered in the simulations [2], even if some authors suggest not doing so in order to avoid misunderstanding [28]. Therefore, before using this method to predict forest plot, or comparing it with simulations, the randomness to be accounted for needs to be clearly defined.

In addition, the calculation of the confidence interval on the ratio is exact only in certain particular cases, otherwise an approximation, for example with the delta method, would be necessary. Moreover, our method is also less straightforward for deriving ratios and CI for secondary parameters such as C_{max} or AUC and would require stochastic computations.

The main limitation of our methodology is the a priori knowledge of covariable distributions to compute the FIM. Indeed, without their distribution in the target population, the design evaluation cannot be performed. Nevertheless, the main use of this approach may be to design a phase III study based on phase II results, and at this stage of pharmaceutical development, expert usually have clues about the covariate distribution in the targeted population. In addition, if methods based on provided distributions is chosen, data are not mandatory as distributions can for instance be chosen according to expert opinion. In the same way, as copula do not require data, we can imagine using copulas from other projects or from other sources but corresponding to the target population of the trial. However, problems arise in the presence of discrete covariates, since the theory of copulas has not yet been fully extended to them, especially for non-ordered categorical data [29]. However, this is an active research area in pharmacometrics, and dedicated research and R package were recently published [29], it is therefore likely that solutions will be proposed in the near future.

Finally, these methods save a considerable amount of time compared with simulations, enabling a wider range of designs to be evaluated. Therefore, this tool can be used to study changes in the power of tests as a function of sample size, IIV, effect size or sampling design before conducting a trial to help in decision making. The speed of calculation could be further increased, for example by using efficient integration schemes (e.g. Quasi-Monte Carlo methods) or by deriving an analytical formula of the linearized FIM in the case where the continuous covariates follow a multivariate Gaussian distribution. These promising results suggest that the method should be more widely implemented in the *PFIM* package. In addition, this work opens the prospect of optimizing the distribution of covariates in clinical trials with the aim of maximizing power of relevance tests.

Acknowledgments

This work was financed by a CIFRE agreement (Conventions Industrielles de Formation par la Recherche) of the ANRT (Association Nationale de la Recherche et de la Technologie). The CIFRE agreement is a partnership between a public laboratory and a company, here the UMR (Unité Mixte de Recherche) 1137 and Ipsen, respectively. The authors would like to thank the entire *PFIM* group, and in particular Romain Leroux, *PFIM6* development manager of *PFIM6*, for their help in adapting the R code for our use.

Conflict of interest

L.F and K.B. are full-/part-time employees of Ipsen.

References

1. Mould, D. R. & Upton, R. Basic concepts in population modeling, simulation, and model-based drug development. *CPT: Pharmacometrics & Systems Pharmacology* **1**, 1–14 (2012).
2. Marier, J.-F., Teuscher, N. & Mouksassi, M.-S. Evaluation of covariate effects using forest plots and introduction to the coveffectsplot R package. *CPT: Pharmacometrics & Systems Pharmacology* **11**, 1283–1293 (2022).
3. *FDA Guidance for Industry Population Pharmacokinetics* <https://www.fda.gov/media/128793/download>. 2022.
4. Menon-Andersen, D *et al.* Essential pharmacokinetic information for drug dosage decisions: a concise visual presentation in the drug label. *Clinical Pharmacology & Therapeutics* **90**, 471–474 (2011).
5. Schuirmann, D. J. 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* **15**, 657–680 (1987).
6. Sanghavi, K. *et al.* Covariate modeling in pharmacometrics: General points for consideration. *CPT: Pharmacometrics & Systems Pharmacology* (2024).
7. Atkinson, A., Donev, A. & Tobias, R. *Optimum experimental designs, with SAS* (OUP Oxford, 2007).
8. Mentre, F., Mallet, A. & Baccar, D. Optimal design in random-effects regression models. *Biometrika* **84**, 429–442 (1997).
9. Strömberg, E. A., Nyberg, J. & Hooker, A. C. The effect of Fisher information matrix approximation methods in population optimal design calculations. *Journal of Pharmacokinetics and Pharmacodynamics* **43**, 609–619 (2016).
10. Mentré, F. *et al.* Current use and developments needed for optimal design in pharmacometrics: a study performed among DDMoRe’s European Federation of Pharmaceutical Industries and Associations members. *CPT: Pharmacometrics & Systems Pharmacology* **2**, 1–2 (2013).
11. Retout, S., Comets, E., Samson, A. & Mentré, F. Design in nonlinear mixed effects models: optimization using the Fedorov–Wynn algorithm and power of the Wald test for binary covariates. *Statistics in Medicine* **26**, 5162–5179 (2007).
12. Nguyen, T. T., Bazzoli, C. & Mentré, F. Design evaluation and optimisation in crossover pharmacokinetic studies analysed by nonlinear mixed effects models. *Statistics in Medicine* **31**, 1043–1058 (2012).
13. Dumont, C. *et al.* PFIM 4.0, an extended R program for design evaluation and optimization in nonlinear mixed-effect models. *Computer Methods and Programs in Biomedicine* **156**, 217–229 (2018).
14. Leroux, R., Seurat, J., Fayette, L., Bach, N. T. & Mentré, F. Design evaluation and optimisation in nonlinear mixed effects models with the R package PFIM 6.0. *PAGE* (2023).
15. Nyberg, J. *et al.* PopED: An extended, parallelized, nonlinear mixed effects models optimal design tool. *Computer Methods and Programs in Biomedicine* **108** (2012).
16. Foracchia, M., Hooker, A. C., Vicini, P. & Ruggeri, A. POPED, a software for optimal experiment design in population kinetics. *Computer Methods and Programs in Biomedicine* **74** (2004).
17. Bauer, R. J., Hooker, A. C. & Mentre, F. Tutorial for \$ DESIGN in NONMEM: Clinical trial evaluation and optimization. *CPT: Pharmacometrics & Systems Pharmacology* **10**, 1452–1465 (2021).
18. 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* **13**, 209–227 (2003).
19. Nguyen, L., Chapel, S., Tran, B. D. & Lacy, S. Updated population pharmacokinetic model of cabozantinib integrating various cancer types including hepatocellular carcinoma. *The Journal of Clinical Pharmacology* **59**, 1551–1561 (2019).
20. Zwep, L. B. *et al.* Virtual patient simulation using copula modeling. *Clinical Pharmacology & Therapeutics* **115**, 795–804 (2024).

21. Geenens, G. Copula modeling for discrete random vectors. *Dependence Modeling* **8**, 417–440 (2020).
22. Philipp, M. *et al.* Model-based bioequivalence approach for sparse pharmacokinetic bioequivalence studies: Model selection or model averaging? *Statistics in Medicine* (2024).
23. Guhl, M. *et al.* Impact of model misspecification on model-based tests in PK studies with parallel design: real case and simulation studies. *Journal of Pharmacokinetics and Pharmacodynamics* **49**, 557–577 (2022).
24. *BEGINNER'S TUTORIAL Case Study 1* <https://ctm.umaryland.edu/#/ms-pharma/model/bgt/cs>.
25. Nagler, T. & Vatter, T. rvinecopulib: High performance algorithms for vine copula modeling. *R package version 0.5 5*. <https://CRAN.R-project.org/package=rvinecopulib> (2021).
26. *Monolix 2023R1*, Lixoft SAS a Simulations Plus company.
27. *Simulx 2023R1*, Lixoft SAS a Simulations Plus company.
28. Jonsson, E. N. & Nyberg, J. Using forest plots to interpret covariate effects in pharmacometric models. *CPT: Pharmacometrics & Systems Pharmacology* (2024).
29. Guo, Y., Guo, T., Knibbe, C. A., Zwep, L. B. & van Hasselt, J. Generation of realistic virtual adult populations using a model-based copula approach. *Journal of Pharmacokinetics and Pharmacodynamics*, 1–12 (2024).

A Technical appendix

A.1 Power of Test

A.1.1 Power computation for significance test

The null hypothesis for the Wald test of comparison is $H_0 : \beta = 0$ and the alternative hypothesis is $H_1 : \beta \neq 0$. The test statistics is given by equation (23).

$$W = \frac{\hat{\beta}}{SE(\hat{\beta})} \sim \mathcal{N}\left(\frac{\beta}{SE_{\beta}}, 1\right) \quad (23)$$

- Under H_0 , $W \sim \mathcal{N}(0, 1)$, therefore H_0 is rejected if $|W| \geq q_{1-\alpha/2}$, where $q_{1-\alpha/2}$ denotes the $1 - \alpha/2$ -quantile of the normal distribution
- Under H_1 , $\beta \neq 0$, $W \sim \mathcal{N}\left(\frac{\beta}{SE_{\beta}}, 1\right)$.

The power is computed for a given β as the probability to reject H_0 under $\beta \in H_1$:

$$\begin{aligned} P_{comp} &= \mathbb{P}_{H_1}(|W| \geq q_{1-\alpha/2}) = \mathbb{P}_{H_1}(W \geq q_{1-\alpha/2}) + \mathbb{P}_{H_1}(W \leq -q_{1-\alpha/2}) \\ &= \mathbb{P}\left(X + \frac{\beta}{SE_{\beta}} \geq q_{1-\alpha/2}\right) + \mathbb{P}\left(X + \frac{\beta}{SE_{\beta}} \leq -q_{1-\alpha/2}\right) \quad \text{with } X \sim \mathcal{N}(0, 1) \\ &= \mathbb{P}\left(X \geq q_{1-\alpha/2} - \frac{\beta}{SE_{\beta}}\right) + \mathbb{P}\left(X \leq -q_{1-\alpha/2} - \frac{\beta}{SE_{\beta}}\right) \\ &= 1 - \Phi\left(q_{1-\alpha/2} - \frac{\beta}{SE_{\beta}}\right) + \Phi\left(-q_{1-\alpha/2} - \frac{\beta}{SE_{\beta}}\right) \end{aligned} \quad (24)$$

- $\beta > 0$: Because $\mathbb{P}(X \leq -q_{1-\alpha/2}) = \alpha/2$ and $\frac{\beta}{SE_{\beta}} > 0$, we have that $\mathbb{P}_{H_1}(W \leq -q_{1-\alpha/2}) < \alpha/2$. Therefore, the second term can be neglected and

$$\begin{aligned} P_{comp} &\approx \mathbb{P}_{H_1}(W \geq q_{1-\alpha/2}) \\ &\approx 1 - \Phi\left(q_{1-\alpha/2} - \frac{\beta}{SE_{\beta}}\right) \end{aligned} \quad (25)$$

- $\beta < 0$: On the contrary,

$$\begin{aligned} P_{comp} &\approx \mathbb{P}_{H_1}(W \leq -q_{1-\alpha/2}) \\ &\approx \Phi\left(-q_{1-\alpha/2} - \frac{\beta}{SE_{\beta}}\right) \end{aligned} \quad (26)$$

A.1.2 Power computation for Relevance test

The null hypothesis is H_0 : "the covariate effect is not relevant", i.e. $r_{l,c(PX)} \in [R_{inf}; R_{sup}]$ while the alternative hypothesis is H_1 : "the covariate is relevant", i.e. $r_{l,c(PX)} \notin [R_{inf}; R_{sup}]$. This two sided null hypothesis can be split into two, respectively $H_{0,inf}$ and $H_{0,sup}$:

$$\begin{cases} H_{0,inf} : r_{l,c(PX)} \geq R_{inf} & \text{and} & H_{0,sup} : r_{l,c(PX)} \leq R_{sup} \\ H_1 : r_{l,c(PX)} < R_{inf} & \text{or} & r_{l,c(PX)} > R_{sup} \end{cases}$$

Thus H_0 is not rejected unless neither $H_{0,inf}$ nor $H_{0,sup}$ is rejected.

For $r_{l,c(PX)} = e^{\beta_{l,c}(z - z_{ref})}$, the hypothesis write: $H_0 : \beta_{l,c} \in [B_{inf}; B_{sup}]$, where depending on the sign of $(z - z_{ref})$, B_{sup} and B_{inf} equals $\frac{\log R_{sup}}{z - z_{ref}}$ or $\frac{\log R_{inf}}{z - z_{ref}}$.

The null hypothesis is rejected if $H_{0,inf}$ is rejected, i.e. $\widehat{\beta}_{l,c} - B_{inf} < s_1$, or if $H_{0,sup}$ is rejected, i.e. $\widehat{\beta}_{l,c} - B_{sup} > s_2$; where s_1 and s_2 are chosen to reach an α type one error. Because these two events are incompatible the following holds:

$$\begin{cases} \min_{\beta \in H_0} \mathbb{P}\left(\widehat{\beta}_{l,c} - B_{inf} < s_1\right) = \alpha \\ \min_{\beta \in H_0} \mathbb{P}\left(\widehat{\beta}_{l,c} - B_{sup} > s_2\right) = \alpha \end{cases} \quad (27)$$

Recalling that $\frac{\widehat{\beta}_{l,c} - \beta_{l,c}}{SE_{\beta_{l,c}}} \sim \mathcal{N}(0, 1)$, the system writes:

$$\left\{ \begin{array}{l} \min_{\beta \in H_0} \mathbb{P}(\beta_{l,c} - B_{inf} < s_1) = \min_{\beta \in H_0} \Phi\left(\frac{s_1 + B_{inf} - \beta_{l,c}}{SE_{\beta_{l,c}}}\right) = \alpha \\ \min_{\beta \in H_0} \mathbb{P}(\beta_{l,c} - B_{sup} > s_2) = \min_{\beta \in H_0} 1 - \Phi\left(\frac{s_2 + B_{sup} - \beta_{l,c}}{SE_{\beta_{l,c}}}\right) = \alpha \end{array} \right. \implies \left\{ \begin{array}{l} s_1 = q_{\alpha} SE_{\beta_{l,c}} = -q_{1-\alpha} SE_{\beta_{l,c}} \\ s_2 = q_{1-\alpha} SE_{\beta_{l,c}} \end{array} \right. \quad (28)$$

Finally, at level $1 - \alpha$, the null hypothesis is rejected if $\widehat{\beta}_{l,c} - B_{inf} < -q_{1-\alpha} SE_{\beta_{l,c}}$ or if $\widehat{\beta}_{l,c} - B_{sup} > q_{1-\alpha} SE_{\beta_{l,c}}$; equivalently, the null hypothesis is rejected if $\widehat{\beta}_{l,c} + q_{1-\alpha} SE_{\beta_{l,c}} < B_{inf}$ or if $\widehat{\beta}_{l,c} - q_{1-\alpha} SE_{\beta_{l,c}} > B_{sup}$. The first inequality involves the upper bound of the $1 - 2\alpha$ CI on the ratio, while the second inequality involves the lower bound.

The power is the probability under H_1 to reject H_0 :

$$P_{relev} = \Phi\left(-q_{1-\alpha} + \frac{B_{inf} - \beta_{l,c}}{SE_{\beta_{l,c}}}\right) + 1 - \Phi\left(q_{1-\alpha} + \frac{B_{sup} - \beta_{l,c}}{SE_{\beta_{l,c}}}\right) \quad (29)$$

On one hand, if $B_{inf} < \beta_{l,c}$, $\Phi\left(-q_{1-\alpha} + \frac{B_{inf} - \beta_{l,c}}{SE_{\beta_{l,c}}}\right) < \Phi(-q_{1-\alpha}) = \alpha$, therefore this term is negligible in power computation. On the other hand, if $B_{sup} > \beta_{l,c}$, $1 - \Phi\left(q_{1-\alpha} + \frac{B_{sup} - \beta_{l,c}}{SE_{\beta_{l,c}}}\right) < 1 - \Phi(q_{1-\alpha}) = \alpha$, therefore this term is negligible in power computation. Consequently, to compute the number of subjects needed to achieve, a given power let us distinguish two cases:

- If $B_{inf} < \beta_{l,c}$: $P_{relev} \approx 1 - \Phi\left(q_{1-\alpha} + \frac{B_{sup} - \beta_{l,c}}{SE_{\beta_{l,c}}}\right) \implies SE_{\beta_{l,c}} = \frac{B_{sup} - \beta_{l,c}}{\Phi^{-1}(1 - P_{relev}) - q_{1-\alpha}}$
- If $B_{sup} > \beta_{l,c}$: $P_{relev} \approx \Phi\left(-q_{1-\alpha} + \frac{B_{inf} - \beta_{l,c}}{SE_{\beta_{l,c}}}\right) \implies SE_{\beta_{l,c}} = \frac{B_{inf} - \beta_{l,c}}{\Phi^{-1}(P_{relev}) + q_{1-\alpha}}$

B Addition details on the evaluation by simulation

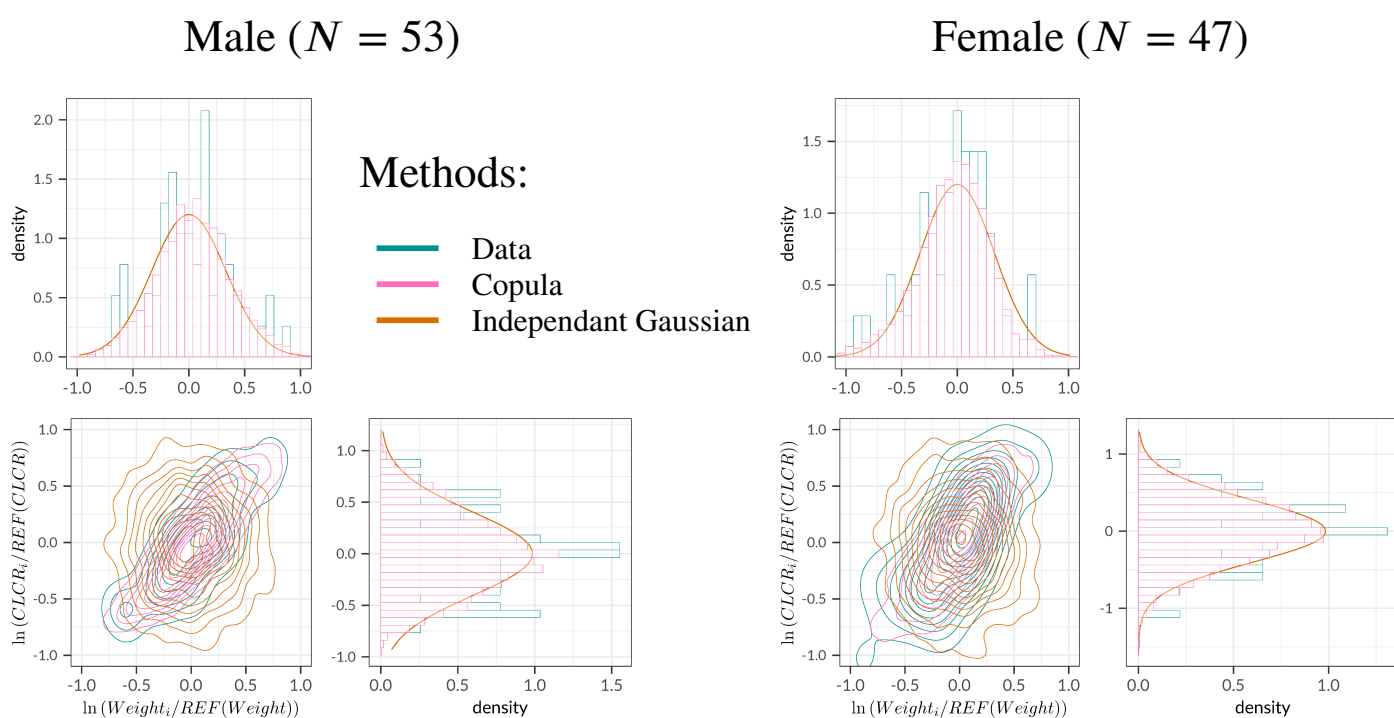


Figure 4: Evaluation - Covariate distributions for Male on the left and Female on the right: Data boxplots, fitted Copulas boxplots and fitted independent Gaussian densities.

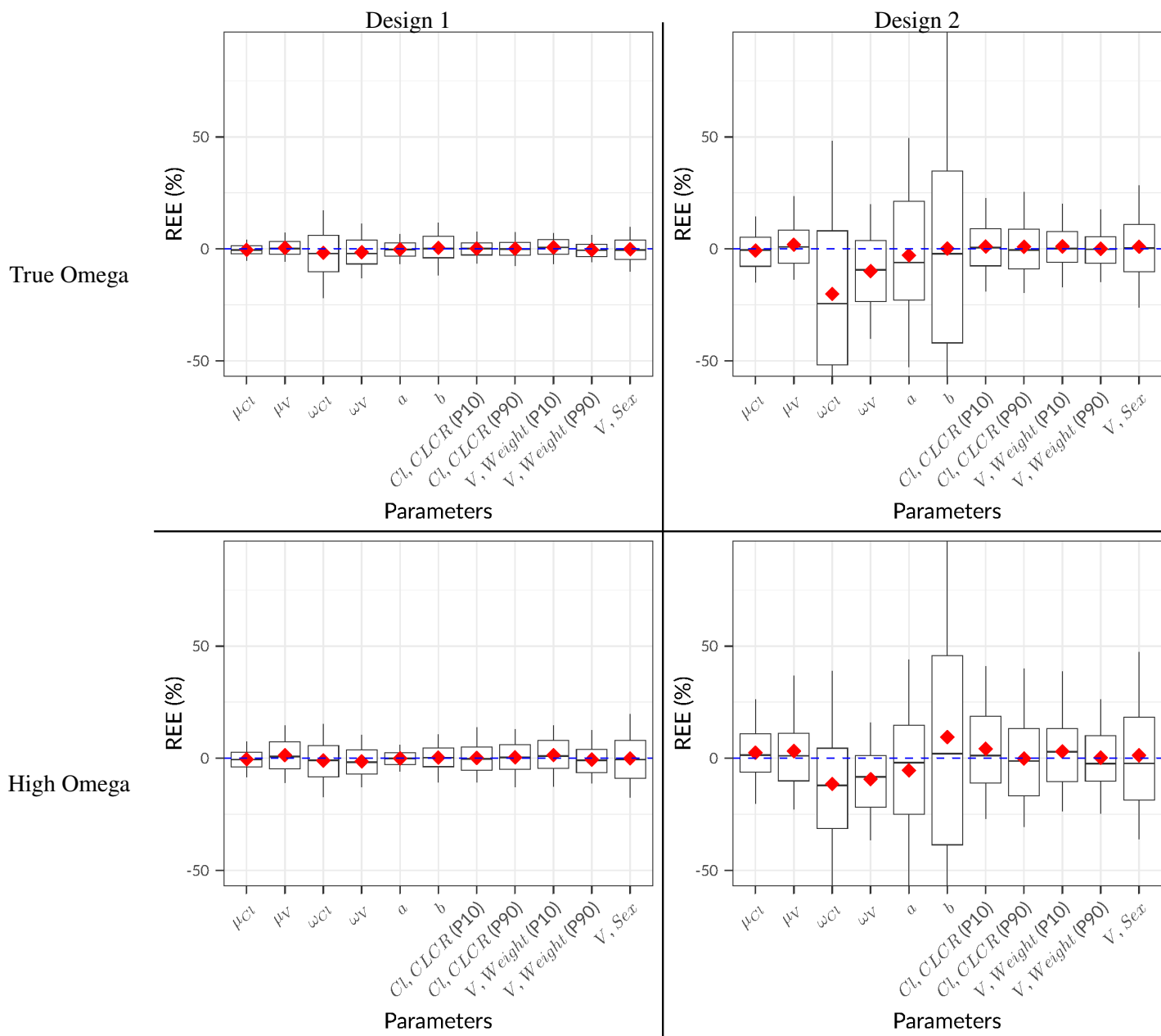


Figure 5: Evaluation - Boxplots of Relative Estimation Error across 200 datasets for the 4 scenarios

The boxplot displays the median, the 25th and 75th percentiles, while the whiskers are 5th and 95th percentiles. The red diamond corresponds to the mean.

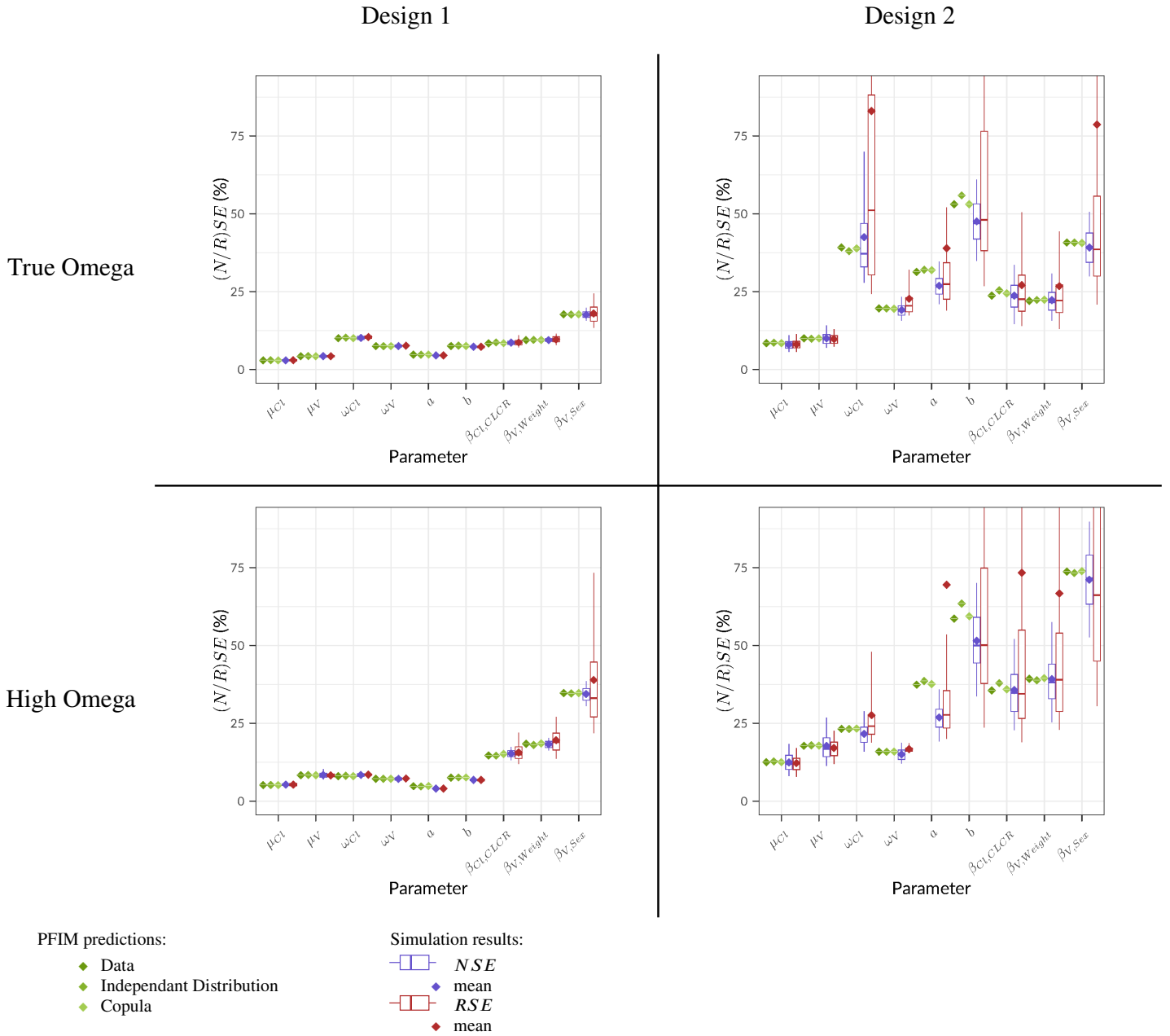


Figure 6: Evaluation - Normalized and Relative Standard Error for the 4 scenarios: PFIM predictions using the three methods for handling covariates and simulation results across 200 datasets

The boxplot displays the median, the 25th and 75th percentiles, while the whiskers are 5th and 95th percentiles.

C Addition details on PK example simulation

Table 8: Application - Theoretical Sampling Design

Study	N	Daily Doses, mg (N)	n	Samplings, hours after first dose
XL184-001	40	140 (35) - 200 (5)	15	0.5, 1, 2, 4, 8, 96, 100, 336, 432, 432.5, 433, 434, 436, 440, 672
XL184-010	77	140 (77)	43	0.5, 1, 2, 3, 4, 5, 6, 8, 10, 12, 14, 24, 48, 72, 120, 168, 240, 288, 336, 408, 504, 744, 744.5, 745, 746, 747, 748, 749, 750, 752, 754, 756, 758, 768, 792, 816, 864, 912, 984, 1032, 1080, 1152, 1248
XL184-020	63	20 (21) - 40 (21) - 60 (21)	21	0.5, 1, 2, 3, 4, 5, 6, 8, 10, 12, 14, 24, 48, 72, 120, 168, 240, 288, 336, 408, 504
XL184-201	39	140 (39)	7	336, 340, 672, 676, 1008, 1012, 1344
XL184-203 - NRE	136	40 (55) - 60 (36) - 100 (45)	5	504, 1008, 2016, 3024, 4032
XL184-203 - RDT	185	60 (35) - 100 (150)	5	2016, 2352, 2688, 3024, 3360
XL184-301	210	100 (1) - 140 (209)	7	2, 4, 6, 672, 674, 676, 678
XL184-306	41	20 (1) - 40 (1) - 60 (39)	3	504, 1008, 2016
XL184-307	498	20 (18) - 40 (41) - 60 (439)	2	504, 2016
XL184-308	282	20 (5) - 40 (33) - 60 (244)	2	672, 1344
XL184-309	452	40 (7) - 60 (445)	3	336, 672, 1344

N : refers to the number of subjects included in the study, n : the number of PK samplings per subject

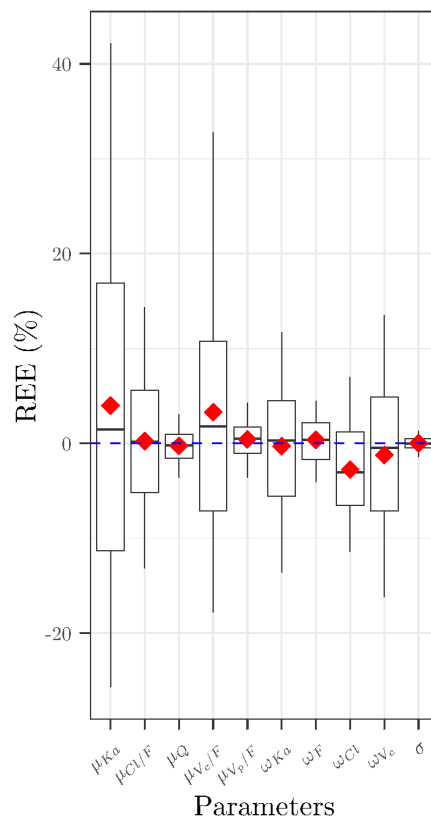


Figure 7: Application - Boxplots of Relative Estimation Error across 200 datasets - Base parameters

The boxplot displays the median, the 25th and 75th percentiles, while the whiskers are 5th and 95th percentiles. The red diamond corresponds to the mean.

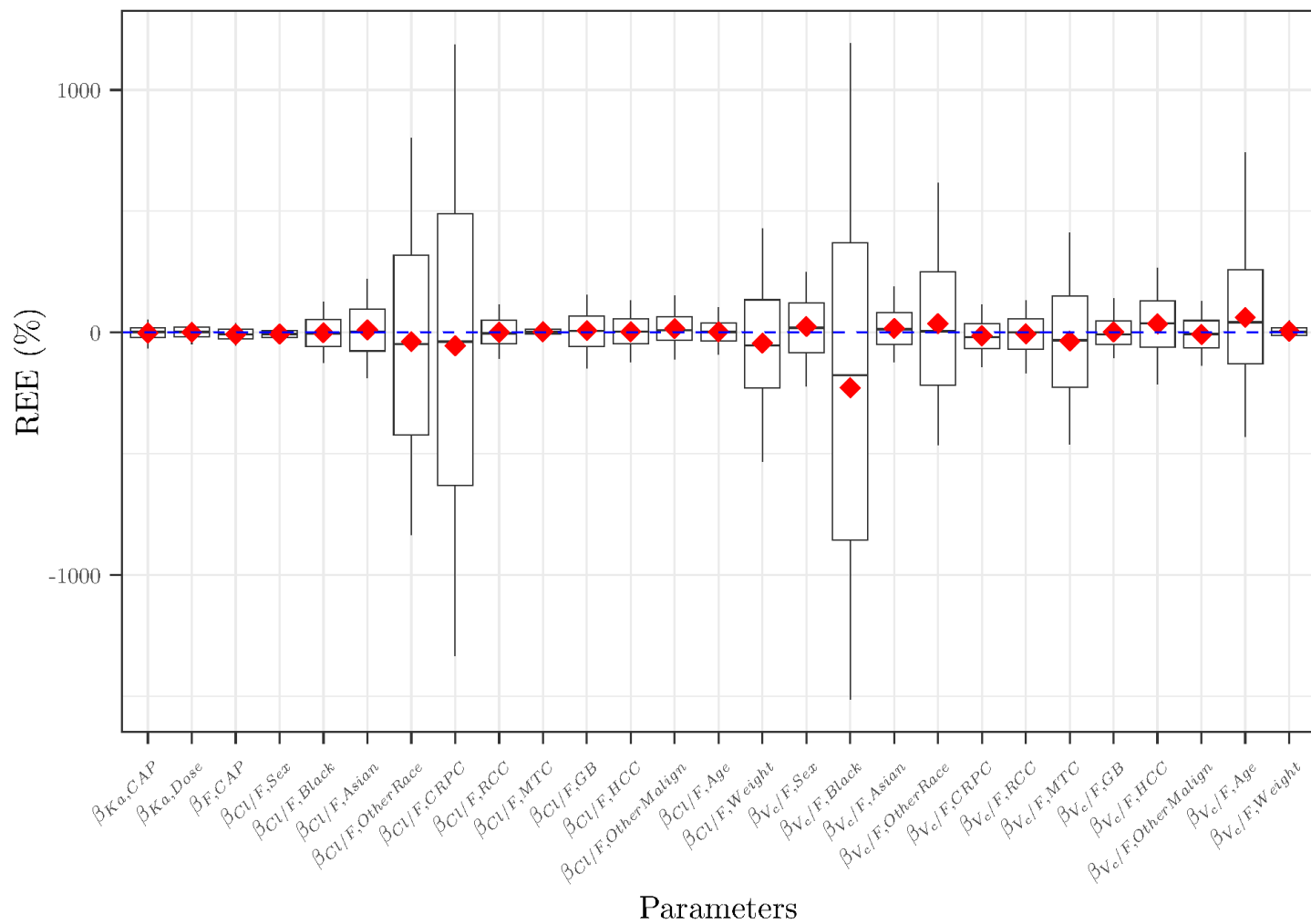


Figure 8: Application - Boxplots of Relative Estimation Error across 200 datasets - Covariate effects parameters

The boxplot displays the median, the 25th and 75th percentiles, while the whiskers are 5th and 95th percentiles. The red diamond corresponds to the mean.

Table 9: Application - *RB* and *RRMSE* across 200 datasets - Base parameters

Parameter	RB (%)	RRMSE (%)
μ_{K_a}	3.97	20.9
$\mu_{Cl/F}$	0.237	7.90
μ_Q	-0.282	2.02
$\mu_{V_c/F}$	3.27	16.1
$\mu_{V_p/F}$	0.418	2.37
ω_{K_a}	-0.305	7.85
ω_F	0.36	2.58
ω_{Cl}	-2.77	6.47
ω_{V_c}	-1.24	9.15
σ	0.00625	0.797

Table 10: Application - Simulation Value, Estimate, *RB* and *RRMSE* across 200 datasets - Covariate effects parameters

Parameter	Value	Estimate	RB (%)	RRMSE (%)
$\beta_{K_a,CAP}$	-0.911	-0.892	-2.11	33.9
$\beta_{K_a,Dose}$	0.734	0.738	0.578	29.1
$\beta_{F,CAP}$	-0.166	-0.153	-7.74	30.8
$\beta_{Cl/F,Sex}$	-0.274	-0.257	-6.17	23.6
$\beta_{Cl/F,Black}$	0.162	0.16	-1.08	83.0
$\beta_{Cl/F,Asian}$	-0.0668	-0.0741	10.9	127
$\beta_{Cl/F,OtherRace}$	0.0279	0.0174	-37.8	537
$\beta_{Cl/F,CRPC}$	-0.0115	-0.00522	-54.6	808
$\beta_{Cl/F,RCC}$	-0.139	-0.14	0.361	67.0
$\beta_{Cl/F,MTC}$	0.643	0.661	2.80	16.2
$\beta_{Cl/F,GB}$	0.178	0.192	7.73	96.7
$\beta_{Cl/F,HCC}$	-0.13	-0.134	3.38	76.0
$\beta_{Cl/F,OtherMalign}$	0.171	0.199	16.3	83.4
$\beta_{Cl/F,Age}$	-0.157	-0.161	2.41	61.0
$\beta_{Cl/F,Weight}$	-0.0393	-0.0221	-43.8	285
$\beta_{V_c/F,Sex}$	0.0939	0.117	24.4	148
$\beta_{V_c/F,Black}$	0.0441	-0.0564	-228	930
$\beta_{V_c/F,Asian}$	-0.363	-0.423	16.5	95.4
$\beta_{V_c/F,OtherRace}$	-0.126	-0.172	36.5	362
$\beta_{V_c/F,CRPC}$	-0.297	-0.256	-13.9	82.2
$\beta_{V_c/F,RCC}$	-0.422	-0.397	-5.88	91.4
$\beta_{V_c/F,MTC}$	-0.0657	-0.0425	-35.3	278
$\beta_{V_c/F,GB}$	-0.735	-0.75	2.09	76.0
$\beta_{V_c/F,HCC}$	-0.166	-0.225	35.7	158
$\beta_{V_c/F,OtherMalign}$	-0.272	-0.248	-8.8	85.5
$\beta_{V_c/F,Age}$	0.0644	0.104	62.1	347
$\beta_{V_c/F,Weight}$	1.19	1.25	4.74	25.6

CRPC: Castration-resistant prostate cancer, RCC: Renal cell carcinoma, MTC: Metastatic medullary thyroid cancer, GB: Glioblastoma multiforme, HCC: Hepatocellular carcinoma, OtherMalign: other malignancy, CAP: Capsule formulation

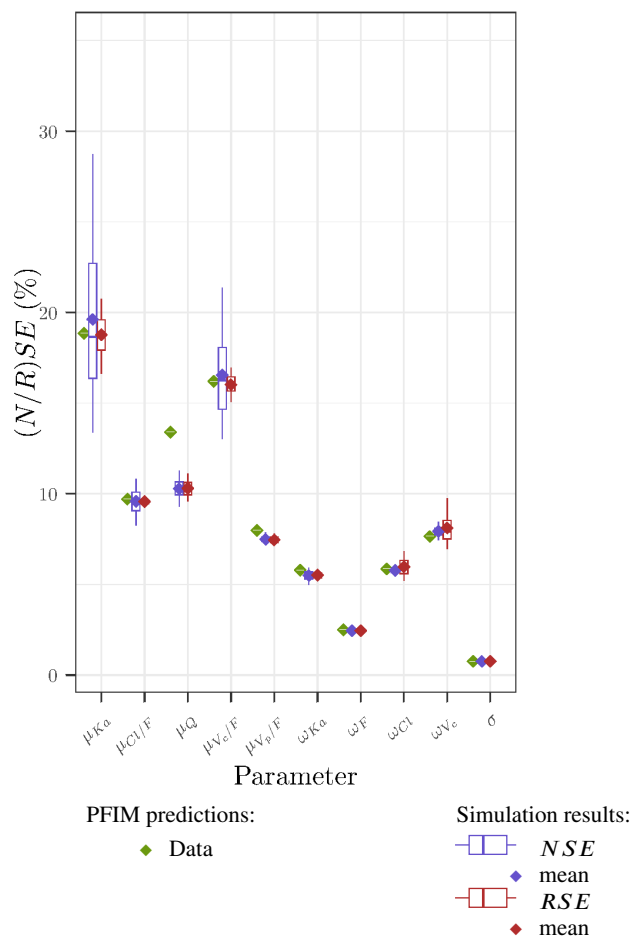


Figure 9: Application - Normalized and Relative Standard Error - Base parameters: PFIM predictions using the Data method for handling covariates and simulation results across 200 datasets

The boxplot displays the median, the 25th and 75th percentiles, while the whiskers are 5th and 95th percentiles.

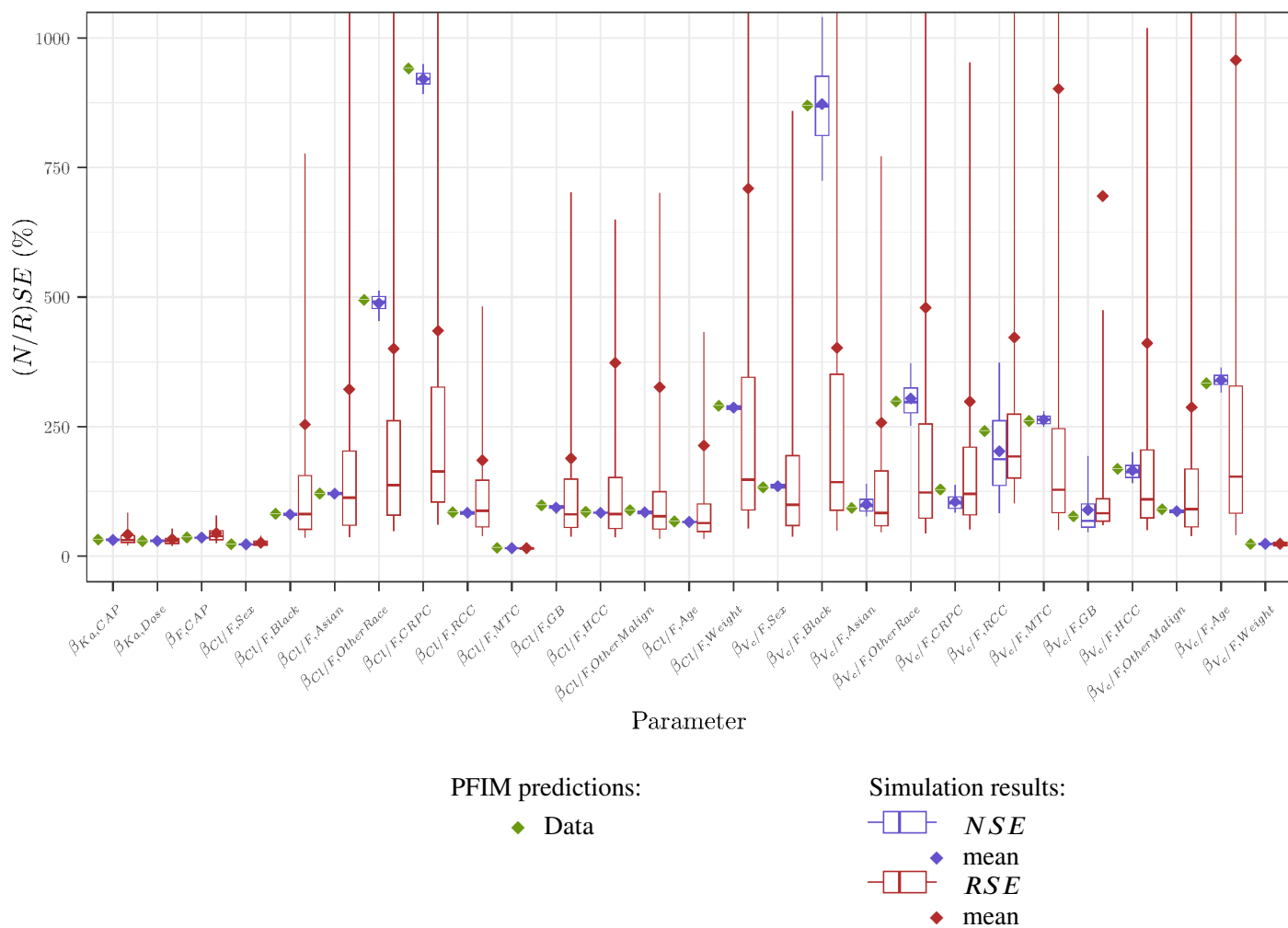


Figure 10: Application - Normalized and Relative Standard Error - Covariate effects: PFIM predictions using the Data method for handling covariates and simulation results across 200 datasets

The boxplot displays the median, the 25th and 75th percentiles, while the whiskers are 5th and 95th percentiles.