HAL
open science

Methods and software tools for design evaluation for population pharmacokinetics-pharmacodynamics studies.

Joakim Nyberg, Caroline Bazzoli, Kay Ogungbenro, Alexander Aliev, Sergei Leonov, Stephen Duffull, Andrew C. Hooker, France Mentré

To cite this version:

Joakim Nyberg, Caroline Bazzoli, Kay Ogungbenro, Alexander Aliev, Sergei Leonov, et al.. Methods and software tools for design evaluation for population pharmacokinetics-pharmacodynamics studies.: Softwares for design evaluation in population PKPD. British Journal of Clinical Pharmacology, 2015, 79 (1), pp.6-17. 10.1111/bcp.12352 . inserm-00978789

HAL Id: inserm-00978789

https://inserm.hal.science/inserm-00978789

Submitted on 14 Apr 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Methods and software tools for design evaluation for population pharmacokinetics-pharmacodynamics studies

Authors

Joakim Nyberg (1), Caroline Bazzoli (2), Kay Ogungbenro (3), Alexander Aliev (4), Sergei Leonov (5), Stephen Duffull (6), Andrew C. Hooker (1), France Mentré (7)

(1) Department of Pharmaceutical Biosciences, Uppsala University, Uppsala, Sweden;

(2) Laboratoire Jean Kuntzmann, Département Statistique, University of Grenoble, France ;

(3) Centre for Applied Pharmacokinetic Research, School of Pharmacy and Pharmaceutical Sciences, University of Manchester, Manchester, United Kingdom;

(4) Institute for Systems Analysis, Russian Academy of Sciences, Moscow, Russia;

(5) AstraZeneca, Wilmington, DE, USA;

(6) School of Pharmacy, University of Otago, Dunedin, New Zealand;

(7) INSERM U738 and University Paris Diderot, Paris, France;

Submitting Author and Corresponding Author

France Mentré, france.mentre@inserm.fr

Running head: Softwares for design evaluation in population PKPD

Keywords: optimal design, population design, Fisher information matrix, nonlinear mixed effect models, population PKPD, PFIM, PkSaMp, PopDes, POpED, POPT.

Word count*:* 3800

Numbers of tables and figures: 5 tables and 2 figures

Summary

Population Pharmacokinetic (PK)-Pharmacodynamic (PD) (PKPD) models are increasingly used in drug development and in academic research. Hence designing efficient studies is an important task. Following the first theoretical work on optimal design for nonlinear mixed effect models, this research theme has grown rapidly. There are now several different software tools that implement an evaluation of the Fisher information matrix for population PKPD. We compared and evaluated five software tools: PFIM, PkStaMP, PopDes, PopED, and POPT. The comparisons were performed using two models: i) a simple one compartment warfarin PK model; ii) a more complex PKPD model for Pegylated-interferon (peg-interferon) with both concentration and response of viral load of hepatitis C virus (HCV) data. The results of the software were compared in terms of the standard error values of the parameters (SE) predicted from the software and the empirical SE values obtained via replicated clinical trial simulation and estimation. For the warfarin PK model and the peg-interferon PKPD model all software gave similar results. Of interest it was seen, for all software, that the simpler approximation to the Fisher information matrix, using the block diagonal matrix, provided predicted SE values that were closer to the empirical SE values than when the more complicated approximation was used (the full matrix). For most PKPD models, using any of the available software tools will provide meaningful results, avoiding cumbersome simulation and allowing design optimisation.

Introduction

Estimation of pharmacokinetic (PK) parameters for an individual using nonlinear regression techniques started in the 1960's, followed by estimation of dose-response and of pharmacodynamics (PD) models. At around the same time mathematical approaches to defining the problem of optimal design for parameter estimation in nonlinear regression was addressed (1-3). However this did not reach the PK literature until some 20 years later (4). The problem was not only to draw inference from data but also to define the best design(s) for estimation of parameters using maximum likelihood or other estimation methods. For this purpose, the Fisher Information matrix (FIM) was used to describe the informativeness of a design, i.e. how much information the design has in relation to parameter estimation. Typically in PK the FIM is summarized by its determinant and maximising the determinant, termed D-optimality, is equivalent to minimising the asymptotic confidence region of the parameters, i.e. getting the most precise parameter estimates (5-9). However, beyond theoretical developments, a limitation of individualised optimised designs of PKPD studies is that those designs do not acknowledge population information and hence cannot have fewer sampling times per individual than parameters to estimate. In addition, optimal designs with a large number of observations per patient will have replicated optimal sampling times; which were not favoured by pharmacologists interested in exploring complex PK models. Some later work also explored Bayesian designs, where *a priori* distributions of the parameters were considered, and individual parameters were estimated using maximum *a posteriori* probability (MAP). Optimal designs for MAP estimation optimise individual designs given prior population information and are suitable for e.g. therapeutic drug monitoring designs (10, 11). Since 1985, the software Adapt (https://bmsr.usc.edu/software/adapt/) has included methods for optimal design in nonlinear regression using several criteria for MAP estimation.

The population approach was introduced by Sheiner et al. (12) for PK analyses in the late 1970's and since the 1980's there has been a large increase in the use of this approach as well as extensions to PKPD. Estimation was mainly based on maximum likelihood using nonlinear mixed effects models (NLMEM) thanks to the software NONMEM. To our knowledge the first article studying the impact of a 'population design' on properties of estimates was performed in early 1990's by Al Banna et al. (13) for a population PK and a population PKPD example. In this work the author used clinical trial simulation (CTS) to explore possible designs. The authors studied the influence of the balance of number of patients, number of sampling times and locations of the sampling times on the precision of the parameter estimates. Several papers, all using CTS, were published (14-16) showing that some designs could be rather poor, and that very sparse designs also performed poorly. The FDA's

Guidance for Industry Population Pharmacokinetics (17) from 1999 includes a specific section on design, and suggests that simulation, based on preliminary information, should be performed to "anticipate certain fatal study designs, and to recognize informative ones".

Using CTS for design evaluation requires a large number of data sets to be simulated and then fitted under each proposed design which is computationally expensive. However, since CTS is a user driven, heuristic approach, then it can miss important regions of the design space because only a fixed number of designs are investigated. Subsequently it was suggested to use the FIM in NLMEM to predict asymptotic standard errors (SE) and define optimal designs without the need for intensive simulations. Because the population likelihood has no closed-form expression the proposed approach for defining the population FIM was to use a first-order linearisation of the model around the random effects (which is the same as used for the first-order (FO) estimation methods). This approximation results in a mixed effect model where the random effects enter the model linearly (rather than nonlinearly) and hence has properties that are similar to linear mixed effects model. The expression for the population FIM was first published in Biometrika in 1997 (18). In this work the FIM was derived for a population PK example and an algorithm was proposed to optimise designs based on the population FIM. This paper launched the new field of optimal design for nonlinear mixed effects models. It has been quoted in the section 'other influential papers of the 1990's' in a review in Biometrika (19).

Since 1997 several methodological papers from various academic teams have published different extensions, for instance robust designs, sampling windows, compound designs, multiple response models, methods for discrete longitudinal data, and other approximations of the FIM, etc. Most importantly, the derivation of the expression of the FIM was implemented in several software tools, the first one PFIM (20) in 2001 appeared simultaneously in both R (http://www.r-project.org/) and Matlab (http://www.mathworks.fr/products/matlab/). This was followed by POPT (21), and later to incorporate an interface version WinPOPT, PopED (22), PopDes (23) and PkStaMp (24). There are now five different software tools, all implementing the first-order approximation, with some tools implementing one or several other approximations. These tools for designing population PKPD studies are gaining popularity. In a recent study performed among European Federation of Pharmaceutical Industries and Associations members' (25), it was found that 9 out of 10 pharmaceutical companies are using one of these software tools for design evaluation or optimisation, mainly in phases I and II.

The computation of the FIM is complex and depends on the numerical implementation. The purpose of the present work was therefore to compare the results provided by those different software tools in terms of FIM and predicted SE values. The same basic approximations were used in each software, and the comparison was performed for two examples: (1) a simple PK example described by a one-compartment model with first-order absorption and linear elimination and (2) a more complex PKPD example where the PD component is defined by a system of nonlinear ordinary differential equations (ODE). The objective was to explore the results from different software tools and to compare results against those obtained using CTS. We wanted to show the user community that similar results would be obtained with any software tool although programmed in different languages and by different authors. This was also studied in the case of a multiple responses ODE model where the numerical imbrication between ODE solver and numerical differentiation is complex. The results were provided by the software developers, all authors of this article, who were given the equations of the models, the values of the parameters and the designs to be evaluated. Results were compared to those obtained by CTS.

The article is organized as follows: first the description of the population FIM for NLMEM, second a description of the various software tools, and then an evaluation of the two examples. As no design optimisation was performed in the present study, no optimisation characteristics or algorithms are described.

Statistical methods for design in NLMEM

A design for a multi-response NLMEM is composed of N subjects each with an associated elementary design ξ_i ($i = 1,...,N$). Hence a design for a population of N subjects can be described as

$$\Xi = \left(\xi_1,...,\xi_N\right) \tag{1}$$

Each elementary design ξ_i can be further divided into sub-designs

$$\xi_i = \left(\xi_{i1},\ldots,\xi_{iK}\right) \tag{2}$$

with ξ_{ik}, $k = 1,...,K$ being the design associated with the kth response. (e.g. drug concentration, metabolite concentration, effect). It may thus be possible to have all responses measured at different times, termed an unbalanced design.

A design for subject *i* at a response $k = 1,...,K$ often consists of several design variables which might be constant between observations, e.g. the drug dose, or vary between observations, e.g. the times at which the response variable is measured.

An elementary design ξ_l can be the same within a group l of N_l subjects ($l = 1,\ldots,L$). Using a similar notation for the complete population design Ξ in a limited number of L groups of different elementary designs gives:

$$\Xi = \left(\left[\xi_1, N_1\right],...,\left[\xi_L, N_L\right]\right) \tag{3}$$

where the total number of subjects in the design, N, is equal to the sum of the subjects in the L elementary designs. At the extreme, each subject may have a different design, $L = N$, or each subject may have the same design, $L = 1$.

In a NLMEM framework with multiple response, the vector of observations Y_i for the ith subject is defined as the vector of K different responses:

$$Y_i = \left[y_{i1}^T, y_{i1}^T, \ldots, y_{iK}^T \right]^T \tag{4}$$

where y_{ik}, $k = 1, \ldots, K$ is the vector of n_{ik} observations for subject *i* and response *k* modelled as

$$y_{ik} = \mathrm{f}_k\left(\theta_i, \xi_{ik}\right) + \mathrm{h}_k\left(\theta_i, \xi_{ik}, \varepsilon_{ik}\right) \tag{5}$$

where $f_k(.)$ is the structural model for the kth response, θ_i is the ith subject's parameter vector, $h_k(.)$ is the residual error model for response *k*, often additive (h=ε_{ik}), proportional (h=$\mathrm{f}_k\left(\cdot\right)\varepsilon_{ik}$) or a combination of both, ε_{ik} is the residual error vector for response *k* in subject *i*. In this paper additive (homoscedastic) or proportional (heteroscedastic) error models will be used in the examples so that only one residual variance parameter is defined for each response. To simplify notation we assume that ε_{ik} are normally distributed and independent between responses (which is not necessary, see e.g. (26, 27)) with mean zero and variance Σ_k=diag(σ_k^2). The individual parameter vector θ_i, with parameter(s) that might be shared between responses, is described as

$$\theta_i = \mathrm{g}\left(\beta, b_i\right) \tag{6}$$

where β is the *u*-vector of fixed effects parameters, or typical subject parameter and b_i, the vector of the v random effects for the subject *i* defining the subject deviation from the typical value of the parameter. We assume that b_i is normally distributed with a mean of zero and a covariance matrix Ω of size $v \times v$. Again, to simplify notation we assume a diagonal (which is not necessary, see e.g. (18, 27-29)) interindividual covariance matrix (Ω) with diagonal elements ($\omega_1^2,...,\omega_v^2$). The vector of population parameters is thus defined as

$$\psi = [\beta, \lambda] = \left[\beta, \omega_1^2,...,\omega_v^2, \sigma_1^2,...,\sigma_K^2\right] \tag{7}$$

where $\lambda = \left[\omega_1^2,...,\omega_v^2, \sigma_1^2,...,\sigma_K^2\right]$ is the vector of all variance components.

The population Fisher information matrix $FIM(\psi, \Xi)$ for multiple response models with the population design Ξ is given by:

$$FIM(\psi, \Xi) = E\left(-\frac{\partial^2 L(\psi, Y)}{\partial \psi \partial \psi^T}\right) \tag{8}$$

where $L(\psi, Y)$ is the log-likelihood of all the observations *Y* given the population parameters ψ.

Assuming independence across subjects, the log-likelihood can be defined as the sum of the individual contribution to the log-likelihood: $L(\psi, Y) = \sum_{i=1}^{N} L(\psi, Y_i)$. Therefore, the population Fisher information matrix (calculated using the second derivative of the log-likelihood) for *N* subjects can also be defined as the sum of the *N* elementary information matrices $FIM(\psi, \xi_i)$ computed for each subject *i:*

$$FIM(\psi, \Xi) = \sum_{i=1}^{N} FIM(\psi, \xi_i) \tag{9}$$

In the case of a limited number of L groups (where each individual in a group share the same design), as in Equation (3), the population FIM is expressed by:

$$FIM\left(\psi,\Xi\right)=\sum_{l=1}^{L}N_l\,FIM\left(\psi,\xi_l\right) \tag{10}$$

For one subject, given the design variables ξ_i and the NLMEM model, the FIM is a block matrix defined as:

$$\mathrm{FIM}\left(\psi,\xi_i\right)=\frac{1}{2}\begin{bmatrix}A & C\\ C & B\end{bmatrix} \tag{11}$$

where $FIM\left(\beta,\xi_i\right)=\frac{1}{2}A$ is the block of the Fisher matrix for the fixed effects β and $FIM\left(\lambda,\xi_i\right)=\frac{1}{2}B$ is the block of the Fisher matrix for the variance components λ.

When a standard FO approximation of the model is performed (see appendix), then the distribution of the observations in patient *i* with design ξ_i is approximated by $Y_i \sim \mathrm{N}\left(E_i,V_i\right)$. Expressions for the population mean E_i and population variance V_i are given in the appendix. Then the following expression for blocks A, B and C are obtained (18, 30, 31), ignoring indices *i* for simplicity :

$$A_{pq}\cong 2\frac{\partial E^T}{\partial\beta_p}V^{-1}\frac{\partial E}{\partial\beta_q}+\mathrm{tr}\left(\frac{\partial V}{\partial\beta_p}V^{-1}\frac{\partial V}{\partial\beta_q}V^{-1}\right)\ \text{with}\ p,q=1..\dim\left(\beta\right)$$

$$B_{pq}\cong\frac{\partial V}{\partial\lambda_p}V^{-1}\frac{\partial V}{\partial\lambda_q}V^{-1}\quad\text{with}\ p,q=1..\dim\left(\lambda\right) \tag{12}$$

$$C_{pq}\cong\mathrm{tr}\left(\frac{\partial V}{\partial\lambda_p}V^{-1}\frac{\partial V}{\partial\beta_q}V^{-1}\right)\ \text{with}\ p=1..\dim\left(\lambda\right),q=1..\dim\left(\beta\right)$$

This expression of the FIM (eq. 12) will be referred to as the full FIM in this paper.

If the approximated variance V is assumed independent of the typical population parameters β, the matrix C will be zero and the matrices A and B will instead be defined as:

$$A_{pq}\cong 2\frac{\partial E^T}{\partial\beta_p}V^{-1}\frac{\partial E}{\partial\beta_q}\quad\text{with}\ p,q=1..\dim\left(\beta\right) \tag{13}$$

$$B_{pq} \cong \frac{\partial V}{\partial \lambda_p} V^{-1} \frac{\partial V}{\partial \lambda_q} V^{-1} \quad \text{with } p,q = 1..\dim(\lambda)$$

which will be termed the block diagonal FIM in the following. The explicit formula for $FIM(\beta, \xi_i)$ using the block diagonal form is given in the appendix. More information about the derivation of the FIM or other approximations are reported in (27, 28, 30, 32-34).

Software description

There are presently five software tools that implement experimental design evaluation and optimisation of the FIM for multiple response population models. The five software tools are (in alphabetical order) : PFIM (35), PkStaMp (24), PopDes (23), PopED (27, 31) and POPT (21). Four of them have been developed by academic teams.

PFIM (Population Fisher Information Matrix) is the only tool that is using the software R, the other software packages have been developed under the numerical computing environment MATLAB. The first version of PFIM appeared in 2001 and since this date several releases have been issued. It is available at www.pfim.biostat.fr. A graphical user interface (GUI) package using the R software (PFIM Interface) is also available but does not include recent methodological developments. Pk StaMP (Pharmacokinetic Sampling Times Allocation – Matlab Platform) is a library compiled as a single executable file which does not require a MATLAB license. The developers can share the stand-alone version with anyone interested. PopDes (Population Design) has been developed at the University of Manchester and this application software is available at www.capkr.man.ac.uk/home since 2007. PopED (Population optimal Experimental Design), freely available at poped.sf.net, consists of two parts, a script version, responsible for all optimal design calculations, and a GUI. The script version can use either MATLAB or Freemat (http://www.freemat.sf.net)(a free alternative to MATLAB) as an underlying engine. Some advanced PopED features such as automatic and symbolic differentiation, Laplace approximation of Bayesian criteria and mode base linearisation are not available in FreeMat, however all features presented in Table 1 are available in PopED using either FreeMat or Matlab. POPT (Population OPTimal design) was developed from PFIM (MATLAB) in 2001 and is constructed as a set of MATLAB scripts. POPT requires MATLAB and can run on FreeMat. This tool can be downloaded on the website www.winpopt.com. All the software tools run on any common operating system platform (e.g. Windows, Linux, Mac).

Comparison of software for design evaluation

As we focus on design evaluation and not design optimisation, we first compared the software tools with respect to a) required programming language, b) availability, c) library of PK and PD models, and ability to deal with: d) multiple response models, e) models defined by differential equations, e) unbalanced multiple response designs, f) correlations between random effects and/or residuals, g) models including inter-occasion variability, h) models including fixed effects for the influence of discrete covariates on the parameters, i) computation of the predicted power. Table 1 is a summary of the comparison of the software with respect to these different aspects. Globally, for all software tools, the library of PK models includes one, two and three compartment models, with bolus, infusion and first-order (e.g. oral) administration, after a single dose, multiple doses and at steady state. PK models with first-order elimination and models Michaelis-Menten elimination are available. Regarding PD models, immediate linear and Emax models and turnover response models are available.

Over recent years, those tools have included various improvements in terms of model specification and calculations of the FIM. For all of them, design evaluation can be performed for single or multiple response models either using libraries of standard PK and PD models or using a user-defined model. For the latter, regardless of the software used, the model can be written using an analytical form or using a differential equation system. In the case of multiple response models, population designs can be different across the responses for all the software. Regarding the calculations of the information matrix, the majority of the software can handle either a block diagonal Fisher information matrix (block FIM) or the full matrix (full FIM). Otherwise, only PopDes and PopED allow for calculations for a model with both correlation between random effects (full covariance matrix Ω) and correlation between residuals (full covariance matrix Σ), PKStamp allows full covariance matrix Ω. It is possible in PFIM, PopDes and PopED to use models with inter-occasion variability (IOV) and models including fixed effects for the influence of discrete covariates on the parameters. The computation of the predicted power of the Wald test (30, 36) for a given distribution of a discrete covariate can be evaluated in PFIM**,** PopDES and PopED frameworks.

Examples

Two different examples were used to illustrate the performance of the five population design software tools. Note that the examples evaluated the prediction for a given design, by evaluating the FIM and the predicted asymptotic SEs, without design optimisation. This was done to evaluate the core calculations of the FIM. The FIM is evaluated with the full and the block diagonal derivation (eq. 12, 13) with the different software tools.

In the first example a one compartment PK model (based on a warfarin PK model) with first-order absorption was used (35). The design of that study consisted of 32 subjects with a single dose of 70 mg (a dose of 1mg/kg and a weight of 70 kg), and with 8 sampling times post-dose (in hours):

$$\Xi = \left(\left[\xi_1, N_1\right]\right) = \left(\xi_1, 32\right)$$

$$\xi_1 = \left(t_i\right) = \left(0.5, 1, 2, 6, 24, 36, 72, 120\right)$$

The residual error model was proportional (h $= \mathrm{f} \cdot \varepsilon$) with a coefficient of variation of 10% ($\sigma^2 = 0.01$) and exponential random effects were assumed for all parameters ($\mathrm{g} = \beta e^{b_i}$). Table 2 reports the model parameters and their values. The dose and design are based on (34, 37).

For the second example a multiple response PKPD model with repeated dosing was selected with the same design across responses (38). The model describes hepatitis C virus (HCV) kinetics, or more specifically, the effect of peg-interferon dose of 180 µg/week administered as a 24 hour infusion once a week for 4 weeks. The same sequence of 12 sampling times for both PK and PD measurements (in days, post-first-dose) was used for 30 subjects:

$$\Xi = \left(\left[\{\xi_{PK}, \xi_{PD}\}, N_{PKPD}\right]\right) = \left(\xi_{PK}, \xi_{PD}, 30\right)$$

$$\xi_{PK} = \xi_{PD} = \left(t_i\right) = \left(0, 0.25, 0.5, 1, 2, 3, 4, 7, 10, 14, 21, 28\right)$$

The HCV model is described by the following system of ODEs:

$$\frac{dX}{dt} = -k_a X(t) + r(t), \qquad X(0) = 0$$
$$\frac{dA}{dt} = k_a X(t) - k_e A(t), \qquad A(0) = 0$$
$$\frac{dT}{dt} = s - T(t)\left(eW(t) + d\right), \qquad T(0) = \frac{c\delta}{pe}$$
$$\frac{dI}{dt} = eW(t)T(t) - \delta I(t), \qquad I(0) = \frac{sep - dc\delta}{p\delta e}$$
$$\frac{dW}{dt} = p\left(1 - \frac{C(t)^n}{C(t)^n + EC_{50}^n}\right) I(t) - cW(t), \quad W(0) = \frac{sep - dc\delta}{c\delta e}$$

where $C(t) = A(t)/V_d$ is the drug concentration at time *t* and *r(t)* is the constant infusion rate. The viral dynamics model considers target cells, T, productively infected cells, I and viral particles, W. Target cells are produced at a rate s and die at a rate d. Cells become infected with de-novo infection rate *e*. After infection, these cells are lost with rate δ. In the absence of treatment, virus is produced by infected cells at a rate p and cleared at a rate c, for more details see (38, 39). The model for each response in subject *i* is defined as

$$y_{i,PK} = C(t_i) + \varepsilon_{iPK}$$
$$\log_{10}\left(y_{i,PD}\right) = \log_{10}\left(W(t_i)\right) + \varepsilon_{iPD}$$

An additive error model was assumed for both PK and PD (log viral load) compartments from which observations were drawn with a standard deviation of 0.2. Some of the parameters in the model are fixed (p, d, e, s). For the other seven parameters (k_a, k_e, V_d, EC_{50}, n, δ, c), log transformation was made with additive random effects on the log fixed effect with a variance ω^2 of 0.25. All parameters and their values are listed in Table 3.

Methods

For each example using each software tool, we computed the FIM based on the FO linearisation, given the parameters and the design. We used both the block-diagonal and the full FIM (not available in POPT). From the FIM, we computed the predicted SE values for each parameter and the

information D-criterion which is defined as the determinant of the FIM to the power of one over the number of parameters: $\left|\mathrm{FIM}\right|^{1/\dim(\psi)}$.

To investigate the FIM predictive performance, the empirical SE values were also estimated using CTS. More precisely, for each example, multiple data sets were simulated and then fitted using the Stochastic Approximation Expectation Maximisation (SAEM) algorithm in MONOLIX 2.4 (www.lixoft.eu) and, for the PK example also with the FOCEI algorithm in NONMEM 7 (http://www.iconplc.com/technology/products/nonmem/). Empirical standard errors were derived from the estimated parameters. The empirical D-criterion was computed from the normalized empirical variance-covariance matrix of all estimated parameters, $\left|\mathrm{cov}\left(\psi\right)^{-1}\right|^{1/\dim(\psi)}$. Because the CTS was much more time consuming for the HCV PKPD model, we did not perform the estimation with NONMEM and we did only 500 replicates, whereas we simulated 1000 replicates for the warfarin PK model.

For the CTS, to compute the empirical covariance matrix, the full variance-covariance matrix of all the estimated vectors was computed, not as two separate blocks for fixed effects and random components.

Results

For the PK model, the results show no differences between the optimal design software tools when evaluating the FIM using the block diagonal and full form. In the same way, all software reported the same expected D-criterion (Figure 1), and the same expected relative standard errors (RSE) values expressed in % (Table 4).

In this example, the block diagonal FIM calculations gave an expected D-criterion that was very similar to the observed D-criterion based on the inverse of the empirical covariance matrix (Figure 1). However, for all software, the block diagonal D-criterion is slightly smaller than the NONMEM FOCEI based criterion. Note that the result from MONOLIX is lower than the expected D-criterions, in line with theoretical expectations the Cramer-Rao inequality (FIM is an asymptotic upper bound on the information). The full FIM predicts considerably more information compared to the simulations (expected D-criterions are larger than the observed values), and predicts total information that is farther from the empirical values than the block diagonal calculations. The same trends are evident when looking at the RSE values, reported in Table 4. Good agreement between the CTS and the block

diagonal FIM was found, while the full FIM predicted considerably higher precision in β_{Ka} and $\beta_{CL/F}$.

For the more complicated PKPD model, results are summarized In Figure 2 and Table 5 where RSE (%) are reported. The D-criterion reveals negligible differences between any of the software (Figure 2) and also almost no difference between predicted SE values (Table 5). In this example, as in the PK example, using the block diagonal FIM gave D-criterion predicted values that were very similar to the D-criterion based on the inverse of the empirical covariance matrix (Figure 2). The full FIM predicts considerably more information compared to the simulations (expected D-criterions larger than the observed values) and predicts total information (D-criterion) that is farther from the empirical values than the block diagonal calculations. The same trends are evident when looking at SE values for each parameter (Table 5). We found good agreement between CTS and the block diagonal FIM, while the full FIM predicted higher precision in numerous parameters than observed.

Discussion

The first statistical developments for the evaluation of the FIM for NLMEM to compare and evaluate population designs without simulation were performed in the late 90's. Since then, five different software tools have been developed. We have compared these tools in terms of design evaluation. Optimisation was not considered in the present work. It should be noted that most software are under active development with regular addition of new features.

We compared the expression of the FIM computed by the five different optimal design software packages for two examples. The first example was a simple PK model for which the algebraic solution could be written analytically. When using the same approximation, all optimal design software packages achieved the same D-efficiency criterion and predicted RSE values (%). The second example was more complex, had two responses (both PK and PD measurements) and the model was written as a series of five differential equations. For this example, the D-criterion and RSE comparisons revealed negligible differences between software. The differences could potentially be explained by the use of different differential equation solvers, methods of implementing multiple response calculations, methods for computing numerical derivatives, tolerance levels for ODEs and numerical implementations of e.g. matrix inverses and solving of linear systems, etc. These small differences could be seen even across the MATLAB computations of the FIM. In this work we did not impose the same implementation of the various steps across software, hence the importance of the present comparison.

In both examples the expected SE values from the block diagonal FIM were close to the empirical SE values obtained from CTS. The runtimes for all software tools were a few seconds compared to minutes (warfarin example) or days (HCV example) for the CTS evaluation. Although computational speed has increased dramatically since the 1990's, a significant speed advantage is seen with the developed software tools even without considering design optimisation. For instance for the HCV PKPD model the CTS took several days for one design, so that optimization of doses and sampling times would be difficult.

In both examples investigated, the block diagonal FIM calculations give an expected D-criterion that is very similar to the observed D-criterion based on the inverse of the empirical covariance matrix and RSE(%) values for parameter match well. In contrast, the full FIM predicts more information compared to the simulations (expected D-criterions larger than the observed values). More discussion on the assumptions beyond the block or full matrix can be found in (33) together with

380 suggestions of other stochastic approaches. It seems that when using a FO approximation for
381 computation of the FIM, linearisation around some fixed values for the fixed effects which are then
382 no longer considered as estimable parameters and therefore corresponds to the block diagonal
383 matrix, provides the best approach. Also higher order approximations to the FIM are available that
384 may give better prediction of RSE(%) values (27).

385 Results using the simple FO approximation and the block diagonal FIM are very close to those
386 obtained by CTS using both FOCEI and SAEM estimation methods in the two examples. However,
387 since the expected FIM calculation is computing an asymptotically lower bound of the covariance of
388 the parameters, and the calculations are based on approximations, the authors suggest that a CTS
389 study of the proposed final design be performed in order to evaluate the likely performance of the
390 design in the setting in which it is proposed to be used. Since this would be a single CTS at a specified
391 design then this should not be computationally onerous compared to attempting to "optimise"
392 designs using CTS. In addition, using a CTS study of the final design makes it possible to assess the
393 bias which is not evaluated by the FIM.

394 In this first comparison between the software, we did only design evaluation for continuous data and
395 using the simpler FO approximation of the FIM. This first step was necessary before the next work
396 where we will compare results of design optimisation. Indeed now that we know that similar
397 criterion across software are obtained, we can compare the rather different optimisation algorithms
398 implemented. In principle any design variable that is present in the model can be optimised within an
399 optimal design framework. Examples of design variables that can be optimised are measurement
400 sampling times, doses, distribution of subjects between elementary designs, number of
401 measurement samples in an elementary design, etc. How this is done and which design variables can
402 be optimised varies between software, but the independent variable (e.g. measurement sampling
403 times) and the group assignment can be optimised in all software presented in this paper. Results will
404 depend on the assumptions about the model and the parameter values, so that sensitivity studies
405 should be performed to implement 'robust' designs, i.e. designs that are robust to the assumed *a*
406 *priori* values of the parameters. Approaches for design optimisation using a priori distribution of the
407 parameters were suggested and implemented for standard nonlinear regression and extended to
408 population approaches and should also be compared in further studies.

409 In conclusion, optimal design software tools allow for direct evaluation of population PKPD designs
410 and are now widely used in industry (25). Choice of software can depend on what platform the user
411 has available and what features they are looking since the FIM calculation in the different software

gives similar results. Population approaches are increasingly used and for more complex/ physiological PD models. It is very difficult to guess, without using one of these tools, what are the good designs for those complex ODE models and whether the study will be reliable. We suggest that before performing any population PKPD study, the design should be evaluated with a good balance between the approach based on the Fisher Matrix (for optimising the design) and CTS (for evaluating the final design).

Appendix : Development of the FIM in NLMEM for multiple responses using FO approximation

For each subject i with design ξ_i, the elementary Fisher information matrix is defined as

$$FIM\left(\psi,\xi_i\right)=E\left(-\frac{\partial^2 L_i\left(\psi;Y_i\right)}{\partial\psi\,\partial\psi^T}\right) \tag{14}$$

where $L_i\left(\psi;Y_i\right)$ is the log likelihood of the vector of observations Y_i given the population parameters ψ.

Let $F\left(\theta_i,\xi_i\right)=F\left(g\left(\beta,b_i\right),\xi_i\right)$, and $H\left(\theta_i,\xi_i,\varepsilon_i\right)=H\left(g\left(\beta,b_i\right),\xi_i,\varepsilon_i\right)$, be the vector composed of the K vectors of n_{ik} predicted responses $f_k\left(\theta_i,\xi_{ik}\right)$, and error $h_k\left(\theta_i,\xi_{ik},\varepsilon_{ik}\right)$ respectively. Then equation (5) can be written

$$Y_i=F(g(\beta,b_i),\xi_i)+H(g(\beta,b_i),\xi_i,\varepsilon_i)\,. \tag{15}$$

As there is no analytical expression for the log-likelihood $L_i\left(\psi;Y_i\right)$ for nonlinear models, a first-order Taylor expansion around the expectation of b_i is used:

$$F\left(g\left(\beta,b_i\right),\xi_i\right)\cong F(g(\beta,0),\xi_i)+\left(\left.\frac{\partial F(g(\beta,b_i),\xi_i)}{\partial b^T}\right|_{b_i=0}\right)b\,. \tag{16}$$

Then equation (16) can be approximated as:

$$Y_i\cong F(g(\beta,0),\xi_i)+\left(\left.\frac{\partial F\ \ (g(\beta,b_i),\xi_i)}{\partial b^T}\right|_{b_i=0}\right)b+H(g(\beta,0),\xi_i,\varepsilon_i)\,, \tag{17}$$

Therefore, $Y_i\sim \mathrm{N}\left(E_i,V_i\right)$ approximately, with marginal expectation E_i and variance V_i given by:

$$E(Y_i) \cong E_i = F(g(\beta,0),\xi_i) \tag{18}$$

$$Var\left(Y_i\right) \cong V_i = \left(\left. \frac{\partial F(g(\beta,b_i),\xi_i)}{\partial b^T} \right|_{b_i=0} \right) \Omega \left(\left. \frac{\partial F(g(\beta,b_i),\xi_i)}{\partial b^T} \right|_{b_i=0} \right)^T + \Sigma\left(\beta,\xi_i\right) \tag{19}$$

where $\Sigma\left(\beta,\xi_i\right)$ is the variance of $H(g(\beta,0),\xi_i,\varepsilon_i)$. $\Sigma\left(\beta,\xi_i\right)$ has a simple expression for usual error models where ε_i enters linearly, otherwise it can be computed using a first-order linearisation of H around the expectation of ε_i.

Then the elementary FIM for the fixed effects using the bock diagonal form (equation 13), has the following expression:

$$FIM(\beta,\xi_i) \cong J(\beta,\xi_i)^T V_i^{-1} J(\beta,\xi_i) \tag{20}$$

where $J(\beta,\xi_i) = \dfrac{\partial F(g(\beta,0),\xi_i)}{\partial \beta^T}$

Of note, when $\Omega = 0$, $FIM(\beta,\xi_i)$ reduces to the FIM for individual nonlinear regression with parameters β.

447 **References**

448 1. Box GEP, Lucas HL. DESIGN OF EXPERIMENTS IN NON-LINEAR SITUATIONS. Biometrika.
449 1959;46(1-2):77-90.

450 2. Draper NR, Hunter WG. The use of prior distributions in the design of experiments for
451 parameter estimation in non-linear situations: multiresponse case. Biometrika. 1967;54(3):662-5.
452 Epub 1967/12/01.

453 3. Atkinson AC, Hunter WG. The Design of Experiments for Parameter Estimation.
454 Technometrics. 1968;10(2):271-89.

455 4. D'Argenio DZ. Optimal sampling times for pharmacokinetic experiments. J Pharmacokinet
456 Biopharm. 1981;9(6):739-56.

457 5. Atkinson AC, Donev AN. Optimum experimental designs. Oxford: Clarendon Press; ed1992.

458 6. D'Argenio DZ. Advanced Methods of Pharmacokinetic and Pharmacodynamic Systems
459 Analysis. Inc. S-VNY, editor. New-York: Plenum Press; 1991. 318 p.

460 7. Fedorov VV, Leonov SL. Optimal Design for Nonlinear Response Models. Boca Raton:
461 Chapman & Hall/CRC Biostatistics Series; ed2013.

462 8. Landaw EM. Optimal design for individual parameter estimation in pharmacokinetics. New
463 York Raven Press; 1985. 181-8 p.

464 9. Pronzato L, Walter E. Robust experimental design via maximin optimization. Math Biosci.
465 1988;89(2):161-76.

466 10. Hennig S, Nyberg J, Fanta S, Backman JT, Hoppu K, Hooker AC, et al. Application of the
467 optimal design approach to improve a pretransplant drug dose finding design for ciclosporin. Journal
468 of clinical pharmacology. 2012;52(3):347-60. Epub 2011/05/06.

469 11. Merle Y, Mentre F, Mallet A, Aurengo AH. Designing an optimal experiment for Bayesian
470 estimation: application to the kinetics of iodine thyroid uptake. Stat Med. 1994;13(2):185-96. Epub
471 1994/01/30.

472 12. Sheiner LB, Beal SL. Evaluation of methods for estimating population pharmacokinetic
473 parameters. III. Monoexponential model: routine clinical pharmacokinetic data. J Pharmacokinet
474 Biopharm. 1983;11(3):303-19. Epub 1983/06/01.

475 13. Al-Banna MK, Kelman AW, Whiting B. Experimental design and efficient parameter
476 estimation in population pharmacokinetics. J Pharmacokinet Biopharm. 1990;18(4):347-60.

477 14. Ette EI, Kelman AW, Howie CA, Whiting B. Analysis of animal pharmacokinetic data:
478 performance of the one point per animal design. J Pharmacokinet Biopharm. 1995;23(6):551-66.
479 Epub 1995/12/01.

480 15. Jonsson EN, Wade JR, Karlsson MO. Comparison of some practical sampling strategies for
481 population pharmacokinetic studies. J Pharmacokinet Biopharm. 1996;24(2):245-63.

482 16. Ette EI, Sun H, Ludden TM. Balanced designs in longitudinal population pharmacokinetic
483 studies. Journal of clinical pharmacology. 1998;38(5):417-23. Epub 1998/05/29.

484 17. U.S. Department of Health and Human Services FaDA. Guidance for Industry. Population
485 Pharmacokinetics. 1999.

486 18. Mentré F, Mallet A, Baccar D. Optimal design in random effect regression models.
487 Biometrika. 1997;84(2):429-42.

488 19. Titterington DM, Cox DR. Biometrika: One Hundred Years. Oxford: Oxford University Press;
489 2001.

490 20. Retout S, Duffull S, Mentre F. Development and implementation of the population Fisher
491 information matrix for the evaluation of population pharmacokinetic designs. Comput Methods
492 Programs Biomed. 2001;65(2):141-51.

493 21. Duffull S, Waterhouse T, Eccleston J. Some considerations on the design of population
494 pharmacokinetic studies. J Pharmacokinet Pharmacodyn. 2005;32(3-4):441-57. Epub 2005/11/15.

495 22. Foracchia M, Hooker A, Vicini P, Ruggeri A. POPED, a software for optimal experiment design
496 in population kinetics. Comput Meth Prog Bio. 2004;74(1):29-46.

497 23. Gueorguieva I, Ogungbenro K, Graham G, Glatt S, Aarons L. A program for individual and
498 population optimal design for univariate and multivariate response pharmacokinetic-
499 pharmacodynamic models. Comput Methods Programs Biomed. 2007;86(1):51-61.

500 24. Aliev A, Fedorov V, Leonov S, McHugh B, Magee M. PkStaMp Library for Constructing Optimal
501 Population Designs for PK/PD Studies. Commun Stat Simul Comput. 2012;41(6):717-29.

502 25. Mentré F, Chenel M, Comets E, Grevel J, Hooker A, Karlsson MO, et al. Current use and
503 developments needed for optimal design in pharmacometrics : a study performed amongst
504 DDMoRe's EFPIA members. 2012.

505 26. Gueorguieva I, Aarons L, Ogungbenro K, Jorga KM, Rodgers T, Rowland M. Optimal design for
506 multivariate response pharmacokinetic models. J Pharmacokinet Pharmacodyn. 2006;33(2):97-124.
507 Epub 2006/03/22.

508 27. Nyberg J, Ueckert S, Stromberg EA, Hennig S, Karlsson MO, Hooker AC. PopED: an extended,
509 parallelized, nonlinear mixed effects models optimal design tool. Comput Methods Programs
510 Biomed. 2012;108(2):789-805. Epub 2012/05/30.

511 28. Gagnon R, Leonov S. Optimal population designs for PK models with serial sampling. J
512 Biopharm Stat. 2005;15(1):143-63. Epub 2005/02/11.

513 29. Ogungbenro K, Graham G, Gueorguieva I, Aarons L. Incorporating correlation in
514 interindividual variability for the optimal design of multiresponse pharmacokinetic experiments. J
515 Biopharm Stat. 2008;18(2):342-58. Epub 2008/03/11.

516 30. Retout S, Mentré F. Further developments of the Fisher information matrix in nonlinear
517 mixed effects models with evaluation in population pharmacokinetics. J Biopharm Stat.
518 2003;13(2):209-27.

519 31. Foracchia M, Hooker A, Vicini P, Ruggeri A. POPED, a software for optimal experiment design
520 in population kinetics. Comput Methods Programs Biomed. 2004;74(1):29-46.

521 32. Leonov S, Aliev A. Optimal design for population PK/PD models. Tatra Mt Math Publ.
522 2012;51:115-30.

523 33. Mielke T, Schwabe S. Some Considerations on the Fisher Information in Nonlinear Mixed
524 Effects Models. In: Springer-Verlag, editor. mODa 9 – Advances in Model-Oriented Design and
525 Analysis Berlin: Physica-Verlag HD; 2010. p. 129-36.

526 34. O'Reilly RA, Aggeler PM, Leong LS. STUDIES ON THE COUMARIN ANTICOAGULANT DRUGS:
527 THE PHARMACODYNAMICS OF WARFARIN IN MAN. J Clin Invest. 1963;42:1542-51. Epub 1963/10/01.

528 35. Bazzoli C, Retout S, Mentre F. Design evaluation and optimisation in multiple response
529 nonlinear mixed effect models: PFIM 3.0. Comput Methods Programs Biomed. 2010;98(1):55-65.
530 Epub 2009/11/07.

531 36. Retout S, Comets E, Samson A, Mentre F. Design in nonlinear mixed effects models:
532 optimization using the Fedorov-Wynn algorithm and power of the Wald test for binary covariates.
533 Stat Med. 2007;26(28):5162-79. Epub 2007/05/09.

534 37. O'Reilly RA, Aggeler PM. Studies on coumarin anticoagulant drugs. Initiation of warfarin
535 therapy without a loading dose. Circulation. 1968;38(1):169-77. Epub 1968/07/01.

536 38. Guedj J, Bazzoli C, Neumann AU, Mentre F. Design evaluation and optimization for models of
537 hepatitis C viral dynamics. Stat Med. 2011;30(10):1045-56. Epub 2011/02/22.

538 39. Neumann AU, Lam NP, Dahari H, Gretch DR, Wiley TE, Layden TJ, et al. Hepatitis C viral
539 dynamics in vivo and the antiviral efficacy of interferon-alpha therapy. Science (New York, NY).
540 1998;282(5386):103-7. Epub 1998/10/02.

541

542

Table 1. Available features in the software tools available for population design evaluation

	Software				
	PFIM	**PkStaMp**	**PopDes**	**PopED**	**POPT**
Language	R	Matlab	Matlab	Matlab FreeMat	Matlab FreeMat
Available on website	✓		✓	✓	✓
Library of PKPD models	✓	✓	✓	✓	✓
User defined models	✓	✓	✓	✓	✓
Multi-response models	✓	✓	✓	✓	✓
Designs differ across responses	✓	✓	✓	✓	✓
ODE models	✓	✓	✓	✓	✓
Full FIM	✓	✓	✓	✓	
Full covariance matrix for Ω		✓	✓	✓	
Full covariance matrix for Σ			✓	✓	
IOV	✓		✓	✓	
Discrete covariates/ power	✓/✓		✓ /	✓/✓	✓/

FIM: Fisher Information matrix; GUI: Graphical User Interface; IOV: Inter-Occasion Variability; ODE: Ordinary Differential Equation; Ω: interindividual covariance matrix ; Σ: residual covariance matrix ;

Table 2 - Model parameters of warfarin PK model

Parameter	Value
$\beta_{CL/F}$ (L/h)	0.15
$\beta_{V/F}$ (L)	8.00
β_{ka} (1/h)	1.00
$\omega^2_{CL/F}$	0.07
$\omega^2_{V/F}$	0.02
ω^2_{ka}	0.60
σ^2	0.01

CL/F: apparent clearance of the warfarin; *V/F*: apparent volume of the warfarin; *ka*: constant of absorption of the warfarin; β : fixed effects ; ω^2 : interindividual variance; σ^2 : residual variance.

Table 3 - Model parameters for HCV PK/VK model

Parameter	Value
p (fixed) [1]	100
d (1/d) (fixed) [1]	0.001
e (mL/d) (fixed) [1]	1E-07
s (mL^{-1}/d) (fixed) [1]	20 000
β_{ka} (1/d)	0.80
β_{ke} (1/d)	0.15
β_{Vd} (mL)	100 000
$\beta_{EC_{50}}$ (µg/mL)	0.00012
β_n	2
β_δ (1/d)	0.20
β_c (1/d)	7
ω^2_{ka}	0.25
ω^2_{ke}	0.25
ω^2_{Vd}	0.25
$\omega^2_{EC_{50}}$	0.25
ω^2_n	0.25
ω^2_δ	0.25
ω^2_c	0.25
σ^2_{PK}	0.04
σ^2_{PD}	0.04

[1] Parameters defined in the section 3.

ka: rate constant of absorption; *ke*: rate constant of elimination; *Vd*: volume of distribution; EC_{50}: drug concentration in the blood at which the drug is 50% effective; *n*: Hill coefficient; *δ*: rate constant of elimination of infected celss; *c*: rate constant of elimination of viral particles; β : fixed effects; ω^2 : interindividual variance; σ^2_{PK} : residual variance for the PK response; σ^2_{PD} : residual variance for the PD response.

Table 4 - Fisher Information Matrix (FIM) predicted RSEs (%) for warfarin PK model with the various software tools compared to empirical RSEs (%)

	Block diagonal FIM	Full FIM	Simulations	
Parameter	**PFIM/PkStaMp/PopDes/PopED/POPT**	**PFIM/PkStaMp/PopDes/PopED**	**NONMEM**	**MONOLIX**
β_{ka}	13.9	4.8	13.6	13.8
$\beta_{CL/F}$	4.7	3.6	4.9	4.8
$\beta_{V/F}$	2.8	2.6	2.7	2.8
ω^2_{ka}	25.8	26.5	26.6	28.1
$\omega^2_{CL/F}$	25.6	26.3	26.1	26.6
$\omega^2_{V/F}$	30.3	30.9	32.4	30.8
σ^2	11.2	12.4	10.9	11.0

CL/F: apparent clearance of the warfarin; *V/F*: apparent volume of the warfarin; *ka*: constant of absorption of the warfarin; β : fixed effects ; ω^2 : interindividual variance; σ^2 : residual variance.

Table 5 – Fisher Information Matrix (FIM) predicted RSEs (%) for the HCV model parameters with the various software tools compared to empirical RSEs

	Block diagonal FIM			Full FIM		Simulations
Parameter	**PFIM**	**PkStaMp/PopDes/PopED**	**POPT**	**PFIM**	**PkStaMp/PopDes/PopED**	**MONOLIX**
β_{ka}	12.0	12.1	13.2	8.6	8.6	12.2
β_{ke}	10.4	10.5	11.1	6.8	6.9	10.4
β_{Vd}	9.9	10.0	11.2	8.3	8.4	9.9
$\beta_{EC_{50}}$	15.8	15.8	15.7	13.6	13.5	14.5
β_{n}	10.5	10.4	10.4	7.4	7.5	10.6
β_{δ}	9.5	9.4	9.4	8.7	8.5	10.1
β_{c}	11.1	11.0	11.0	8.8	8.7	10.3
ω^2_{ka}	39.6	40.0	42.0	42.8	43.2	41.6
ω^2_{ke}	30.4	30.8	31.6	36.4	37.2	34.4
ω^2_{Vd}	28.4	28.8	31.6	32.8	33.2	30.4
$\omega^2_{EC_{50}}$	60.8	60.4	60.0	66.4	66.4	53.2
ω^2_{n}	28.8	28.8	28.8	32.8	32.8	31.6
ω^2_{δ}	27.2	27.2	27.2	32.4	31.6	31.6
ω^2_{c}	32.8	32.8	32.4	34.0	33.6	30.0
σ^2_{PK}	9.0	8.5	8.3	9.3	8.5	10.0
σ^2_{PD}	8.0	9.0	9.0	8.5	9.3	9.0

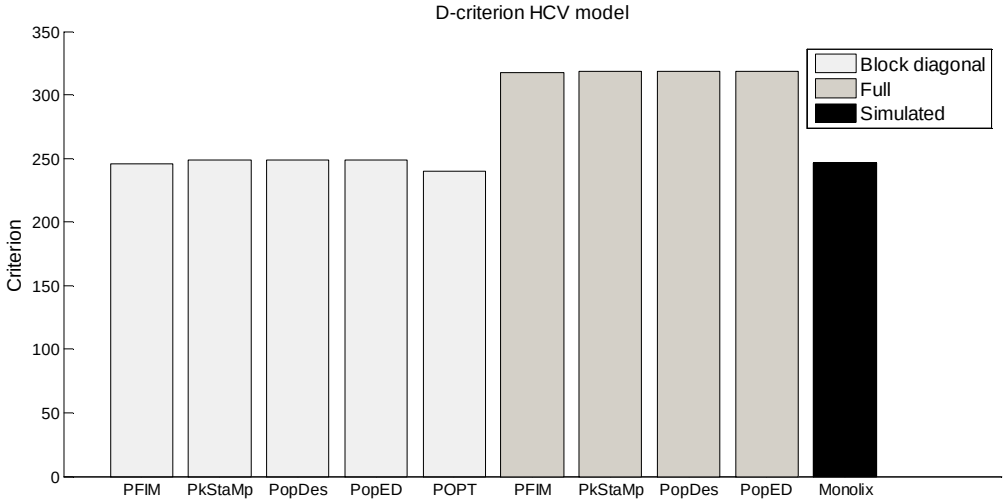
ka: rate constant of absorption; *ke*: rate constant of elimination; *Vd*: volume of distribution; EC_{50}: drug concentration in the blood at which the drug is 50% effective; *n*: Hill coefficient; *δ*: rate constant of elimination of infected celss; *c*: rate constant of elimination of viral particles; β : fixed effects; ω^2 : interindividual variance; σ_{PK}^2 : residual variance for the PK response; σ_{PD}^2 : residual variance for the PD response.

Figures

Figure 1. D-criterion predicted by the different software tools for the warfarin PK model compared to simulated D-criterion calculated from the inverse of the empirical covariance matrix.

NM FOCEI is calculated from the estimates using the first-order conditional estimation method with interaction in NONMEM. The Monolix criterion is calculated from the estimates using the SAEM algorithm in Monolix.

Figure 2. D-criterion predicted by the different software tools for the HCV model. Simulated D-criterion calculated from the inverse of the empirical covariance matrix.



The Monolix criterion is calculated from the estimates using the SAEM algorithm in Monolix.