Journal of Pharmacokinetics and Pharmacodynamics (2022) 49:557–577
https://doi.org/10.1007/s10928-022-09821-z

ORIGINAL PAPER

Impact of model misspecification on model-based tests in PK studies with parallel design: real case and simulation studies

Mélanie Guhl[1] · François Mercier[2] · Carsten Hofmann[3] · Satish Sharan[4] · Mark Donnelly[4] · Kairui Feng[4] · Wanjie Sun[5] · Guoying Sun[5] · Stella Grosser[5] · Liang Zhao[4] · Lanyan Fang[4] · France Mentré[1] · Emmanuelle Comets[1,6] · Julie Bertrand[1]

Received: 1 February 2022 / Accepted: 11 August 2022 / Published online: 16 September 2022

Abstract

This article evaluates the performance of pharmacokinetic (PK) equivalence testing between two formulations of a drug through the Two-One Sided Tests (TOST) by a model-based approach (MB-TOST), as an alternative to the classical non-compartmental approach (NCA-TOST), for a sparse design with a few time points per subject. We focused on the impact of model misspecification and the relevance of model selection for the reference data. We first analysed PK data from phase I studies of gantenerumab, a monoclonal antibody for the treatment of Alzheimer's disease. Using the original rich sample data, we compared MB-TOST to NCA-TOST for validation. Then, the analysis was repeated on a sparse subset of the original data with MB-TOST. This analysis inspired a simulation study with rich and sparse designs. With rich designs, we compared NCA-TOST and MB-TOST in terms of type I error and study power. With both designs, we explored the impact of misspecifying the model on the performance of MB-TOST and adding a model selection step. Using the observed data, the results of both approaches were in general concordance. MB-TOST results were robust with sparse designs when the underlying PK structural model was correctly specified. Using the simulated data with a rich design, the type I error of NCA-TOST was close to the nominal level. When using the simulated model, the type I error of MB-TOST was controlled on rich and sparse designs, but using a misspecified model led to inflated type I errors. Adding a model selection step on the reference data reduced the inflation. MB-TOST appears as a robust alternative to NCA-TOST, provided that the PK model is correctly specified and the test drug has the same PK structural model as the reference drug.

Keywords Equivalence test · Pharmacokinetics · Non-compartmental analysis · Non-linear mixed effects models · Sparse design

✉ Mélanie Guhl
melanie.guhl@inserm.fr

[1] Université Paris Cité and Université Sorbonne Paris Nord, Inserm, IAME, 75018 Paris, France

[2] Department of Biostatistics, Roche Innovation Center Basel, Basel, Switzerland

[3] Department of Clinical Pharmacology, Roche Innovation Center Basel, Basel, Switzerland

[4] Division of Quantitative Methods and Modeling, Office of Research Standards, Office of Generic Drugs, Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, MD 20993, USA

[5] Office of Biostatistics, Office of Translational Sciences, Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, MD 20993, USA

[6] Univ Rennes, Inserm, EHESP, Irset - UMR_S 1085, 35000 Rennes, France

Introduction

In bioequivalence (BE) studies with pharmacokinetic (PK) endpoints (for generics), or PK similarity studies (for biologicals), we aim to compare the exposure after administration of different drug formulations by comparing two PK parameters of interest: the area under the curve (AUC) of the plasma concentration as a function of time, and the maximal concentration (C_{max}).

BE studies are an essential part of drug development and still an active research field. Currently, a key science and research priority at the U.S. Food and Drug Administration (FDA) is to "improve quantitative pharmacology and BE trial simulation to optimise the design of BE studies for generic drug products and establish a foundation for model-based BE study designs" [1].

The classical statistical test used to assess BE is the Two One Sided Tests (TOST) proposed by Schuirmann in 1987 [2]. It consists of two t tests, on PK parameters of interest, comparing the difference of treatment effects computed to a threshold δ. The FDA as well as the European Medicines Agency (EMA) fix this threshold to $\delta = log(0.8)$ and $\delta = log(1.25)$ [3, 4].

FDA and EMA recommend estimating BE treatment effects via non-compartmental analysis (NCA) for both crossover and parallel study designs [3, 4]. However, assessment of PK equivalence may be challenging for PK BE studies with sparse sampling, such as in participants receiving ophthalmic or oncology drug products. PK BE studies for ophthalmic drug products typically involve a sparse design with one sampling time point per subject (or per treatment group per subject in a crossover design). In such studies, FDA recommends BE to be assessed using a non-parametric bootstrap NCA-based approach or a parametric method [5, 6]. This type of sparse study design may be useful for certain drug products or may occur from study interruptions due to the COVID-19 pandemic or other causes.

An alternative proposed by Dubois et al. [7] is to use a model-based (MB) approach, using the empirical Bayes estimated (EBE) individual parameters of a non-linear mixed effects model instead of NCA parameters. They showed that this method leads to an increase in type I error when the EBE shrinkage is above 20%, which is frequent in case of sparse design. Dubois et al. [8] also proposed a MB approach, this time inferring on the population parameters. They showed that this MB approach works as well as the NCA on rich designs and can be applied on sparser designs. Currently, it is unclear when MBBE methods would be preferred over traditional BE approaches. As such, FDA has actively supported research focused on MBBE approaches for PK BE studies with sparse designs [9–11]. Indeed, MB tests can lead to an inflation of the type I error because of an underestimation of the standard error (SE) of treatment effects on sparse designs in presence of large variability, which led Loingeville et al. to propose and evaluate methods of correction of the standard errors in MB studies [10]. Shen et al. [12] also proposed a MB alternative to traditional BE tests. In this MBBE approach, rich individual PK profiles are simulated from the model and NCA is performed to estimate individual AUC and C_{max} values. Since TOST was based on individual predicted values, the authors assessed distributional assumptions.

MB approaches involve the selection of a PK model to fit the data, which raises the question of the impact of model misspecification on the results of the equivalence tests.

In this study, we define a "sparse" design as any study with only a few sampling points and that challenges the identifiability of the model, which means that the sparse nature of data depends on the complexity of the model of interest.

Our work was based on data collected during the development of gantenerumab, a monoclonal antibody for the treatment of Alzheimer's disease. As this drug has a very long half-life, the clinical trials were conducted using a parallel design (more than 13 weeks of follow up), which is not the classical design for PK equivalence studies that are usually conducted using a crossover design.

In this real case, we compared the PK data gathered in participants treated with two formulations of gantenerumab. Then, we evaluated the performance of the MB approach on simulations based on data from this study and assessed the impact of study design, model misspecification, and the relevance of a model selection step. Although this assessment was based on PK data from a monoclonal antibody, our novel method may potentially be used to evaluate BE studies in generic drug development when there is sparse PK sampling.

We first present the theoretical background, i.e., the NCA and MB approach for equivalence TOST tests. We then describe the observed data, the methodology to analyse it and the results of this real case study. We finally present the design, methods and results of the simulation study, and discuss our findings in the last section.

Theoretical background

Two One-Sided Tests

Showing the PK equivalence of two drug formulations, one reference (R) and one test (T), means showing their exposure is equivalent.

In PK BE studies, drug exposure is typically characterised by two PK parameters, variables of the plasma concentration versus time profiles : the Area Under the Curve (AUC), which can be computed from 0 to the last sampling point (AUC_{tlast}) or extrapolated to infinity (AUC_{∞}), and the maximum plasma concentration (C_{max}). Treatment effects on AUC and C_{max}, namely θ_{AUC} and $\theta_{C_{max}}$, are defined as the difference of the expectation of the log individual values of these variables under test and reference treatment. For instance:

$$\theta_{AUC} = \mathbb{E}(log(AUC_T)) - \mathbb{E}(log(AUC_R)) \tag{1}$$

Since we wish to reject the assumption that the two formulations have different exposures, we write the null hypothesis as [2]:

$$H_0 : \{\theta \leq -\delta \text{ or } \theta \geq \delta\} \tag{2}$$

where δ is the tolerance. The regulatory guidances for equivalence studies fix the threshold $\delta = log(1.25)$ [3, 4].

By decomposing this null hypothesis in two, we perform Two One-Sided Tests (TOST):

$$H_{0,-\delta} : \{\theta \leq -\delta\} \text{ and } H_{0,\delta} : \{\theta \geq \delta\} \tag{3}$$

The two t test statistics are rejected at $\alpha = 5\%$ if:

$$Z_{-\delta} = \frac{\theta + \delta}{SE(\theta)} \geq q_{1-\alpha} \text{ and } Z_{\delta} = \frac{\theta - \delta}{SE(\theta)} \leq q_{\alpha} \tag{4}$$

with q_{α} the quantile of order α of a reference distribution.

Equivalently, we can reject the null hypothesis if the confidence interval of θ is within $[-\delta, \delta]$, that is if the confidence interval of the exponential of θ is within [0.8 ; 1.25]. The exponential of θ is often shown in the results of the test and is called the geometric mean ratio (GMR).

Non-compartmental analysis

The standard method for PK equivalence studies is to compute individual AUC and C_{max} and use an ANOVA or a linear mixed model to estimate the treatment effect. AUC_{tlast} can be computed using the trapezoidal method and AUC_{∞} can be estimated by linear extrapolation. For this, FDA recommends that sampling continues for at least three or more terminal elimination half-lives of the drug and there are at least three sampling points after the peak [3]. C_{max} is defined as the maximal concentration measured among the study sampling times.

Depending on the study design, there can be a period and a sequence effect on the variables of interest. In parallel studies, there is only one period: each group of participants receives one treatment only. Our present work focuses on a drug with a long half-life which warrants a parallel study design instead of the classical crossover design for PK equivalence studies. In this case, there is no period or sequence effect and intra-individual variability cannot be properly evaluated. The models to fit are simply:

$$log(AUC_i) = \mu_{AUC} + \theta_{AUC} T_i + \epsilon_{AUC_i} \tag{5}$$

$$log(C_{max_i}) = \mu_{C_{max}} + \theta_{C_{max}} T_i + \epsilon_{C_{max_i}} \tag{6}$$

with:

- μ: mean value of variable for the reference treatment;
- T_i: treatment covariate variable for individual i;
- θ: coefficient of treatment effect;
- $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$: residual error.

The treatment effects on the variables of interest and their standard errors are obtained directly from the linear model inference.

The geometric mean ratio is, e.g. for AUC:

$$\begin{aligned} GMR &= \frac{exp(\mathbb{E}(log(AUC_T)))}{exp(\mathbb{E}(log(AUC_R)))} \\ &= \frac{exp(\mu_{AUC} + \theta_{AUC})}{exp(\mu_{AUC})} \\ &= exp(\theta_{AUC}) \end{aligned}$$

In non-compartmental PK equivalence analyses (hereafter called NCA-TOST), the standard error is obtained with the Fisher Information Matrix (FIM), which is asymptotically the inverse of the lower bound of the variance-covariance matrix of regression coefficients. With balanced groups, the reference distribution to use in NCA-TOST is a Student's t distribution with N-2 degrees of freedom, N being the number of participants in the study.

Model-based approach

Regulatory requirements may not be met in studies with sparse sampling design, and NCA-TOST may then become less accurate. Indeed, it can be hard to compute individual AUC and C_{max} if we only have a few points per subject. In an effort to leverage population data over time to inform predictions for individuals, a model-based alternative has been proposed [8, 10], in which we build a structural PK model and use a non-linear mixed effect model (NLMEM) to estimate the treatment effect. The corresponding statistical model can be written as follows in the case of parallel studies:

$$y_{ij} = f(t_{ij}, \phi_i) + g(t_{ij}, \phi_i)\epsilon_{ij} \tag{7}$$

$$log(\phi_{il}) = log(\mu_l) + \theta_l T_i + \eta_{il} \tag{8}$$

with:

- t_{ij}: time j for individual i;
- y_{ij}: concentration for individual i at time t_{ij};

- ϕ_i: vector of parameters for individual i (typically of size 3 to 10);
- $f(t_{ij}, \phi_i)$: non-linear structural PK model depending on ϕ_i;
- $g(t_{ij}, \phi_i)$: error model;
- $\epsilon_{ij} \sim \mathcal{N}(0,1)$: residual error;
- μ_l: fixed effect for parameter l;
- T_i: treatment covariate variable;
- θ_l: coefficient of treatment effect for parameter l;
- $\eta_{il} \sim \mathcal{N}(0, \omega_l)$: between subject random effect for parameter l;
- ω_l: standard deviation of the inter-individual random effect for parameter l.

$g()$ describes the error model, with usual models being:

- Additive error model: $g(t_{ij}, \phi_i) = \sigma_a$;
- Multiplicative error model:$g(t_{ij}, \phi_i) = \sigma_b\, f(t_{ij}, \phi_i)$;
- Combined error model: $g(t_{ij}, \phi_i) = \sigma_a + \sigma_b\, f(t_{ij}, \phi_i)$.

In the context of BE studies, we usually have previous knowledge on the underlying PK characteristics of the reference product, which could be described by a subset of structural PK models *f*().

In this study, we only fitted and compared PK models that differed in terms of number of compartments, order of absorption, and presence of an absorption delay. A description of all the models used in this study can be found in Appendix 1, defining the vector μ of l parameters related to each model.

Computation of standard errors

In this study, we used and compared three different methods of computation of SE in the MB approach, that are described below, and called ”Asympt”, ”Gallant” and ”Post”. These three methods have also been evaluated in the context of BE studies by Loingeville et al. [10].

Asympt

AUC and C_{max} are secondary PK parameters of the models, i.e., functions derived from the PK model direct parameters, and their treatment effects are also functions of the PK model direct parameters and treatment effect: $\theta = h(\mu_{PK}, \theta_{PK})$. For instance, for all PK models with a linear elimination, $AUC_\infty = \frac{FD}{CL}$, where D is the dose administered, F the bioavailability of the drug and CL the clearance, so the treatment effect on AUC_∞ can be simply derived from the model as $\theta_{AUC_\infty} = -\theta_{CL/F}$ and $SE(\theta_{AUC_\infty}) = SE(\theta_{CL/F})$. In one compartment models, there are analytical solutions for all secondary PK parameters, so the delta-method can be used to compute the standard errors of treatment effects. In two-compartment models, there is no analytical solution for C_{max}, so we need to compute $\theta_{C_{max}}$ and its standard error by simulation. This method consists of sampling parameters from a multinormal distribution with maximum likelihood estimates as the mean vector and the inverse of the FIM as the variance-covariance matrix, to simulate rich concentration profiles for reference and test treatments (see Appendix 2 for a more precise description of the method).

In this approach (which will be designated hereafter by MB-TOST Asympt), the standard error computed in NLMEM is also obtained with the FIM, using a linearisation of the PK model.

The reference distribution we use in MB-TOST Asympt is a Gaussian distribution with zero mean and a standard deviation equal to 1.

In the MB approach, an underestimation of the asymptotic standard errors of the treatment effects has been observed which resulted in an inflation of type I error when performing PK equivalence tests [8]. To address this, several methods of correction of the asymptotic standard errors have been suggested. Here, we use two methods of correction, designated Gallant and Post, which were proposed for equivalence tests by Loingeville et al. [10].

Gallant

The Gallant correction [13] (MB-TOST Gallant) aims to take into account the number of parameters estimated towards the available data to correct for the underestimation of the standard errors of treatment effects. It involves re-weighting the standard errors using the following formula:

$$SE_{Gallant} = SE\sqrt{\frac{N}{N-p}} \tag{9}$$

with N the number of participants in the study and p the number of fixed and covariate effects (here, we only have the treatment as a covariate).

We also switch the reference distribution used in the tests from a Gaussian distribution to a Student’s t distribution with $N-p$ degrees of freedom.

Post

This method (MB-TOST Post) uses posterior distribution samples to compute the standard errors of treatment effects [10].

Samples of population parameters are generated by Bayesian inference, with the Hamiltonian Monte Carlo algorithm. Maximum likelihood estimates obtained with NLMEM are used as initial values. Uniform priors are used for the fixed and treatment effects and Half-Cauchy distributions with zero mean and a standard deviation equal to 1 for the random effects and residual error variance parameters.

When the data are not informative enough given the number of model parameters to estimate, these priors can result in chains with low N_{eff} and high $\hat{R}$. When $N_{eff} \leq 400$ and $\hat{R} \geq 1.05$, log normal priors can be used for the fixed effects, with mean equal to the maximum likelihood estimation and a standard deviation equal to 0.5 and normal priors with zero mean and standard deviation equal to 0.5 for the treatment effects as in [10].

The standard errors of treatment effects are computed using samples from the posterior distribution.

The reference distribution, as for MB-TOST Asympt, is a Gaussian distribution with zero mean and a standard deviation equal to 1.

Case study: gantenerumab

Data

In our analysis, PK data was collected from two phase I randomised clinical trials on healthy male or female subjects between 40-70 years of age. These trials investigated the relative bioavailability, tolerability, and dose-exposure relationship of a high concentration liquid formulation (HCLF G3) versus a lyophilised formulation (LyoF G2) of gantenerumab, a monoclonal antibody used for the treatment of Alzheimer's disease. Hereafter we considered the high concentration liquid formulation as the reference formulation. Both formulations were administered by subcutaneous injection. The first study (NCT01636531, here called S1) was composed of five parallel arms with 24 participants each: three reference arms at different dose levels (105, 225 and 300 mg) and two test arms (105 and 225 mg). In the second study (NCT02133937, here called S2), composed of one reference arm of 25 participants and one test arm of 23 participants, the dose tested was 225 mg. PK sampling was performed in participants for up to 13 weeks using the following scheme: 0.25, 1, 2, 3, 4, 7, 13, 20, 42, 63, and 84 days post dose. There was one additional sampling time in S2, one hour post dose (0.04 days). We evaluated PK equivalence of the two formulations in terms of C_{max} and AUC_{∞}.

Methods

We performed separate analyses for each study and dose tested, hereafter called S1-105, S1-225 and S2-225, discarding the 300 mg arm of S1 as this study did not include a test treatment arm at this dose.

On the original rich design data (11 sampling points per subject), different structural PK models and residual error models were fitted on the reference arms, and compared for selection purposes. The structural PK models tested differed in terms of number of compartments (one or two), order of absorption (zero or one) and presence of an absorption delay. A description of all these models can be found in Appendix 1. As we work on a drug administered by sub-cutaneous injection, the parameters of the PK models used are apparent parameters scaled by the bioavailability of the drug F. Inter-individual variability followed a log-normal distribution for all parameters. Three types of error models were tested: additive, multiplicative and combined. Models were compared using the Bayesian Information Criterion (BIC) computed by Importance Sampling, combined with a second criteria of a relative SE (RSE) below 50% for all parameters. Inter-individual variability parameters that did not meet this second criteria were removed. We also explored the relevance of adding a correlation between the inter-individual variabilities. Goodness of fit was assessed with Visual Predictive Checks (VPC) and Normalised Prediction Distribution Errors (NPDE) [14]. The selected PK model was then fitted on both the reference and test arms and treatment effects were estimated on all parameters. We compared the results of MB-TOST, using only the Asympt computation method for the SE, with results obtained with NCA-TOST which usually performs well on such rich designs.

MB analyses were also run on a sparse subset of the data to explore the impact of the study design. The sparse subset for each study contained 5 points per subject because it is the maximum number of population parameters that we needed to estimate, in order to make the model identifiable. These points were obtained by optimisation of the design with *PFIM* [15] (Population Fisher Information Matrix, an algorithm for the evaluation and optimisation of designs), using the model fitted on the rich reference and test arms. Given that this manuscript focuses on the investigation of MB methods as an alternative for sparse design, we tested the PK equivalence only with MB-TOST, selecting again the PK structural model on the reference arm. Three methods to compute the SE were used: Asympt, Gallant and Post.

Implementation

Analyses were run on R version 4.0.2. Parameters of the PK models were estimated by maximising the likelihood using the Stochastic Approximation of Expectation Maximisation algorithm (SAEM) [16], in the *saemix* R package [17] (development version: https://github.com/saemixdevelopment/saemixextension). For NCA-TOST, AUC_∞ was computed by extrapolation with the *PKNCA* R package [18] version 0.9.4, using the observed concentration at t_{last}. Sampling points for the sparse designs were chosen with the *PFIM* [15] R package version 4.0 which enables to optimise population design using the Fedorov–Wynn algorithm.

Results

Figure 1 shows spaghetti plots of the plasma concentrations of gantenerumab versus time in log-scale, for the two lower doses in each study.

The same model, a two-compartment model (V_1/F: apparent volume of the principal compartment, V_2/F:

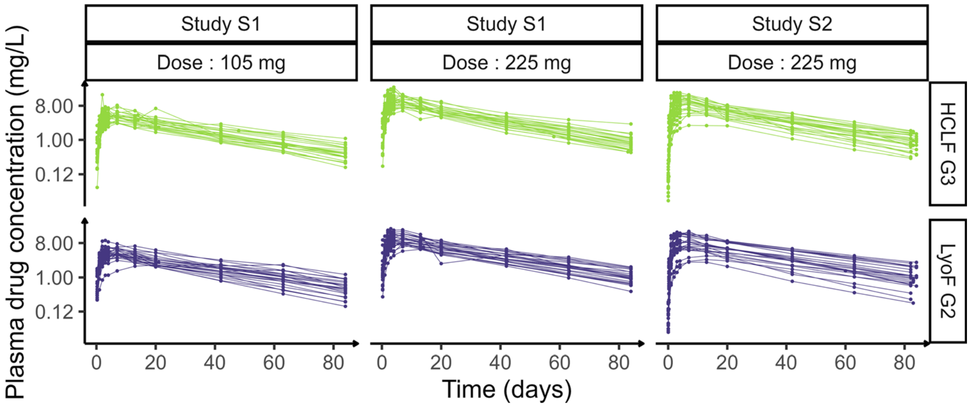


Fig. 1 Individual concentration versus time profiles, in log scale, in studies S1 and S2 per dose (105 and 225 mg), in the reference (HCLF G3) and test (LyoF G2) treatment arms (colour figure online)

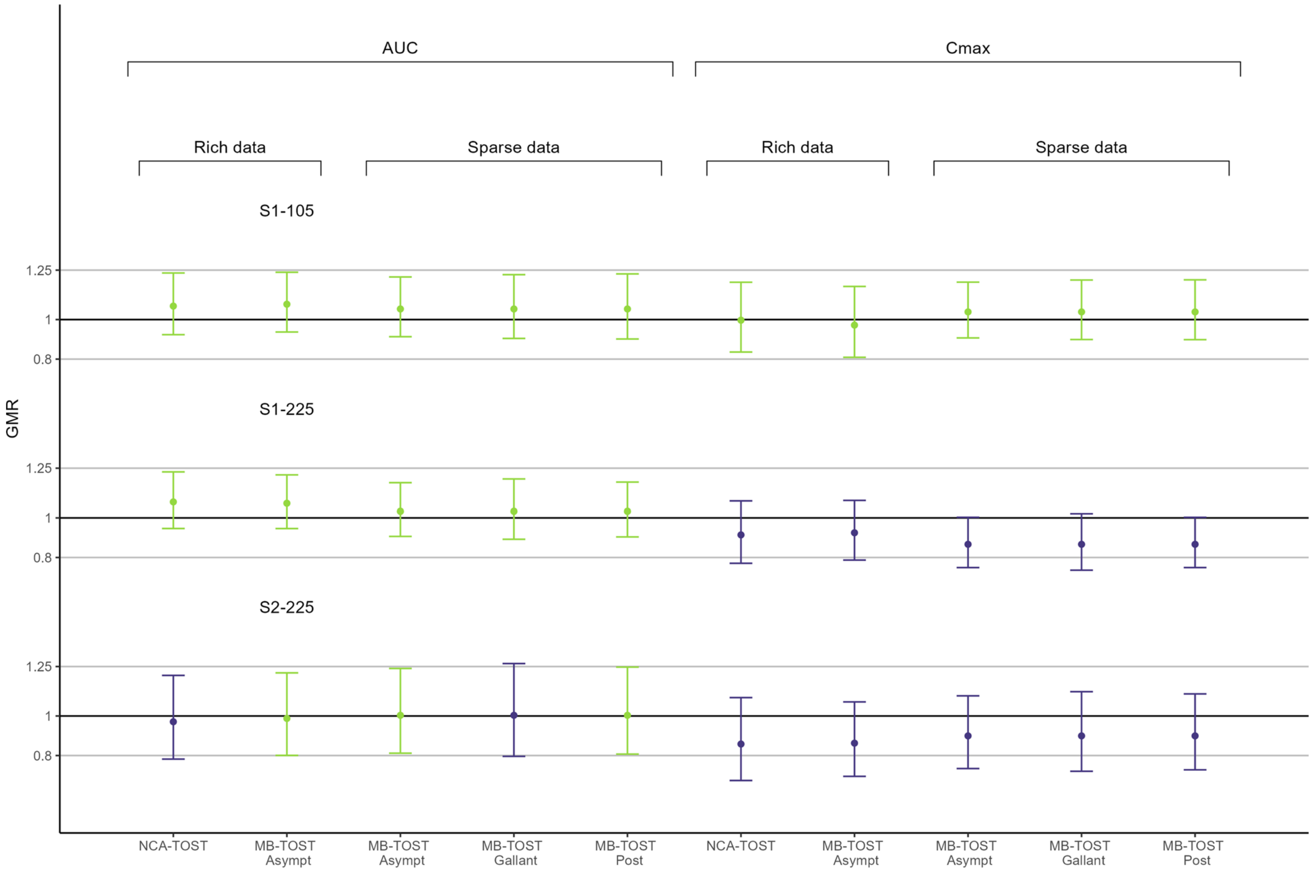


Fig. 2 Geometric mean ratios (GMR) and their 90% confidence intervals for AUC and C_{max}, with NCA-TOST and MB-TOST Asympt on observed data and with MB-TOST Asympt, Gallant and Post on sparse data S1-105 denotes Study 1 with dose=105mg reference and treatment arms and similarly for S1-225 and S2-225. *Grey lines are the limits of the null hypothesis interval, $GMR = 0.8$ and $GMR = 1.25$, and the black line represents $GMR = 1$. PK equivalence is shown as green intervals while blue intervals highlight the parameters and datasets for which PK equivalence was not established*

apparent volume of the peripheral compartment, *Q*/*F*: apparent inter-compartmental clearance) with linear absorption (*ka*: absorption constant) and elimination (*CL*/*F*: apparent clearance constant) with an absorption delay (T_{lag}), was selected to be the best (among the considered candidates) at describing the drug PK across studies/arms (taken as three separate datasets). A treatment effect was estimated on all 6 parameters (θ_{Tlag}, θ_{ka}, $\theta_{CL/F}$, $\theta_{V1/F}$, $\theta_{Q/F}$, and $\theta_{V2/F}$). On all datasets, based on BIC, the inter-individual random effect on V_2/F was withdrawn, and a correlation between the inter-individual random effects of *CL*/*F* and V_1/F was estimated. On S1-105 and S1-225, the error model was multiplicative. On S1-225, no inter-individual random effect was kept on *Q*/*F*. On S2-225, the error model was combined. The models selected were therefore very similar. Table 4 in Appendix 3 gives the parameter estimates obtained across datasets. As shown in Fig. 2, illustrating the GMR and their confidence intervals in the different datasets investigated, the different methods gave consistent results: for S1-105, with both NCA-TOST and MB-TOST Asympt, the 90% confidence interval of the GMR of AUC and C_{max} fell within [0.8; 1.25], but for S1-225, equivalence could not be shown on C_{max} with either of the two methods. On S2-225, equivalence could not be shown on C_{max} with both methods. For AUC, equivalence was shown using MB-TOST but not using NCA-TOST, although the estimates were close (MB-TOST Asympt: 90% CI=[0.801;1.218], p-value=0.049; NCA-TOST: 90% CI=[0.782;1.205], p-value=0.070). The data used to produce Fig. 2 are provided in Table 5 in Appendix 3.

The sparse design optimised using *PFIM* led to the following sampling scheme: 0.25, 3, 7, 20, 84 days post dose for S1-105, 0.25, 4, 20, 42, 84 days for S1-225, and 0.04, 4, 13, 42, 84 days post dose for S2-225. The selected PK model was a one compartment model with linear absorption and an absorption delay on the two S1 datasets, and a one compartment model with zero order absorption and no absorption delay on S2. Again, a treatment effect was estimated on all apparent parameters in each case. On all datasets, a correlation between the inter-individual random effects of *CL*/*F* and *V*/*F* was selected. On S1-105 and S1-225, the error model selected was multiplicative. On S2-225, the error model selected was combined. On S1-225 and S2-225, no inter-individual random effect was kept on T_{lag}. Table 4 in Appendix 3 gives the parameters estimated on all these subsets. Although the PK models selected on the sparse data were different from the ones selected on the observed data, the results of the equivalence study using MB-TOST were consistent, across all computation methods of SE, and comparable to those obtained on rich design (Fig. 2).

Fig. 6 shows the VPC and Fig. 7 reports the normality of residuals for S1-225 original and sparse design. These goodness of fit plots have also been checked for S1-105 and S2 (not shown).

Simulation study

Methods

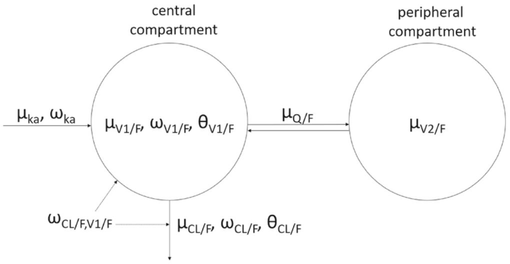
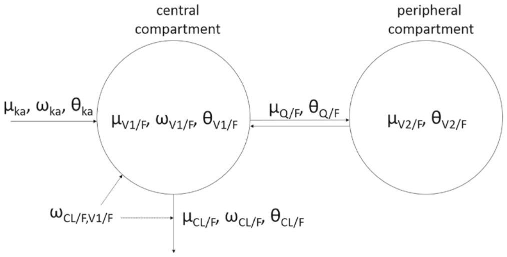
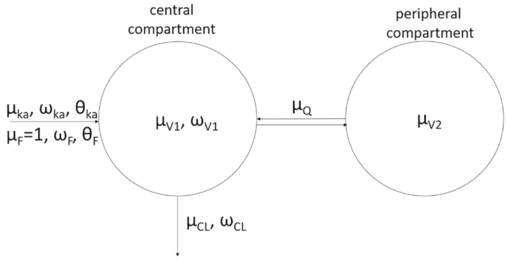
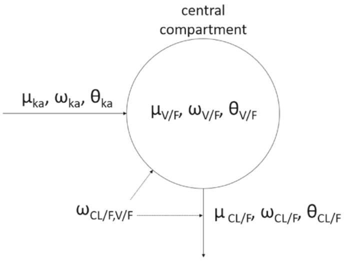
The real case study inspired our simulation settings with rich and sparse design. We simulated parallel studies with reference and test treatment arms, 24 participants per arm. The vector of rich sampling times was taken from S1-225 : 0.25, 1, 2, 3, 4, 7, 13, 20, 42, 63, and 84 days post dose.

The PK model used to simulate data was the one selected to describe the data of the reference arm of S1-225, corresponding to a two-compartment model with linear absorption and elimination. We removed the absorption delay. Moreover, the simulation study was performed prior to the availability of the data for publication. At the time, we only had access to scaled values of the doses that were divided by 15. Table 1 gives a graphical representation of the model simulated, and Table 2 gives the values of the fixed, random and error parameters simulated that were taken from the fit of S1-225.

Different levels of treatment effects were simulated on the apparent parameters, in order to get a treatment effect on AUC and C_{max} at the desired levels. To compute type I errors, we simulated data with treatment effects on AUC and C_{max} at boundaries of the null hypothesis, log(0.8) and log(1.25). These scenarios are denoted as $H_{0:0.8}$ and $H_{0:1.25}$, respectively. To study the power, we simulated data with treatment effects on AUC and C_{max} at and close to 0 (log(0.9), log(1) and log(1.11)). These scenarios are denoted as $H_{1:0.9}$, $H_{1:1}$ and $H_{1:1.11}$. The treatment effects were simulated on clearance (*CL*/*F*) and central volume (V_1/F), with no treatment effect on *ka*, *Q*/*F* and V_2/F. In practice, the treatment effect on *CL*/*F* was fixed (e.g. $\theta_{CL/F} = log(0.8)$ to get $\theta_{AUC} = log(1.25)$) and then the treatment effect on *V*1/*F* was varied to obtain the desired treatment effect on C_{max} without impacting the treatment effect on AUC. Table 3 gives the values of the different levels of treatment effects simulated. For each of the 5 treatment effects, 1000 datasets were simulated.

On rich design simulations, we compared the performances of NCA-TOST and MB-TOST Asympt in terms of type I error and study power. We first fitted the simulated structural PK model, estimating treatment effects on all 5 apparent parameters (referred to as model 2cpt_par). We also explored the performance of MB-TOST Asympt when modeling the treatment effects differently, i.e., two-

Table 1 Graphical representation of the model simulated and the models fitted on the rich and sparse design simulations, with the corresponding fixed and treatment effects and inter-individual variability parameters

<table>
<tr><td>Model simulated</td><td colspan="2"> </td></tr>
<tr><td>Models fitted on rich design</td><td rowspan="2">2cpt_par
 </td><td>2cpt_F
 </td></tr>
<tr><td>Models fitted on sparse design</td><td>1cpt_par
 </td></tr>
</table>

The graphical representation 1*cpt_par* corresponds to the third model presented in Appendix 1 (one compartment model with linear absorption and elimination) and the three other graphical representations correspond to the fifth model presented in Appendix 1 (two compartment model with linear absorption and elimination)

Table 2 Fixed coefficient values for fixed effects and standard deviations of the inter-individual random effects and residual errors, under which data were generated in the simulation study

μ_{ka} (d)	$\mu_{CL/F}$ (L.d^{-1})	$\mu_{V_1/F}$ (L)	$\mu_{Q/F}$ (L.d^{-1})	$\mu_{V_2/F}$ (L)
0.45	0.04	0.96	0.03	0.34
ω_{ka} (%)	$\omega_{CL/F}$ (%)	$\omega_{V_1/F}$ (%)	ρ_{CL/V_1}	σ_b (%)
57	26	36	0.8	15

Table 3 Treatment effects simulated on *CL*/*F* and V_1/F and GMR obtained on *AUC* and C_{max} on each simulation scenario

Scenario	Treatment effect on		GMR on	
	CL/*F*	V_1/F	*AUC*	C_{max}
$H_{0:0.8}$	log(1.25)	log(1.279)	0.8	0.8
$H_{1:0.9}$	log(1.11)	log(1.124)	0.9	0.9
$H_{1:1}$	log(1)	log(1)	1	1
$H_{1:1.11}$	log(0.9)	log(0.889)	1.11	1.11
$H_{0:1.25}$	log(0.8)	log(0.778)	1.25	1.25

compartment model with treatment effects estimated on the absorption parameter only, i.e., *ka*, and an additional scale/bioavailability parameter defined by *F*, with μ_F fixed to 1, and ω_F estimated (called hereafter 2cpt_F). Table 1 represents the structure of both models fitted to the rich design data.

In a second step, we also analysed sparse optimal design subsets, using PFIM: we selected 5 time-points, assuming 2cpt_par was true. The same 5 time points were selected

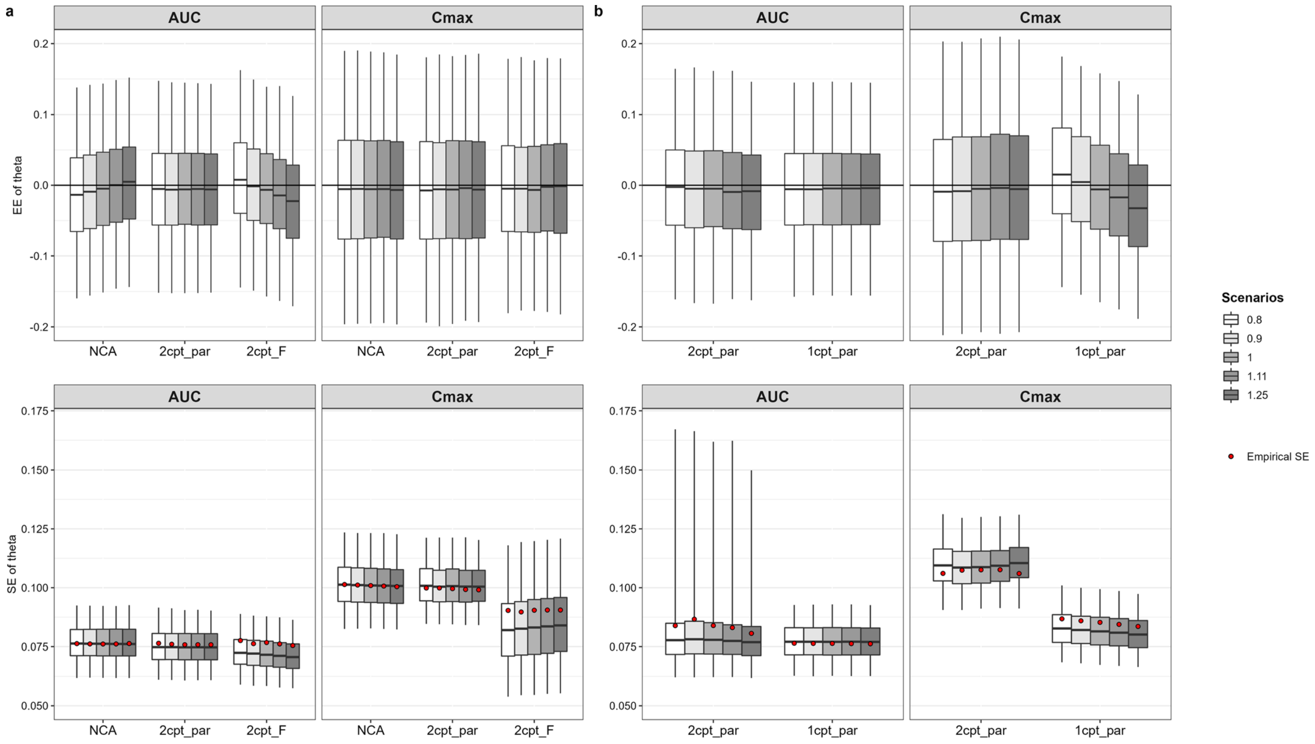


Fig. 3 Boxplots of estimation errors (EE) (top row) and standard errors (SE) (bottom row) of the treatment effects estimated on *AUC* and C_{max}, on (a) rich design simulations with NCA-TOST and MB-TOST Asympt, using the simulated PK structural model and treatment effects estimated and all apparent parameters (2cpt_par) or only on *ka* and *F* (2cpt_F), and (b) sparse design simulations with MB-TOST Asympt using the simulated PK structural model (2cpt_par) or a misspecified one compartment model (1cpt_par), with treatment effects estimated on all apparent parameters

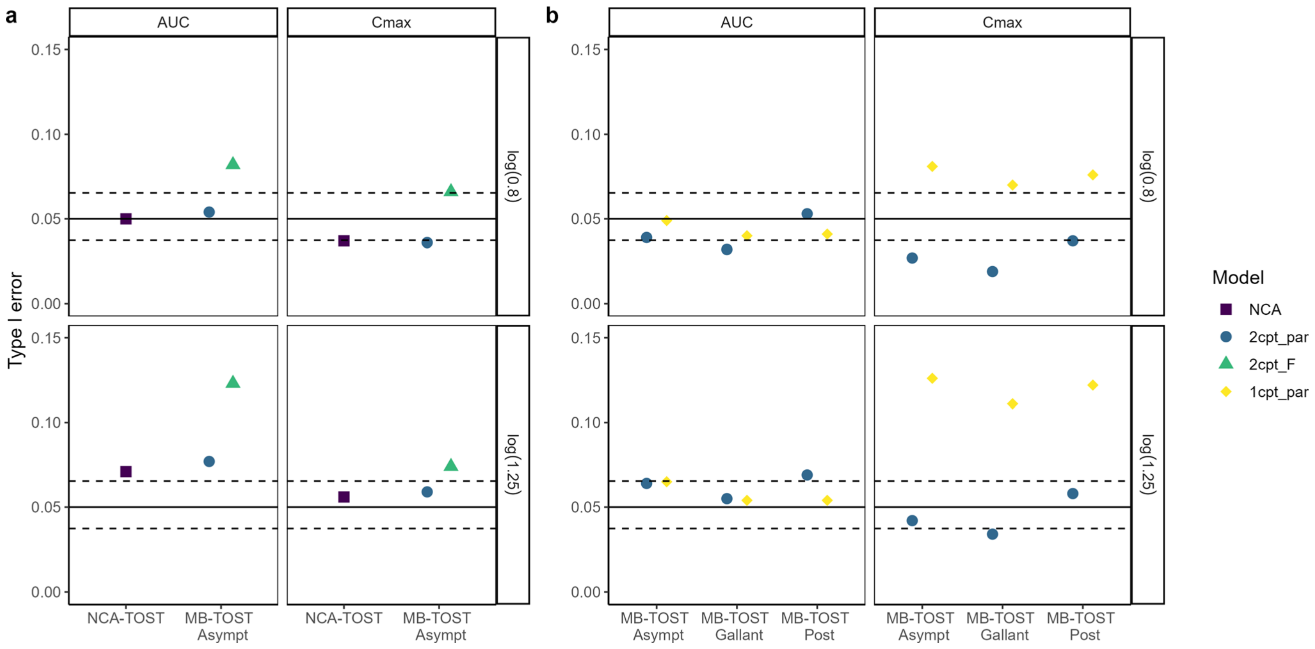


Fig. 4 Type I errors for *AUC* and C_{max}, under $H_{0:0.8}$ and $H_{0:1.25}$, on (a) rich design simulations with NCA-TOST and MB-TOST Asympt, and on (b) sparse design simulations with MB-TOST Asympt, Gallant and Post

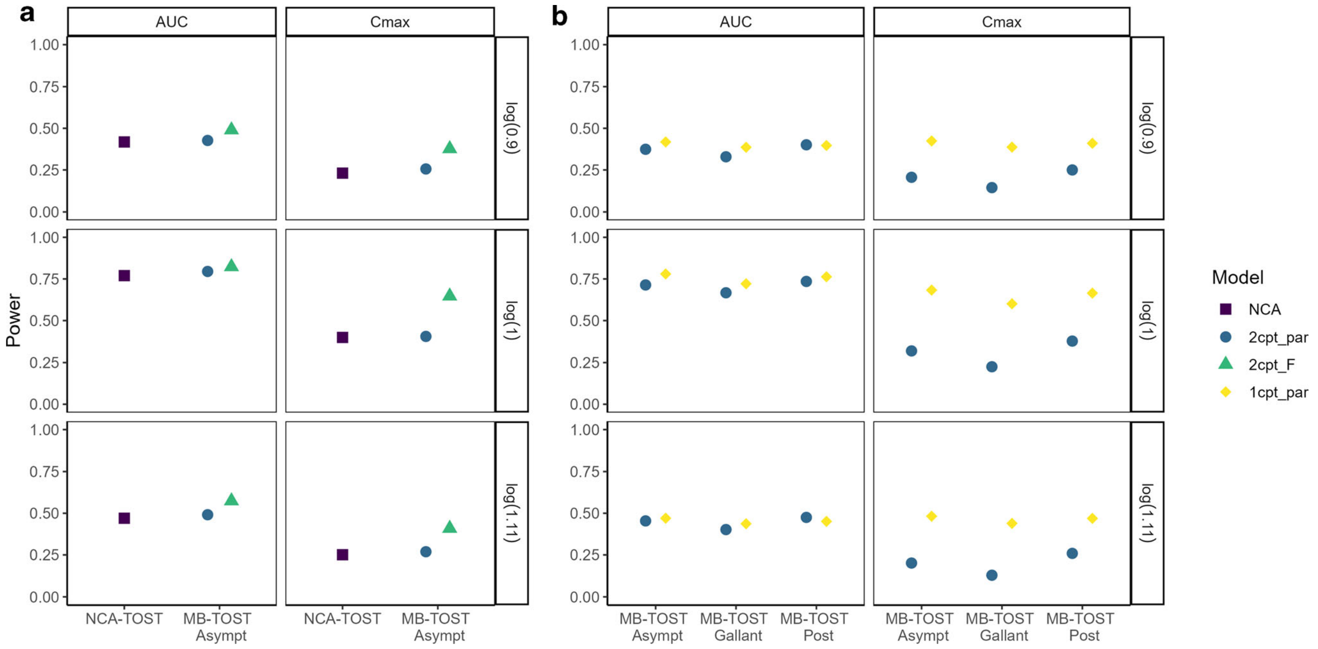


Fig. 5 Study power for AUC and C_{max}, under $H_{1:0.9}$, $H_{1:1}$ and $H_{1:1.11}$, on (a) rich design simulations with NCA-TOST and MB-TOST Asympt, and on (b) sparse design simulations with MB-TOST Asympt, Gallant and Post

regardless of the level of treatment effect considered: 0.25, 7, 20, 42, and 84 days post dose. On these sparse design simulations, we challenged MB-TOST by exploring the impact of a structural PK model misspecification: the model used to fit the data was either 2cpt_par or a misspecified one-compartment model with treatment effects estimated on all apparent parameters (1cpt_par). Table 1 represents the models fitted on sparse design simulations. As on the case study, three methods of computation of the SE were used on the sparse design simulations: Asympt, Gallant and Post.

We also explored the relevance of a PK model selection step, on the reference arm, on the BIC, prior to the equivalence test, on rich and sparse design simulations (two models to compare in each case). We observed the impact of this approach in terms of type I error.

Estimation Errors (EE) and Standard Errors (SE) of treatment effects were computed to evaluate the agreement between the estimations of NLMEM and the real values under which we simulated the data. Empirical SE were computed as the standard deviation on the 1000 estimates of each parameter in each scenario.

Implementation

A script detailing the analysis of one simulated dataset with *saemix* and *stan* is available on Zenodo (https://doi.org/10.5281/zenodo.6500556).

Results

Rich design

Figure 3a shows the boxplots of estimation errors (EE, top) and standard errors (SE, bottom) of the treatment effects on AUC and C_{max} in the different simulation scenarios with a rich design. We see that the treatment effects estimated with 2cpt_par (the structure of which is similar to the one of the model we simulated except treatment effects are estimated on all parameters) showed no bias and good precision : the EE were close to 0 and the estimated SE were close to the empirical SE. As expected on this rich design, NCA also provided good estimations of the treatment effects.

Figure 4a shows the type I errors of the TOST for AUC and C_{max} using NCA or a MB approach on rich design. The type I errors obtained with MB-TOST Asympt, using 2cpt_par, were similar to those obtained with NCA-TOST and close to the nominal value of 5%.

When we modelled the treatment effects differently from how they were simulated (i.e., using the misspecified model 2cpt_F), the model misspecification led to unsatisfactory results: the graph of EE (Fig. 3a Top) shows that the treatment effect on AUC was underestimated. In the scenario $H_{0:0.8}$, the relative bias in the estimation of the treatment effect on AUC is − 0.038, 0.016, and 0.016 for 2cpt_F, 2cpt_par, and NCA, respectively. In the scenario $H_{0:1.25}$, the relative bias in the estimation of the treatment effect on AUC is − 0.104, − 0.021, and − 0.030 for

2cpt_F, 2cpt_par, and NCA, respectively. The asymptotic SE boxplots appear lower than the empirical SE, though the relative root mean square errors (RMSE) are approximately − 0.35 for both 2cpt_par, 2cpt_F, and NCA, respectively. Increasing bias led to inflated type I errors we see in Fig. 4a.

A selection step using the BIC, prior to the test, on reference data helped in correcting the bias. Indeed, the difference of BIC between 2cpt_par and 2cpt_F ranged from − 22.1 to 10.3 with a median of − 3.4. The simulation model was found in 85% of the cases thanks to the selection procedure. Consequently, the type I error of MB-TOST was within the 95% prediction interval of the nominal value of 0.05 for each simulated level of treatment effect .

Study power for each study design, with NCA or MB-TOST, was low due to the parallel design of the clinical trial and the sample size (N=24 per arm, see Fig. 5).

Sparse design

On the simulations with sparse design, the treatment effects were still well estimated using 2cpt_par (Fig. 3b). Figure 4b shows the type I errors on sparse simulations with the MB approach where MB-TOST Asympt led to type I errors close to the 95% prediction interval of the nominal value of 0.05 with 2cpt_par.

When the structural PK model was misspecified, with only one compartment for the drug to distribute to, we observed a large inflation of the type I error on C_{max}, which we infer from Fig. 3b to be due to an underestimation of both the treatment effect and its SE. Indeed, in the scenario $H_{0:0.8}$, the relative bias in the estimation of the treatment effect on C_{max} is − 0.079 and 0.037 for 1cpt_par and 2cpt_par, respectively. In the scenario $H_{0:1.25}$, the relative bias in the estimation of the treatment effect on C_{max} is − 0.139 and − 0.015 for 1cpt_par and 2cpt_par, respectively. In the scenario $H_{0:0.8}$, the relative RMSE in the estimation of the treatment effect on C_{max} is − 0.40 and − 0.48 for 1cpt_par and 2cpt_par, respectively. In the scenario $H_{0:1.25}$, the relative RMSE in the estimation of the treatment effect on C_{max} is 0.40 and 0.47 for 1cpt_par and 2cpt_par, respectively.

MB-TOST Post gave results similar to MB-TOST Asympt (Fig. 4b). MB-TOST Gallant corrected the inflation of type I errors partly but could not correct for the bias in the estimations.

The numbers used to produce Figs. 4 and 5 are provided in Tables 6 and 7 in Appendix 3.

Here, a selection step using the BIC, prior to the test, to choose the number of compartments of the structural PK model on reference data, led to the selection of the simulated structural model in most cases (at least 99.0%). The difference of BIC between 2cpt_par and 1cpt_par ranged from -69.1 to 6.0 with a median of − 20.8. This allowed for control of type I error with MB-TOST.

We checked the assumption of normality of the test statistics under the null with Asympt in both rich and sparse design (data not shown).

Discussion

In this article, we compare the PK data gathered in participants treated with two formulations of gantenerumab, a monoclonal antibody for the treatment of Alzheimer's disease. The data used was originally collected to study the relative bioavailability of these two formulations. In this work, we use the data to compare the conventional NCA-TOST to the MB-TOST approach for PK equivalence testing. The data evaluated in our study is based on a parallel design instead of the more conventional crossover design in equivalence studies. The data is then used to generate a simulation study to explore the impact of sparse design and of model misspecification on the MB approach to test for PK equivalence.

After finding a dose effect on the pooled data, we performed the analyses separately on each study and dose evaluated. In our evaluation of these PK BE studies, we assume that the PK characteristics of the reference drug are well-known and the change of treatment does not affect the underlying PK structural model. In our simulations, we assumed the residuals are independent of the treatment covariate in the model. Also, we assumed that the study population would be adequately randomised to avoid imbalance between the treatment arms, so we did not evaluate the impact of covariates in our MB approaches. However, it is important to acknowledge that covariates would likely have a greater impact on a PK BE study with a parallel study design as compared to a crossover design. Moreover, adding covariates that affect the PK would decrease the between subject variability. Thus, future research may be warranted in this area. Using the real data, we evaluated the models selected to assess the assumptions made on the residuals as part of model building process. Also, the distributions of the MB-TOST statistics under the null from our simulations were verified as recommended by Shen et al. [12].

Using the original data, the NCA and MB-TOST approaches generally provide consistent results with the original rich design and the MB-TOST approach provides consistent results after sparsifying the data.

Previous studies by Dubois et al. and Reijers et al. have shown that MB approaches evaluating studies with a crossover design [19] and a parallel design [20], respectively, have performed as well as NCA methods for biosimilarity studies in the case of rich sampling. Dubois et al. also explored MB approaches on a sparse version of their data.

In the present study, we performed a simulation study to explore the influence of the design and model specification on the performance of the approaches and the relevance of the model selection.

Here, as in the previous works [8, 10], we only considered average BE. With average BE, by contrast with individual and population BE [21], we only take into account the average treatment effect at population level. Population BE would also take into account the variability of this effect, and individual BE would take into account the within-subject and subject-by-formulation variabilities. In this parallel study, population BE could be done, because variability is not correctly accounted for. Individual BE requires replicated cross-over studies so this approach would not be feasible on our data.

In the simulation study, when using the simulated model, MB-TOST Asympt achieved controlled type I errors that were similar to those obtained with NCA-TOST on rich designs. These results complement previous studies showing the efficiency of MB approaches for equivalence tests [8]. In general, regulatory authorities recommend that PK sampling includes 12–18 samples with at least three sampling points after the peak [3, 4]. These recommendations present unique challenges for PK studies with sparse designs. Indeed, the sparse design we extracted from the full design did not comply with those requirements, consequently we did not apply NCA-TOST to the datasets simulated with the sparse design. In this setting, we used MB-TOST as it relies on NLMEM which demonstrated improved accuracy of the estimates in particular when dealing with sparse designs [22]. However, Dubois et al. [8] showed that MB tests can lead to an inflation of the type I error because of an underestimation of the standard error of treatment effects when it is estimated asymptotically on sparse design with high variability. As such, Loingeville et al. proposed and evaluated methods of correction of the standard errors in MB studies, with satisfying results [10]. Notably, they compared the three methods we present here, along with a bootstrap method, but considering one model only (one-compartment) and without exploring the interest of model selection. One of the correction methods for SE in MB studies, Gallant, has been used outside the context of BE. To illustrate, Bertrand et al. [23] considered various methods of correcting the number of degrees of freedom in a Student distribution and found that the Gallant correction was a good compromise in NLMEM to handle the information carried by the number of subjects. In this research, the use of Gallant leads to the same reference distribution in MB-TOST Gallant as in NCA-TOST, instead of the Gaussian distribution used in MB-TOST Asympt and MB-TOST Post.

Our results showed that MB-TOST Asympt was adequate with sparse designs, with a slightly conservative type I error for C_{max} that was corrected using MB-TOST Post. Here, the Post method was used only as an alternative to produce SE. This algorithm is sensitive to the choice of prior distribution, and this could be further investigated. Nevertheless, the performance of the different MB methods were very similar on the sparse design in our work. Actually, we obtained asymptotic SE close to the empirical SE which explains that the results of the tests were not affected by the correction methods.

As the treatment effect on C_{max} is not directly linked to the parameters, we estimated it via simulations. We used an approximation simulating the treatment effect on a profile using the mean parameters; in Appendix 2, we provide a more computationally intense method. In this example, the first approximation gave equivalent results, but the second approximation should be used in the presence of higher variability.

A sparse design is commonly seen in PK BE studies for ophthalmic drug products where only one sample of aqueous humor is collected from the eye at a single time point. Currently, FDA recommends a non-parametric bootstrap NCA-based approach or a parametric method in the BE assessment for these drug products [5, 6]. In our assessment, we evaluated a study design with only five sampling points, which were optimally selected using PFIM. One limitation of this work is that we did not evaluate the performance of the classical NCA-TOST approach on sparse design as our focus was to evaluate the MB-TOST approach. The limitation of few sampling points per subject apply to both approaches as the NCA-TOST approach may become less accurate when there are few sample points whereas the MB-TOST approach may select a wrong PK structure model-based. Indeed, in our application study, the model parameter estimates varied considerably between the rich and sparse design (see Table 4 in Appendix 3).

The MB approach was previously evaluated only in simulations assuming the true model to be known [8–10]. In our present study, we investigate this question by fitting PK models different from the one used to simulate the data. The two-compartment model with treatment effects estimated on *ka* and *F* only, fitted on the rich designs, is the same structural PK model as the simulated one but with an alternative way of parameterising the treatment effects. It has already been used in other studies as the simulated model [11]. Here, it cannot properly fit the data as θ_F reflects a treatment effect on all distribution and elimination parameters, which does not agree with the way we simulated the data (i.e., without an effect on the peripheral clearance and volume). This explains why the effect on AUC is underestimated. With biosimilars, differences in the PK characteristics of a drug may be due to factors other than differences in the absorption phase. In contrast, the misspecified one-compartment model with treatment effects estimated on all apparent parameters, fitted on sparse designs, is a different PK structural model than the simulated one. The choice of the number of compartments is an essential step in structural PK model building. It is very sensitive to the study design and is therefore highly susceptible to misspecification. A less complex model would more likely be selected on a real study in case of non-optimised sparse design because of the lack of information. The treatment effect on AUC is still quite well estimated because it is a mean PK parameter, unlike C_{max} which is more sensitive to the misspecification because it is driven by only one point.

Adding a step of model selection on the reference data allowed to select the simulated model in most cases. When the simulated model is not selected, the difference of BIC between the models is very low. In this case, we assume that the misspecified model can adequately describe the data because the overall type I errors are controlled after the selection step. Most importantly, we mimic a real model development setting, where model selection is always part of the PK analysis. The selection of the model is based on data from the reference product only in order to avoid a bias in the MBBE evaluation from using test product data to fit the model used in the BE assessment. However, it is possible that using the reference arm for the model selection, and then for the assessment of a treatment effect, could inflate the type I error of the BE assessment. Therefore, this issue may warrant further investigation. Moreover, this can cause a problem if the underlying PK model is different in the test arm. Another limit of our simulation study is that we only selected between two different PK models. We could extend this approach to test and compare more features of the PK structural (absorption and elimination phases) and/or variability (random effects and residual errors) models as we performed in the real case study. We could also consider more complex data exhibiting, for example, double peaks which can be very challenging to evaluate, or that the magnitude of the variability depends on the treatment arm. It is likely that, in this case, the simulated model would not be recovered as often, potentially affecting the type I errors. However, the impact may not be very large if there were more candidate models in the selection step, as the models retained would have adequate goodness of fit. Hence, the estimated AUC and C_{max} would all be acceptable despite the diversity of underlying structural PK models. It would therefore be interesting to further evaluate the impact of small model variations on the model selection process and the ensuing ability to estimate C_{max} and AUC and the associated treatment effects. Competing models could also be taken into account via model averaging, which has been shown to work at least as well as model selection in dose finding studies using NLMEM [24, 25], as it allows to take into account the uncertainty on the model.

The methods presented in our study may be applied to PK similarity for large molecules (i.e., biologics) as well as PK BE studies for small molecules. By re-scaling the time frame, we could transpose our simulation settings and results to a BE study framework. In both cases, the test product or new drug contains the same active substance as the reference product, for which the PK is likely well characterised. To shorten the development phase of the new drug, it is recommended to demonstrate that there is no difference of treatment effect on the PK. In both cases, MB approaches may serve as an alternative method to NCA for sparse designs, and thus, are increasingly explored [26]. However, it is acknowledged that the performance of NCA and MB methods will drop in case of large inter-individual variability in PK or deviations from working assumptions.

Thus, we propose the use of MB-TOST when NCA-TOST may not be feasible or reasonable, as MB approaches are more informative and flexible than NCA.

This is consistent with recent proposals for MB approaches to serve as an alternative BE approach in generic drug development in situations for which conventional BE approaches are not feasible [27].

Conclusions

Our novel MB BE approach appears to be a robust alternative to the conventional NCA approach provided that the PK model is correctly specified and the test drug has the same PK structural model as the reference drug. Our simulation studies show that the selection of the PK model is a key step in the implementation of a model-based approach for PK equivalence studies. However, MB methods rely on numerous assumptions which need further investigation to determine when MB could offer a viable alternative to NCA in the context of PK BE studies.

Appendix 1 : Pharmacokinetic models equations and parameters

One compartment model with zero-order absorption, linear elimination

$$C(t) = \begin{cases} \frac{D}{Tk_0 CL}\left(1 - \exp(-\frac{CL}{V_1}t)\right) & \text{if } t \leq Tk_0 \\ \frac{D}{Tk_0 CL}\left(1 - \exp(-\frac{CL}{V_1}Tk_0)\right)\exp\left(-\frac{CL}{V_1}(t - Tk_0)\right) & \text{if } t > Tk_0 \end{cases}$$

with:

- $C(t)$ the concentration at time t;
- D the dose administered;
- Tk_0 the absorption duration;
- V_1 the volume of distribution of the compartment;
- CL the clearance of the drug;
- $k = \frac{CL}{V_1}$ the elimination rate constant.

Here, there are $l = 3$ parameters: $\mu = c(Tk_0, V_1, CL)$.

One compartment model with zero-order absorption, linear elimination, with a lag time

$$C(t) = \begin{cases} 0 & \text{if } t \leq T_{lag} \\ \frac{D}{Tk_0 CL}\left(1 - \exp(-\frac{CL}{V_1}(t - T_{lag}))\right) & \text{if } T_{lag} < t \leq T_{lag} + Tk_0 \\ \frac{D}{Tk_0 CL}\left(1 - \exp(-\frac{CL}{V_1}Tk_0)\right)\exp\left(-\frac{CL}{V_1}(t - T_{lag} - Tk_0)\right) & \text{if } t > Tk_0 \end{cases}$$

The lag time T_{lag} adds a period of latency before the concentration starts rising. It works the same for all models.

Here, there are $l = 4$ parameters: $\mu = c(Tk_0, V_1, CL, T_{lag})$.

One compartment model with first-order absorption, linear elimination

The model $1cpt\_par$ represented in Table 1 corresponds to the equation:

$$C(t) = \frac{D}{V_1}\frac{ka}{\frac{CL}{V_1} - ka}\left(\exp(-kat) - \exp(-\frac{CL}{V_1}t)\right) \tag{10}$$

with ka the absorption constant rate.

Here, there are $l = 3$ parameters: $\mu = c(ka, V_1, CL)$.

Two compartment model with zero-order absorption, linear elimination

$$C(t) = \begin{cases} \frac{D}{Tk_0}\left(\frac{A}{\alpha}(1 - \exp(-\alpha t)) + \frac{B}{\beta}(1 - exp(-\beta t))\right) & \text{if } t \leq Tk_0 \\ \frac{D}{Tk_0}\left(\frac{A}{\alpha}(1 - \exp(-\alpha Tk_0))\exp(-\alpha(t - Tk_0))\right. & \text{if } t > Tk_0 \\ \qquad \left. + \frac{B}{\beta}(1 - exp(-\beta Tk_0))\exp(-\beta(t - Tk_0))\right) & \end{cases}$$

with:

- $A = \frac{1}{V_1}\frac{k_{21} - \alpha}{\beta - \alpha}$ the first macro-constant;
- $B = \frac{1}{V_1}\frac{k_{21} - \beta}{\alpha - \beta}$ the second macro-constant;
- $\alpha = \frac{k_{21}k}{\beta}$ the first rate constant;
- $\beta = \frac{1}{2}\left(k_{12} + k_{21} + k - \sqrt{(k_{12} + k_{21} + k)^2 - 4k_{21}k}\right)$ the second rate constant;
- $k_{12} = \frac{Q}{V_1}$ the distribution rate constant between the principal and the peripheral compartment;
- $k_{21} = \frac{Q}{V_2}$ the distribution rate constant between the peripheral and principal compartment;
- Q the inter-compartmental clearance;
- V_1 the volume of distribution of the principal compartment;
- V_2 the volume of distribution of peripheral compartment.

Here, there are $l = 5$ parameters: $\mu = c(Tk_0, V_1, CL, V_2, Q)$.

Two compartment model with first-order absorption, linear elimination

The model used to generate the data in the simulation, and the models $2cpt\_par$ and $2cpt\_F$, are represented in Table 1 and correspond to the equation:

$$C(t) = D\big(A\exp(-\alpha t) + B\exp(-\beta t) - (A + B)\exp(-kat)\big) \tag{11}$$

with:

- $A = \frac{ka}{V_1}\frac{k_{21} - \alpha}{(ka - \alpha)(\beta - \alpha)}$;
- $B = \frac{ka}{V_1}\frac{k_{21} - \beta}{(ka - \beta)(\alpha - \beta)}$.

Here, there are $l = 5$ parameters: $\mu = c(ka, V_1, CL, V_2, Q)$.

Note: in the two compartment models under study here, the clearance occurs only from the central compartment via the clearance constant CL. The drug in the peripheral compartment can only return to the central compartment via the inter-compartmental clearance constant Q.

Parameterisation with *F*

Implicit in the equations above is the notion of bioavailability, defined as the fraction of dose reaching the system. Including bioavailability as an explicit parameter F corresponds to replacing D with $D \times F$ in the equations above. We can easily see from equations 10 and 11 that this is equivalent to dividing both CL and V_1 by F, so that the latter, when estimated from data collected after oral absorption, are called apparent clearance and volume, and sometimes denoted *CL*/*F* and V_1/F to show their dependency on *F*.

This leads to an alternative way of parameterising the model, by including *F* in the model. Because *F* cannot be identified without intravenous data, we fix the population value at *F*=1 and only allow for some inter-individual variability. We put a treatment effect only on the absorption parameters and *F*. Also, no correlation is allowed between the random effects of volumes and clearances, as these correlations are assumed to be carried by *F*. This parameterisation allows to compute fewer treatment effect coefficients.

Appendix 2: Method to compute the treatment effect on C_{max} and its SE

As part of a PK equivalence analysis, after fitting a NLMEM, we want to compute treatment effects on the PK parameters of interest and their SE when there is no explicit relationship with the direct parameters of the model. We simulate typical concentration versus time profiles, taking into account the variance covariance matrix of the fixed parameters.

Let $c(\mu, \theta)$ be the vector of fixed effects and treatment effects obtained with the NLMEM and M_F^{-1} the asymptotic variance-covariance matrix of the fixed effects and treatment effects, obtained by solving the Fisher Information Matrix of the model.

We simulate K parameter sets with a multivariate normal distribution.

$$c(\mu_k, \theta_k) \sim \mathcal{N}(c(\mu, \theta), M_F^{-1})$$

with k=1,...,K, here K=1000.

For each parameter set, we compute a profile of concentrations with the population parameters and a short time step, under reference treatment and under test treatment. For example, with a two-compartment model with first order absorption:

$$C_k^R = C(time, ka_k, CL_k, V1_k, Q_k, V2_k)$$

and

$$\begin{aligned} C_k^T =& C(time, ka_k\ e^{\theta_{ka,k}}, CL_k\ e^{\theta_{CL,k}}, V1_k\ e^{\theta_{V1,k}}, \\ & Q_k\ e^{\theta_{Q,k}}, V2_k\ e^{\theta_{V2,k}}) \end{aligned}$$

Then we compute the treatment effect as the log ratio of the PK parameter of interest under test and reference treatment. For instance, with $C_{max,k}^R$ the maximum over the vector C_k^R and $C_{max,k}^T$ the maximum over the vector C_k^T:

$$\theta_{C_{max,k}} = log\left(\frac{C_{max,k}^T}{C_{max,k}^R}\right)$$

We obtain a vector of K estimated treatment effects. We estimate the global treatment effect as the mean of this vector $mean(\theta_{Cmax,k})$, and its standard error as the standard deviation of this vector $sd(\theta_{Cmax,k})$.

Consequently, the geometric mean ratio is computed as the exponential of the point estimate of θ computed, $GMR_{Cmax} = exp(mean(\theta_{Cmax,k}))$ for instance.

We evaluate the performance of this method on the estimated standard error of the treatment effect on AUC, because it has an explicit formulation with direct parameters:

$$\begin{aligned} AUC =& \frac{D}{CL} \\ AUC_k^R =& \frac{D}{CL_k} \\ AUC_k^T =& \frac{D}{CL_k\ e^{\theta_{CL,k}}} \\ \theta_{AUC_k} =& log\left(\frac{AUC_k^T}{AUC_k^R}\right) = -\theta_{CL,k} \\ \mathbb{E}(sd(\widehat{\theta_{AUC}})) =& \mathbb{E}[sd((\theta_{CL,k}))] = sd(\theta_{CL}) \end{aligned}$$

That shows that the method would give an estimate of the standard error of θ_{AUC} consistent with the method based on the explicit link with the direct parameters. However, we compute θ_k as the treatment effect on a concentration profile in the mean parameters. The definition of θ is the mean of the treatment effect on each individual profile. In this example, the relationship between θ_{AUC} and θ_{CL} is linear, so these two quantities are equal, but this does not apply for C_{max}. A more accurate simulation method would take into account the interindividual variability by simulating individual time-concentration profiles using the variance-covariance matrix of the random effects. This method would be much more computationally intensive because simulating too few participants would lead to the poor estimation of the variability of the treatment effect, even more if the random variability of the direct parameters influencing C_{max} is high.

In this study, a comparison between the two approximations showed us that they gave similar results in terms of estimation of θ for C_{max} and its SE: it seems its relationship with the direct parameters treatment effect is close enough to linear, so we decided to use it for its computational conservativeness. It is likely that in a study with more random effects, for instance if there is some intra-individual variability, the second method would be preferable.

We also checked that this simulation method gave results similar from the FIM-based method for θ_{AUC}, but decided to keep the FIM-based method for θ_{AUC} because it is less time-consuming.

Appendix 3: Tables

Table 4 Parameters estimates and treatment effect coefficients (relative standard errors), given by *saemix* on all separate studies, with the original and sparse design

	Parameters (RSE, %)					
	S1-105 Rich	S1-105 Sparse	S1-225 Rich	S1-225 Sparse	S2-225 Rich	S2-225 Sparse
μ_{Tk0}						3.892 (7.6)
θ_{Tk0}						0.076 (141.7)
μ_{ka} (d)	0.327 (16.0)	0.844 (18.0)	0.469 (15.9)	2.947 (57.5)	0.361 (12.6)	
θ_{ka}	0.418 (54.0)	0.271 (110.7)	0.272 (82.6)	-1.220 (50.9)	0.019 (953.5)	
$\mu_{CL/F}$ (L.d^{-1})	0.632 (6.0)	0.622 (6.1)	0.632 (5.4)	0.621 (5.6)	0.681 (8.8)	0.698 (8.9)
$\theta_{CL/F}$	− 0.075 (113.7)	− 0.052 (166.3)	− 0.070 (108.6)	− 0.032 (245.9)	0.013 (1013.4)	− 0.003 (3806.2)
$\mu_{V_1/F}$ (L)	11.611 (14.8)	21.858 (6.0)	14.698 (9.1)	19.709 (6.1)	15.615 (12.8)	23.181 (8.6)
$\theta_{V_1/F}$	0.200 (102.4)	− 0.014 (622.2)	0.194 (64.3)	0.077 (113.6)	0.095 (200.2)	0.108 (114.4)
$\mu_{Q/F}$ (L.d^{-1})	1.882 (30.2)		0.415 (30.9)		0.601 (28.9)	
$\theta_{Q/F}$	0.421 (114.3)		− 0.460 (140.4)		0.773 (51.1)	
$\mu_{V_2/F}$ (L)	9.343 (11.5)		4.828 (15.0)		6.234 (13.3)	
$\theta_{V_2/F}$	− 0.235 (90.9)		− 0.444 (67.1)		0.126 (166.2)	
$\mu_{T_{lag}}$ (d)	0.062 (26.5)	0.129 (13.0)	0.037 (31.8)	0.208 (11.1)	0.050 (23.9)	
$\theta_{T_{lag}}$	− 0.744 (68.9)	− 0.209 (149.1)	− 0.079 (608.5)	− 0.974 (37.5)	− 0.529 (69.7)	
ω_{Tk0}						0.226 (23.4)
ω_{ka}	0.492 (13.1)	0.548 (15.1)	0.659 (11.2)	0.807 (11.0)	0.448 (11.9)	
$\omega_{CL/F}$	0.287 (10.7)	0.283 (11.6)	0.257 (10.7)	0.261 (11.3)	0.438 (10.4)	0.441 (10.4)
$\omega_{V_1/F}$	0.430 (12.2)	0.255 (13.1)	0.328 (11.4)	0.268 (13.0)	0.526 (10.8)	0.419 (10.7)
$\omega_{Q/F}$	1.129 (17.0)				0.412 (41.0)	
$\omega_{T_{lag}}$	0.926 (19.8)	0.258 (42.6)	0.869 (25.0)		1.005 (14.8)	
ρ_{CL/V_1}	0.694 (34.1)	0.843 (31.5)	0.762 (30.5)	0.844 (31.3)	0.931 (26.7)	0.940 (26.3)
σ_a					0.064 (11.8)	0.048 (14.6)
σ_b	0.168 (3.9)	0.194 (7.4)	0.153 (3.7)	0.171 (7.2)	0.112 (4.9)	0.099 (9.5)

Table 5 Gantenerumab analysis—Geometric mean ratios (GMR), their 90% confidence interval and the *p* − *value* of the test, for *AUC* and C_{max}, with NCA-TOST and MB-TOST Asympt on original data and with MB-TOST Asympt, Gallant and Post on sparse data

Dataset	Design	Method	PK parameter	GMR	90% CI	*p*
S1-105	Rich	NCA-TOST	*AUC*	1.068	[0.924 ; 1.236]	**0.038**
			C_{max}	0.997	[0.836 ; 1.189]	**0.021**
		MB-TOST Asympt	*AUC*	1.077	[0.937 ; 1.239]	**0.040**
			C_{max}	0.972	[0.809 ; 1.167]	**0.040**
	Sparse	MB-TOST Asympt	*AUC*	1.054	[0.913 ; 1.215]	**0.025**
			C_{max}	1.039	[0.907 ; 1.189]	**0.012**
		MB-TOST Gallant	*AUC*	1.054	[0.905 ; 1.227]	**0.033**
			C_{max}	1.039	[0.899 ; 1.200]	**0.018**
		MB-TOST Post	*AUC*	1.054	[0.902 ; 1.231]	**0.035**
			C_{max}	1.039	[0.898 ; 1.201]	**0.018**
S1-225	Rich	NCA-TOST	*AUC*	1.080	[0.947 ; 1.231]	**0.034**
			C_{max}	0.914	[0.771 ; 1.085]	0.098
		MB-TOST Asympt	*AUC*	1.073	[0.946 ; 1.216]	**0.023**
			C_{max}	0.925	[0.787 ; 1.087]	0.070
	Sparse	MB-TOST Asympt	*AUC*	1.033	[0.906 ; 1.177]	**0.008**
			C_{max}	0.867	[0.749 ; 1.003]	0.184
		MB-TOST Gallant	*AUC*	1.033	[0.892 ; 1.196]	**0.017**
			C_{max}	0.867	[0.736 ; 1.021]	0.208
		MB-TOST Post	*AUC*	1.033	[0.904 ; 1.180]	**0.009**
			C_{max}	0.867	[0.749 ; 1.003]	0.184
S2-225	Rich	NCA-TOST	*AUC*	0.971	[0.782 ; 1.205]	0.070
			C_{max}	0.858	[0.674 ; 1.093]	0.314
		MB-TOST Asympt	*AUC*	0.988	[0.801 ; 1.218]	**0.049**
			C_{max}	0.863	[0.695 ; 1.071]	0.284
	Sparse	MB-TOST Asympt	*AUC*	1.003	[0.812 ; 1.240]	**0.044**
			C_{max}	0.899	[0.734 ; 1.102]	0.171
		MB-TOST Gallant	*AUC*	1.003	[0.796 ; 1.265]	0.059
			C_{max}	0.899	[0.720 ; 1.123]	0.190
		MB-TOST Post	*AUC*	1.003	[0.807 ; 1.247]	**0.048**
			C_{max}	0.899	[0.728 ; 1.111]	0.182

Significant p-values are highlighted in bold

Table 6 Type I errors for AUC and C_{max}, under $H_{0:0.8}$ and $H_{0:1.25}$, on rich (R) design simulations with NCA-TOST and MB-TOST Asympt, and on sparse (S) design simulations with MB-TOST Asympt, Gallant and Post

			Type I error	
			AUC	C_{max}
$R_{0.8}$	NCA-TOST		0.05	0.037
	2cpt_par	MB-TOST Asympt	0.054	0.036
	2cpt_F	MB-TOST Asympt	0.082	0.066
$R_{1.25}$	NCA-TOST		0.071	0.056
	2cpt_par	MB-TOST Asympt	0.077	0.059
	2cpt_F	MB-TOST Asympt	0.123	0.074
$S_{0.8}$	2cpt_par	MB-TOST Asympt	0.039	0.027
		MB-TOST Gallant	0.032	0.019
		MB-TOST Post	0.054	0.038
	1cpt_par	MB-TOST Asympt	0.049	0.081
		MB-TOST Gallant	0.040	0.070
		MB-TOST Post	0.041	0.077
$S_{1.25}$	2cpt_par	MB-TOST Asympt	0.064	0.042
		MB-TOST Gallant	0.055	0.034
		MB-TOST Post	0.069	0.058
	1cpt_par	MB-TOST Asympt	0.065	0.126
		MB-TOST Gallant	0.054	0.111
		MB-TOST Post	0.055	0.123

Table 7 Study power to detect a treatment effect on AUC and C_{max}, under $H_{1:0.9}$, $H_{1:1}$ and $H_{1:1.11}$, on rich (R) design simulations with NCA-TOST and MB-TOST Asympt, and on sparse (S) design simulations with MB-TOST Asympt, Gallant and Post

			Power	
			AUC	C_{max}
$R_{0.9}$	NCA-TOST		0.418	0.231
	2cpt_par	MB-TOST Asympt	0.427	0.256
	2cpt_F	MB-TOST Asympt	0.490	0.377
R_1	NCA-TOST		0.770	0.401
	2cpt_par	MB-TOST Asympt	0.795	0.407
	2cpt_F	MB-TOST Asympt	0.823	0647
$R_{1.11}$	NCA-TOST		0.470	0.251
	2cpt_par	MB-TOST Asympt	0.491	0.269
	2cpt_F	MB-TOST Asympt	0.574	0.409
$S_{0.9}$	2cpt_par	MB-TOST Asympt	0.374	0.206
		MB-TOST Gallant	0.329	0.144
		MB-TOST Post	0.409	0.251
	1cpt_par	MB-TOST Asympt	0.418	0.424
		MB-TOST Gallant	0.386	0.387
		MB-TOST Post	0.4399	0.411
S_1	2cpt_par	MB-TOST Asympt	0.714	0.320
		MB-TOST Gallant	0.667	0.225
		MB-TOST Post	0.739	0.384
	1cpt_par	MB-TOST Asympt	0.780	0.683
		MB-TOST Gallant	0.721	0.601
		MB-TOST Post	0.762	0.667
$S_{1.11}$	2cpt_par	MB-TOST Asympt	0.454	0.201
		MB-TOST Gallant	0.402	0.128
		MB-TOST Post	0.4731	0.255
	1cpt_par	MB-TOST Asympt	0.470	0.482
		MB-TOST Gallant	0.437	0.439
		MB-TOST Post	0.450	0.467

Appendix 4: Figures

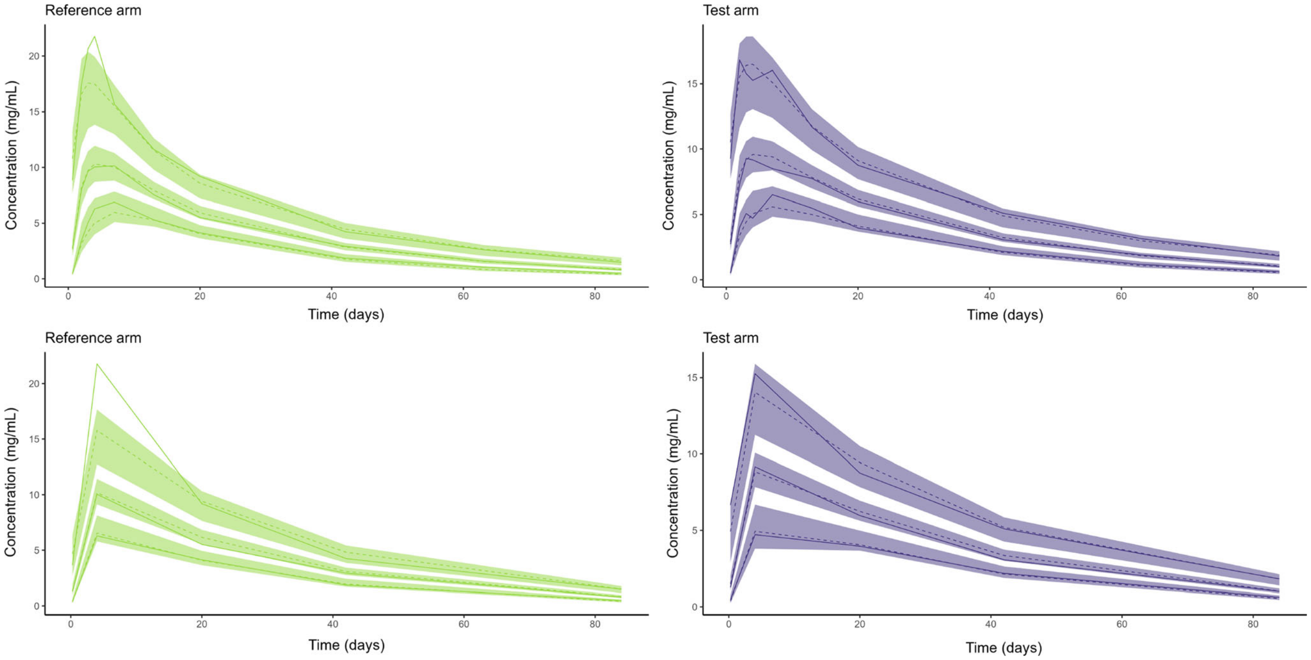


Fig. 6 Visual predictive check for the S1-225 study reference (left) and test (right) arm, on original (top) and sparse (bottom) design. Note: the predicted 5%, 50% and 95% percentiles are shown as dashed lines; the observed percentiles as solid lines (colour figure online)

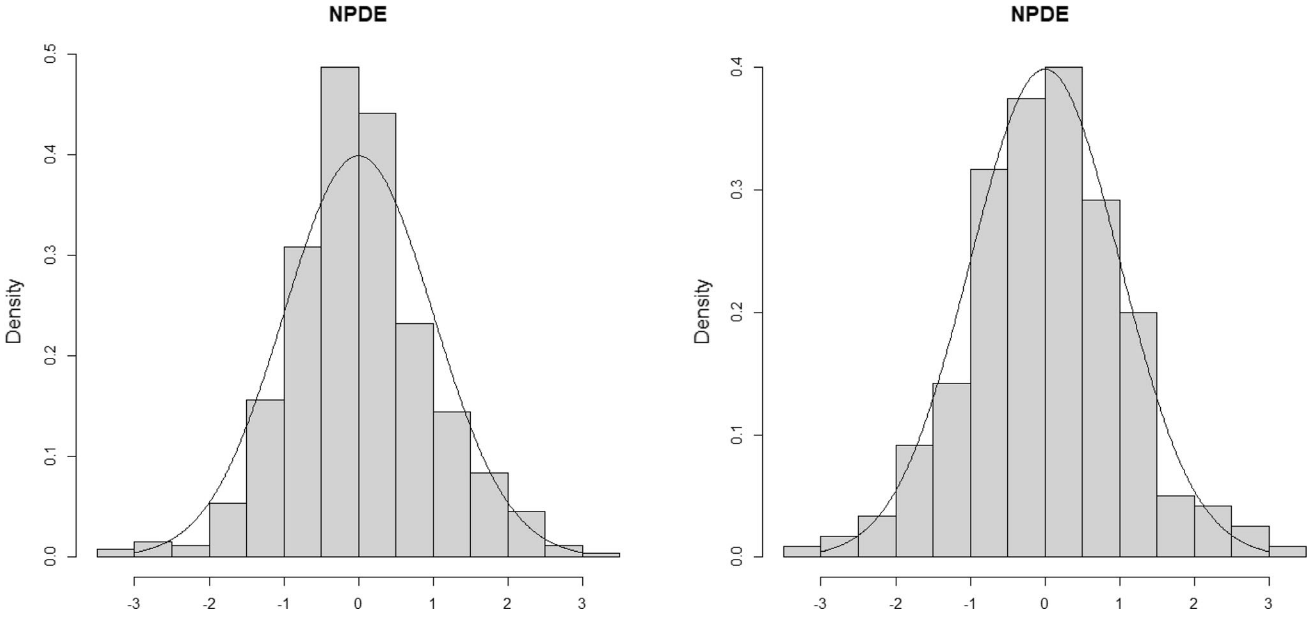


Fig. 7 Distributions of normalised predictive distribution errors (NPDE) for the S1-225 study on original (left) and sparse (right) design

Definition of Visual Predictive Checks (from Monolix documentation) :

The VPC (Visual Predictive Check) offers an intuitive assessment of misspecification in structural, variability, and covariate models. The principle is to assess graphically whether simulations from a model of interest are able to reproduce both the central trend and variability in the observed data, when plotted versus an independent variable (typically time). It summarises in the same graphic the structural and statistical models by computing several quantiles of the empirical distribution of the data after having regrouped them into bins over successive intervals. More precisely, the goal is to compare the two following elements:

Empirical percentiles: percentiles of the observed data, calculated either for each unique value of time, or pooled by adjacent time intervals (bins).

Theoretical percentiles: percentiles of simulated data are computed from multiple Monte Carlo simulations with the model of interest and the design structure of the original dataset (i.e., dosing, timing, and number of samples). For each simulation, the same percentiles are computed across the same bins as for empirical percentiles. Prediction intervals for each percentile are then estimated across all simulated data and displayed as colored areas.

If the model is correct, the observed percentiles should be close to the predicted percentiles and remain within the corresponding prediction intervals.

Definition of normalised prediction distributions errors (NPDE) from Comets et al. [28]:

The cumulative distribution function (cdf) of the predictive distribution of the concentrations observed can be computed using Monte–Carlo simulations.

We define the prediction discrepancies (pd) as the value of this cdf at each observation.

pd are computed as the percentiles of each observation in the empirical distribution of the simulations.

By construction, pd are expected to follow $\mathcal{U}(0,1)$, but only in the case of one observation per subject; within-subject correlations introduced when multiple observations are available for each subject induce an increase in the type I error of the test. To correct for this correlation, we compute the empirical mean and empirical variance-covariance matrix over the simulations.

Decorrelation is performed simultaneously for simulated data and for observed data. Decorrelated pd are then obtained using the same formula but with the decorrelated data, and we call the resulting variables prediction distribution errors (pde).

If the number of Monte-Carlo simulations is large enough, the distribution of the prediction distribution errors should follow a uniform distribution over the interval [0,1] by construction of the cdf. Normalised prediction distribution errors can then be obtained using the inverse function of the normal cumulative density function. By construction, NPDE follow the $\mathcal{N}(0,1)$ distribution without any approximation and are uncorrelated within an individual.

Acknowledgements The authors would like to thank all their collaborators in the project ”Evaluation of model-based bioequivalence (MBBE) statistical approaches for sparse design PK studies” under the FDA contract 75F40119C10111. They also thank Hervé Le Nagard and Lionel de la Tribouille for the use of the CATIBioMed calculus facility. The illustrative example data were obtained from studies sponsored by F Hoffmann-La Roche. We thank the participants and investigators who participated in these studies. We are grateful to the gantenerumab project team, which allowed us to use these data and provided comments to improve the quality of this manuscript.

Disclaimer This work reflects the views of the author and should not be construed to represent the FDA’s views or policies.

Author contributions MG, EC and JB wrote the manuscript. FM, CH, SS, MD, KF, WS, GS, SG, LZ, LF and FM criticially revised the manuscript. FM and JB designed the research. MG, EC and JB performed the research. MG simulated and analysed the data.

Funding This work was supported by the U.S. Food and Drug Administration (FDA) under contract 75F40119C1011. The authors thank the FDA for this funding.

Declarations

Conflict of interest Francois Mercier and Carsten Hofmann are Roche employees. Francois Mercier and Carsten Hofmann declare no conflict of interest. IAME laboratory has PhD students funded by Roche but not for this project.

References

1. GDUFA (2022) Generic Drug User Fee Amendments (GDUFA) science and research priority initiatives for Fiscal Year (FY) 2022. https://www.fda.gov/media/154487/download
2. Schuirmann DJ (1987) A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. J Pharmacokinet Biopharm 15:657–680. https://doi.org/10.1007/BF01068419
3. U.S. Food and Drug Administration (2021) Bioequivalence studies with pharmacokinetic endpoints for drugs submitted under an ANDA guidance for industry. https://www.fda.gov/media/87219/download
4. European Medicines Evaluation Agency (2010) Guideline on the investigation of bioequivalence. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-investigation-bioequivalence-rev1_en.pdf
5. U.S. Food and Drug Administration (2016) Draft guidance on dexamethasone; tobramycin. https://www.accessdata.fda.gov/drugsatfda_docs/psg/Dexamethasone;%20Tobramycin_ophthalmic%20ointment_RLD%20050616_RV06-16.pdf
6. U.S. Food and Drug Administration (2018) Draft guidance on loteprednol etabonate. https://www.accessdata.fda.gov/

drugsatfda_docs/psg/Loteprednol%20Etabonate_draft_Ophthalmic%20drops%20susp_RLD%2020583_RC02-18.pdf

7. Dubois A, Gsteiger S, Pigeolet E, Mentré F (2010) Bioequivalence tests based on individual estimates using non-compartmental or model-based analyses: evaluation of estimates of sample means and type I error for different designs. Pharm Res 27:92–104. https://doi.org/10.1007/s11095-009-9980-5
8. Dubois A, Lavielle M, Gsteiger S, Pigeolet E, Mentré F (2011) Model-based analyses of bioequivalence crossover trials using the stochastic approximation expectation maximisation algorithm. Stat Med 30:2582–2600. https://doi.org/10.1002/sim.4286
9. Möllenhoff K, Loingeville F, Bertrand J, Nguyen TT, Sharan S, Sun G, Grosser S, Zhao L, Fang L, Mentré F et al (2022) Efficient model-based bioequivalence testing. Biostatistics 23. https://doi.org/10.1093/biostatistics/kxaa026
10. Loingeville F, Bertrand J, Nguyen T, Sharan S, Feng K, Sun W, Han J, Grosser S, Zhao L, Fang L, Möllenhoff K, Dette H, Mentré F (2020) New model-based bioequivalence statistical approaches for pharmacokinetic studies with sparse sampling. AAPS J 22:141. https://doi.org/10.1208/s12248-020-00507-3
11. Hooker A (2022) Model averaging for model-based bioequivalence design and analysis. In: World Conference of Pharmacometrics WCOP 2022
12. Shen M, Russek-Cohen E, Slud E (2016) Distributional assumptions for pharmacokinetic summary statistics based on simulations with compartmental models. J Biopharm Stat. https://doi.org/10.1080/10543406.2016.1222535
13. Gallant AR (1975) Seemingly unrelated nonlinear regressions. J Econom 3:35–50. https://doi.org/10.1016/0304-4076(75)90064-0
14. Nguyen T, Mouksassi M, Holford N, Al-Huniti N, Freedman I, Hooker A, John J, Karlsson M, Mould D, Ruixo JP, Plan E, Savic R, van Hasselt J, Weber B, Zhou C, Comets E, Mentré F (2017) Model evaluation of continuous data pharmacometric models: metrics and graphics. CPT Pharmacomet Syst Pharmacol 6:87–109. https://doi.org/10.1002/psp4.12161
15. Dumont C, Lestini G, Le Nagard H, Mentré F, Comets E, Nguyen TT (2018) the PFIM group, PFIM 4.0, an R program for design evaluation and optimisation in nonlinear mixed effect models. Comput Methods Prog Biomed 156:217–229. https://doi.org/10.1016/j.cmpb.2018.01.008
16. Delyon B, Lavielle M, Moulines E (1999) Convergence of a stochastic approximation version of em algorithm. Ann Stat 27:94–128. https://doi.org/10.1214/aos/1018031103
17. Comets E, Lavenu A, Lavielle M (2017) Parameter estimation in nonlinear mixed effect models using saemix, an R implementation of the SAEM algorithm. J Stat Softw 80:1–41. https://doi.org/10.18637/jss.v080.i03
18. Denney W, Duvvuri S, Buckeridge C (2015) Simple, automatic noncompartmental analysis: the PKNCA R package. J Pharmacokinet Pharmacodyn 42(11–107):S65. https://doi.org/10.1007/s10928-015-9432-2
19. Dubois A, Gsteiger S, Balser S, Pigeolet E, Steimer J, Pillai G, Mentré F (2012) Pharmacokinetic similarity of biologics: analysis using nonlinear mixed-effects modeling. Clin Pharmacol Ther 91:234–242. https://doi.org/10.1038/clpt.2011.216
20. Reijers J, van Donge T, Schepers F, Burggraaf J, Stevens J (2016) Use of population approach non-linear mixed effects models in the evaluation of biosimilarity of monoclonal antibodies. Eur J Clin Pharmacol 72:1343–1352. https://doi.org/10.1007/s00228-016-2101-6
21. U.S. Food and Drug Administration (2001) Statistical approaches to establishing bioequivalence. https://www.fda.gov/media/70958/download
22. Hu C, Moore KHP, Kim YH, Sale ME (2004) Statistical issues in a modeling approach to assessing bioequivalence or PK similarity with presence of sparsely sampled subjects. J Pharmacokinet Pharmacodyn 31:321–339. https://doi.org/10.1023/B:JOPA.0000042739.44458.e0
23. Bertrand J, Comets E, Chenel M, Mentré F (2012) Some alternatives to asymptotic tests for the analysis of pharmacogenetic data using nonlinear mixed effects models. Biometrics 68:146–155. https://doi.org/10.1111/j.1541-0420.2011.01665.x
24. Buatois S, Ueckert S, Frey N, Retout S, Mentré F (2018) Comparison of model averaging and model selection in dose finding trials analyzed by nonlinear mixed effect models. AAPS J 20:56. https://doi.org/10.1208/s12248-018-0205-x
25. Gonçalves A, Mentré F, Lemenuel-Diot A, Guedj J (2020) Model averaging in viral dynamic models. AAPS J 22:48. https://doi.org/10.1208/s12248-020-0426-7
26. Yue C, Ozdin D, Selber-Hnatiw S, Ducharme M (2019) Opportunities and challenges related to the implementation of model-based bioequivalence criteria. Clin Pharm Ther 105:350–362. https://doi.org/10.1002/cpt.1270
27. Zhao L, Kim MJ, Zhang L, Lionberger R (2018) Generating model integrated evidence for generic drug development and assessment. Clin Pharmacol Ther 105:338–349. https://doi.org/10.1002/cpt.1282
28. Comets E, Brendel K, Mentré F (2008) Computing normalised prediction distribution errors to evaluate nonlinear mixed-effect models: the NPDE add-on package for R. Comput Methods Programs Biomed. https://doi.org/10.1016/j.cmpb.2007.12.002

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.